

## Learning from human tutoring

Micheline T.H. Chi\*, Stephanie A. Siler, Heisawn Jeong,  
Takashi Yamauchi<sup>1</sup>, Robert G. Hausmann

*Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260, USA*

---

### Abstract

Human one-to-one tutoring has been shown to be a very effective form of instruction. Three contrasting hypotheses, a tutor-centered one, a student-centered one, and an interactive one could all potentially explain the effectiveness of tutoring. To test these hypotheses, analyses focused not only on the effectiveness of the tutors' moves, but also on the effectiveness of the students' construction on learning, as well as their interaction. The interaction hypothesis is further tested in the second study by manipulating the kind of tutoring tactics tutors were permitted to use. In order to promote a more interactive style of dialogue, rather than a didactic style, tutors were suppressed from giving explanations and feedback. Instead, tutors were encouraged to prompt the students. Surprisingly, students learned just as effectively even when tutors were suppressed from giving explanations and feedback. Their learning in the interactive style of tutoring is attributed to construction from deeper and a greater amount of scaffolding episodes, as well as their greater effort to take control of their own learning by reading more. What they learned from reading was limited, however, by their reading abilities. © 2001 Cognitive Science Society, Inc. All rights reserved.

---

### 1. Background and hypotheses

#### 1.1. General findings

Three general findings should be noted about the effectiveness of one-to-one tutoring. First, in one-to-one tutoring, students gain greater understanding, are more motivated, and work faster (Slavin, 1987). The average student in a tutoring situation achieved a performance gain ranging from 0.4 to 2.3 standard deviations (usually referred to as the 2 sigma

---

\* Corresponding author. Tel.: +1-412-624-7456.

E-mail address: [Chi@pitt.edu](mailto:Chi@pitt.edu) (M.T.H. Chi).

<sup>1</sup> Now at the Department of Psychology, Texas A&M University, College Station TX 77843.

effect) above the average student in a traditional transmission-based classroom (with a class size of 30 students; Bloom, 1984; Cohen, Kulik & Kulik, 1982). One main instructional difference between teachers and tutors is that tutors have the opportunity to pursue a given topic or problem until the students have mastered it. Even if we factor out this difference in pursuing mastery, tutoring still produces a gain of about one sigma rather than two (Bloom, 1984).

A second equally important finding, which is often overlooked, is that the tutors also learned the subject matter that they tutored. Cohen et al.'s meta-analysis (Cohen et al., 1982) found that tutors improved in 33 of the 38 studies that they reviewed. Juel's (Juel, 1991) study exemplified this kind of finding in her tutoring program for first and second graders and special education students who had reading difficulties. The tutors Juel used in her study were college-athletes, whose reading abilities were at the 9th grade level. However, after two semesters of tutoring, their mean grade equivalent went up to 13.1, significantly surpassing that of a control group.

A third important but again often overlooked finding is that tutors often do not have formal training in the skills of tutoring (Fitz-Gibbon, 1977). Tutoring skills refer to the pedagogical skills of knowing when to give feedback, scaffoldings, and explanations, when to hold back error corrections and allow the students to infer that an error has been made, and so forth. For example, in educational settings, everyday tutors are often peers and other paraprofessionals with little expertise in tutoring (Cohen et al., 1982), although they often have expertise in the content domain, such as algebra. Nevertheless, tutoring is still effective (Fantuzzo, King & Heller, 1992; Rogoff, 1990).<sup>1</sup>

In some sense, how could the tutors have formal training since we have little understanding of what tutoring skills are? This is confirmed by Cohen et al.'s (Cohen et al., 1982) meta-analysis, showing that the impact of tutoring on learning is not significantly related to the amount of tutor training or age differences between the tutor and the student. Basically, the typical tutor in a formal program (such as those run by colleges) has at best six hours of training. However, most of the training is not geared toward instructional skills, but toward directing the tutor's attention to other factors, such as sensitivity to the students' cultural differences (Reisner, Petry & Armitage, 1990). In short, in tutoring, both the tutors and the tutees gain in understanding; moreover, the tutors do not necessarily have expertise in the pedagogical skills of tutoring, although they do have expertise in the domain knowledge.

## *1.2. Three hypotheses to account for tutoring effectiveness*

### *1.2.1. Does tutoring effectiveness arise from the tutors' pedagogical skills? The tutor-centered pedagogical hypothesis*

According to Graesser, Person and Magliano (1995), a typical tutoring dialogue consists of the five broad steps that they have referred to as the "tutoring frame." A version adapted from their "tutoring frame" is described below:

- (1) Tutor asks an initiating question;
- (2) Student provides a preliminary answer;

- (3) Tutor gives (confirmatory or negative) feedback on whether the answer is correct or not;
- (4) Tutor scaffolds to improve or elaborate the student's answer in a successive series of exchanges (taking 5–10 turns, according to Graesser et al. 1995);
- (5) Tutor gauges student's understanding of the answer.

For instance, in the context of tutoring the topic of statistical design, the tutor may start with an initiating question (Step 1) such as “Now what is a factorial design?” (Graesser et al., 1995). On the basis of the student's response, “The design has two variables” (Step 2), the tutor gives feedback such as, “Right” (Step 3). The dialogue pattern of a classroom typically stops after this third step, consisting of an Initiation by the teacher, a Reply by the child, followed by an Evaluation by the teacher. This is generally referred to as the IRE (Griffin & Humphrey, 1978; Poole, 1990)

However, in tutoring, two additional steps are taken. In the fourth step, the tutor proceeds with a *scaffolding episode*, one that can take 5–10 turns. What exactly is a scaffolding episode? Three kinds of literature discuss the notion of scaffolding: human tutoring research, intelligent tutoring systems (ITS) research, and developmental research. We adapt the notion of scaffolding as discussed in the developmental literature. For example, in the Vygotskian developmental context (Vygotsky, 1978; Rogoff, 1986), scaffolding has been considered to be a pivotal kind of adult-child interchange in which the adult “guides” the child to develop and achieve to the child's fullest potential. Since tutoring is similar to adult-child interaction, scaffolding may be the pivotal step in tutoring as well. Translating this to the tutoring context, “guidance” means that in structuring the task, a tutor might decompose a complex task into simpler ones, do part of the task, initiate a task and let the student do the rest, remind the student of some aspect of the task, and so forth (Brown & Palinscar, 1989; Cole, 1978; Collins, Brown & Newman, 1989; Merrill, Reiser, Ranney & Trafton, 1992; Woods, Bruner & Ross, 1976). Thus, in this sense, we consider scaffolding to be *any kind of guidance that is more than a confirmatory or a negative feedback*. In other words, feedback alone can “guide” students, in the sense of encouraging students to stay on the same track of reasoning or problem solving (in the context of confirmatory feedback), or to change direction or goal (in the context of negative feedback). However, in this paper, we define scaffolding to exclude feedback-type of guidance (i.e., “guidance” is a more general term). A scaffolding episode then, is simply a multiturn dialogue containing scaffoldings and addressing the same concept or topic. Both of these definitions will be operationalized later.

After a scaffolding episode (Step 4), the fifth step, according to the tutoring frame, is for the tutor to gauge the student's understanding by asking questions, such as “Do you understand that now?.” Note that a comprehension gauging question asks the student to evaluate his/her own understanding; it is not a form of diagnostic evaluation on the part of the tutor to see how much the student understands (Chi, Siler & Jeong, in preparation).

The tutoring frame suggests that tutors basically dominate the dialogue, dictate the agenda, craft the appropriate next question, select the next example and problem to be solved, pose the analogy and counterexample, give the feedback, and so forth. It is thus natural to credit tutoring effectiveness to the tutors' pedagogical skills. The typical hypothesis posed about tutoring effectiveness focuses on the tutor, by asking the question “What do tutors do

that is so effective?" (Merrill et al., 1992). By and large, research programs that study human tutoring take this tutor-centered approach in tacitly assuming that tutoring effectiveness derives from the tutors' skill. That is, it is *what the tutors say and when they say it* that causes tutoring to be more effective than unaided studying or classroom instruction. The conception is that tutors are agile at tailoring and microadapting their instruction to the students' understanding. Thus, the methodological approach of such research programs is to focus almost exclusively on identifying the repertoire of tactics or moves available to tutors (Merrill et al., 1992; Evens, Spitkovsky, Boye, Michael & Rovick, 1993; Putnam, 1987), such as giving explanations, giving feedback, and scaffolding. Basically, these studies attempt to determine how tutors decide and choose among these different tactics (McArthur, Stasz & Zmuidzinis, 1990), how and whether they shift from one tactic to another tactic (VanLehn, Siler, Murray & Baggett, in press), how they generate explanations and feedback (Merrill et al., 1995), how they monitor students' understanding (Sleeman, Kelly, Martinak, Ward & Moore, 1989), how they motivate the students (Lepper, Woolverton, Mumme & Gurtner, 1991), and what variety of hints they use (Hume, Michael, Rovick & Evens, 1993). The focus on the tutors, to the total exclusion of the role of the students, is epitomized by the occasional use of simulated students (Putnam, 1987). This tutor-centered approach is further fueled by the need for developers of intelligent tutoring systems (ITS) to know which tutoring heuristics and tactics to implement. Thus, the goal of a tutor-centered approach has been to understand which pedagogical tactics human tutors use in a specific circumstance (e.g., which responses from students trigger a particular tactic), sometimes overlooking the need to relate tutor tactics to learning outcomes.

This kind of research approach does take into account the students' moves to some extent. The most frequent candidate for investigation among students' moves is when they made an error. The research question typically focuses on considering which tactics a tutor would undertake when the student either makes an error or makes a correct move. For example, Merrill, Reiser, Ranney & Trafton (1995) studied natural human tutoring in terms of understanding tutor's moves as a consequence of students' correct or incorrect actions while solving Lisp problems. They found that 66% of students' correct problem solving actions were followed by a tutor's confirmatory feedback (such as "Right, uh huh."). This can be taken to mean that confirmatory feedback is a useful tactic, presumably because it guides the students to continue the same line of correct reasoning. Similarly, using a computer Lisp tutor, Anderson, Corbett, Koedinger & Pelletier (1995) manipulated the kind of feedback students received. In general, students' post-test scores were superior when they received feedback on errors than when they did not receive feedback. Moreover, when the feedback was given immediately after an error, it reduced the amount of time students needed to reach criterion by half (but it did not necessarily improve their post-test scores as compared to the delayed feedback group). Such results again suggest that tutors' moves, such as giving feedback, can be responsible for students' learning, especially for a procedural skill, when examined in the context of one kind of student moves (student errors).

In general, even though the research approach does consider the students' moves as a condition to define the tutors' moves, the emphasis of these studies is to depict the advantage of tutoring as a result of the tutors' pedagogical moves. Basically, our interpretation of a

tutor-centered pedagogical hypothesis (T-hypothesis) is that it makes two general claims (one about the behavior of the tutors and the other one about its effects on learning):

- (1) That tutors dominate, control and craft the tutoring sessions; and
- (2) That it is the tutors' moves (as reflected in *what* they say and *when* they say it) that are responsible for learning.

The findings, as reviewed above, do support these claims, in showing that tutors do give responsive (confirmatory as well as negative) and on-going feedback, in the sense of guiding students and keeping them on track in acquiring the problem solving procedures (Anderson, Corbett, Koedinger & Pelletier, 1995; Graesser, Person & Magliano, 1995). These feedback comments do result in improved learning, at least in a procedural domain. However, such results do not rule out the possibility that other factors, in natural human tutoring, can also be responsible for students' enhanced learning.

The second claim of the T-hypothesis tacitly assumes that there is a systematicity and rationality in the way tutors select and choose a specific tactic in a specific situation (to be reflected in *what* they say and *when* they say it). This means that one should, in principle, be able to determine under what condition (i.e., *when*) a specific tactic or move is adopted; whether one can identify an optimal kind of move (i.e., *what*) for a specific situation; and whether it then results in learning. However, attempts at detailed analyses of this kind have failed to uncover any sort of systematicity in either the kind of moves human tutors undertake in similar situations, when they are delivered, and whether or not learning ensued. VanLehn et al., (in press) examined two "expert" physics tutors ("expert" in the sense that they consider themselves to be professional tutors) to see how they corrected student errors. They examined rules (such as "If a taut string is attached to an object, there is a tension force on the object exerted by the string") that were learned by the students in the course of tutoring, as assessed by gains from the pre- to the post-test. The findings suggest several conclusions. First, there was no systematicity in how the tutors "taught" the students' rules that they needed to solve the problems. For example, when students made an error, sometimes the tutors stated a general form of the rule, and other times they stated an overly specific version of the rule. In fact, for one of the rules that students missed, the tutors used as many as eight varieties of tactics. Thus, there did not appear to be an optimum choice for how a specific content rule ought to be taught (that is, there is no systematicity in *what* the tutors said). Second, the tutors' moves were not consistent in response to the same kind of student actions. For example, when a student made an error, sometimes the tutors explained why the student's error was wrong, sometimes they did not, suggesting that there is no consistency in *when* a specific kind of tactic is undertaken. Third, there did not appear to be any relationship between the use of a specific tactic and learning. For example, the tutors' explanations were not associated with students' learning gains, whether or not the explanations were given in the context of students making or not making errors. Overall, students did not appear to learn on the basis of systematic application of tutoring moves, as reflected in what tutors said and when they said it. Instead, students were more likely to learn a rule if they *made an error* and *recognized that they made an error*.

The preceding VanLehn et al. (in press) analyses examined the correct rules that the students needed to acquire in order to solve the problems successfully. Using a microgenetic

approach (i.e., examining a single student's 2,592 lines of problem solving protocols of the same dataset collected by VanLehn et al., in press), Chi (1996) reanalyzed the protocols from a different perspective. Instead of analyzing the correct rules that students needed to acquire in order to solve the physics problems, she analyzed the correction of 7 misconceptions. A misconception can be confusing mass with weight; or it can be confusing the direction of motion with the direction of acceleration; or it can be not realizing that an unknown force can be one of the forces embedded in the total force, without necessarily having to know an explicit equation containing that force as a variable. These misconceptions were not necessarily displayed in any written equations or utterances, nor were they necessarily elicited by the tutor. They can be captured in coding either by the absence of what students ought to say, or by alternative actions that they took. Learning was assessed by changes in the student's conception in the course of solving these three problems. Of these 7 misconceptions, only two were learned, each from a different tutoring move: hinting (a type of scaffolding) and answering student's question. Thus, consistent with the VanLehn et al. (in press) finding, the learning, although meager, resulted from two different tutoring tactics.

Why were the other 5 misconceptions *not* learned? It seems equally important to find out what tutor moves prevented learning. Three reasons can be postulated from the small sample of data analyzed (Chi, 1996). First, there were many lost opportunities, that is, occasions when the students exhibited confusion and the tutors ignored them. Out of 41 tutorial actions coded for one of the problems, there were 6 occasions when the student manifested confusion. In 5 of these 6 cases, the tutor simply ignored the confusion. Second, instead of addressing their confusion, the tutor was interested primarily in following his own plan, in this case teaching the student how to draw all the forces on the diagram and how to sum all the forces, even though these two procedures had nothing to do with the student's misconceptions and confusion. Third, the tutor often preferred to give long-winded didactic explanations that did not address the student's misconceptions.

In sum, in these detailed and small-scaled microgenetic analyses of tutoring moves and student learning, it has been difficult to capture any systematicity in the tactics used in any given situation, nor has it been possible to pinpoint which tactics are the most effective for learning. Moreover, many of the tutor's moves (such as ignoring student's confusion, following their own plans, and giving long-winded explanations) seem less than optimal. Finally, although it is often difficult to interpret instructional studies when they fail, efforts to teach tutors to use specific strategies have not been completely successful either (Clark, Snow & Shavelson, 1976; Swanson, 1992). With these kinds of results, it seems unwarranted to hypothesize that tutoring effectiveness arises solely from the tutor's skillful use of tutoring tactics.

### *1.2.2. Can tutoring effectiveness arise from the students' active generation? The student-centered constructive hypothesis*

Tutoring effectiveness has been examined largely from the perspective of the tutor, as summarized by the set of studies cited in the preceding section. The skepticism raised in the preceding section about the T-hypothesis suggests that we might also want to consider tutoring effectiveness from the perspective of the student. With a few exceptions (Fox, 1991; Graesser, Person & Magliano, 1995), very little attention has been given to the role the



students take in a tutoring context. However, it is possible that tutoring is effective because it affords opportunities for students to be more generative, that is to engage more actively in *constructive* learning, whereas in traditional classrooms in which instruction consists of the transmission of canonical knowledge by the teacher, the learner can only be a passive receiver of knowledge (Lave, 1988). Although it is possible to be generative when one is studying or learning alone, as in self-explaining (Chi, Bassok, Lewis, Reimann & Glaser, 1989), tutoring may be effective precisely because its interactive nature affords greater opportunities for students to engage in more constructive activities, as compared to a classroom.

Although the notion of “construction” has been discussed in the literature from several perspectives (e.g., Brinton, Jarvis & Harris, 1984; Cobb, 1994; von Glasersfeld, 1984, 1989; Wittrock, 1990), our notion of constructive learning is similar to the current emphasis on fostering more active and self-directed learning on the part of the students. We are thus merely adopting the prevailing notion of the active learner as one that constructs an understanding by interpreting the new, to-be-learned material in the context of prior knowledge. This can be accomplished by making inferences, elaborating the material by adding details, and integrating materials. More complex or deeper forms of construction might be forming hypotheses and justifications (Chan, Burtis, Scardamalia & Bereiter, 1992; Klahr & Dunbar, 1988); forming analogies (Chi et al., 1994); reflecting, summarizing, and predicting (Palinscar & Brown, 1984); justifying, criticizing, and exploring (Collins et al., 1989); or revising one’s existing knowledge (Chi, 2000; Brown & Palinscar, 1989). Such a learner thereby achieves deeper and more profound understanding than the passive learner, who typically avoids such “worrying knowledge” (Brown & Palinscar, 1989, p. 395).

Being constructive can be manifested externally by observable behavioral activities such as spontaneously self-explaining (Chi et al., 1989), asking questions, drawing, taking notes, summarizing, explaining to the tutors, answering the tutor’s questions, and so forth. In contrast, a less externally constructive learner would be one who passively assimilates the information, or someone who simply acknowledges the given instruction without giving more elaborative and deep responses. Being externally constructive of course does not guarantee learning, but there is a correlation between the frequency of external construction and learning. For example, the more students self-explained with content-relevant inferences, the more students learned (Chi et al., 1994). Conversely, a student could be constructive covertly, but since it would be difficult for us to assess such occurrences, we rely on overt manifestations of constructive activities.

The student-centered constructive hypothesis (S-hypothesis) is subtly but importantly different from the T-hypothesis. The T-hypothesis basically claims that tutoring is effective because of specific actions undertaken by the tutors, resulting in a research goal of identifying which moves are important for tutors to undertake and when they should be undertaken. The S-hypothesis suggests that a number of tutor moves may be effective because they promote constructive and effortful responses from the student. This means that it is not a specific tutor’s moves per se that is important (such as whether a tutor states a general rule or an overly specific rule when an error is made), but rather, whether the tutors’ moves encourage the students to construct substantive responses. Encouragement presumably can be accomplished by a number of different tutoring moves.

The assumption of the S-hypothesis is consistent with the detailed and microgenetic analyses that failed to find any systematic relation between specific tutoring moves and learning. (The only exception is feedback, as described earlier, which does seem to correlate with learning a procedural skill such as problem solving. However, feedback, especially in the context of a problem solving ITS, may be beneficial precisely because it often forces the student to respond constructively.) Most importantly, the S-hypothesis is consistent with the second and third general findings of tutoring: that the tutors themselves learn the subject matter (because by tutoring, they were being more constructive), and that tutoring is effective even when the tutors have no special tutoring skills. This latter finding is consistent with the interpretation that the effectiveness of tutoring derives from their elicitation of construction (such as from scaffolding), which requires tutors to have domain knowledge but not any special tutoring skills.

Indirect evidence in support of the constructive hypothesis can be garnered by reinterpreting several findings in the literature that were reviewed above. That is, there is substantial evidence indicating that a great deal of the tutors' tactics can be reframed as prompting or encouraging students to construct knowledge, either through the use of content-free prompts or scaffolding prompts. For example, McArthur et al. (1990) studied tutoring from a tutor-centered approach, in that their goal was to identify which tactics expert tutors used while tutoring high school students on algebra problems that the students had failed. They identified 44 different tactics that tutors used and the methods by which they chose among these tactics. However, these 44 tactics can be collapsed into 3 broad categories, such as: content-free prompting (24%), giving feedback (47%), and some kind of scaffolded prompting (29%). Thus, 53% of the tutors' tactics were forms of prompts that could elicit constructive responses from the students.

The S-hypothesis is consistent with other evidence that portrays the tutors as more relaxed, rather than being driven to keep the students on track in acquiring a particular procedure. Instead, this alternative perspective conceives of the tutor as having the role of guiding the students in a more open-ended way, not necessarily with direct corrective feedback. Rather, they suggest that expert tutors virtually never provide answers to problems. Instead, they *prompt* students to generate their own corrections with statements such as "So, you think it's 126?," *request explanations* ("Now tell me how you got that 6"), or *ask leading questions* ("Now which column is the one's column?"). Lepper et al. (1991) considered these tactics as serving the purpose of motivating the students and evoking curiosity. Nonetheless, even though these tactics are couched in the context of motivation, they do support the idea that tutors often prompt the students to elicit construction from them. Similarly, Evens et al.'s (1993) data show that 44% of the negative responses given by the tutors in face-to-face sessions can be considered general suggestions, rather than direct correction. In the same vein, Fox (1991) noticed that human tutors seldom give direct negative feedback immediately; instead, they pause and ask questions designed to elicit the students to self-correct their own errors and line of reasoning. Thus, these portrayals of more indirect tutoring are consistent with the hypothesis that these tactics give students more opportunities to construct knowledge.

The S-hypothesis emphasizes that tutoring is effective, not necessarily because tutors select and execute a specific move in a precise and skillful manner (in the sense that the right



move is given at the right time, in response to students' comments), but because these moves encourage the students to respond in ways that might cause learning. This hypothesis embodies two specific testable claims (one about the students' behavior and one about students' learning):

- (1) Students have greater opportunities to be externally constructive in tutoring than in a traditional classroom;
- (2) The students' constructive responses should correlate with learning.

There is some direct and indirect evidence to support parts of these claims, but no direct evidence to support both of these claims of the entire S-hypothesis, especially in a tutoring context. There is direct evidence in support of the first claim if we take question asking to be a constructive activity. For example, Graesser and Person (1994) found that the incidence of students asking deep questions increased significantly in tutoring sessions as compared to a classroom. On average, a student asks 8 deep questions per hour in tutoring, as compared to only 0.11 questions per hour in the classroom (including both deep and shallow ones). Although 8 questions per hour is still low, and moreover, students in a classroom context may be formulating questions covertly to themselves, this evidence nevertheless supports the claim that students are being more constructive in tutoring than in a classroom context.

There is also evidence to support a corollary of the second claim, that students who actively construct knowledge learn more than students who do not (Chi et al., 1989). Moreover, if students were elicited to construct knowledge, then they learn more than a control group. In Chi et al. (1994), one group of students was encouraged to be more constructive by elicitations from the experimenter. Students were individually prompted by the experimenter to self-explain while reading each sentence from a passage about the human circulatory system. To self-explain at the behavioral level is merely to elaborate to themselves about whether they understand the sentence and what it might mean. Such elaborations were coded as self-explanations if they contained inferences that went beyond the text sentences. The results showed that students who were prompted to self-explain learned with greater understanding (were more able to answer deeper questions) than a control group of students who were permitted to read the passage twice but not prompted to self-explain. Many other studies have since replicated this basic result, sometimes referred to as the self-explanation effect (Renkl, 1997; also see Chi, 2000, for a review). The prompts the experimenter used were content-free prompts, such as "Did you understand that?," or "What else does it mean to you?." This suggests that being more constructive (by giving self-explanations) yielded greater learning as compared to a control group of students who presumably were less constructive, since the experimenter did not overtly elicit construction from them, but they were permitted to read the same text twice. Thus the results of this study supports part of the second claim of the S-hypothesis, namely, that elicited construction led to greater learning gains than less construction, even though no pedagogical skills were required to employ content-free prompts.

In sum, even though there is evidence to partially support both of the claims of the S-hypothesis, no tutoring study directly tested the student-centered constructive hypothesis.

*1.2.3. Does tutoring effectiveness arise from the joint effort of both the tutors and the students? The interactive coordination hypothesis*

The previous two hypotheses viewed the effectiveness of tutoring from either the perspective of the tutors or the perspective of the students, although each perspective did acknowledge the role of the other (the students and the tutors, respectively). That is, the T-hypothesis examines the kind of moves tutors make in response to specific student moves (such as errors), and the S-hypothesis might consider the effect of students' constructive responses on learning when tutors elicited these constructive responses. However, it might be worthwhile to explore more broadly (beyond cases when students make errors) and more directly the contribution of the tutor-tutee interactions in tutoring effectiveness.

The importance of interaction has often been promoted in the context of the situativity theory (Greeno, 1998; Suchman, 1987) or a social constructivist context (Newman, Griffin & Cole, 1989). One interpretation of the benefit of interaction is that it presumably enables the participants to jointly construct (or co-construct) an understanding or a shared meaning that neither partner initially understands (Barwise & Perry, 1983; Brown, Collins & Duguid, 1989; Perret-Clermont, Perret & Bell, 1991; Rochelle, 1992). For example, Schliemann and Nunes (1990) have suggested that primitive fisherman can develop an understanding of proportions and devise calculation procedure through collaborative everyday practices. However, in tutoring, the dialogues between the tutor and the tutee are not symmetrical, so that the interaction is less likely to be a co-construction of a shared understanding, and more likely to be one in which the tutor tries to convey information about X in a way that the student can understand it, and the student displays or signals whether or not s/he does understand X; and moreover, the tutor may assess whether or not the student understands X as well as understands what the tutor has said. Despite its asymmetry, the processes in tutoring nevertheless involve joint coordination of interacting activities. However, no studies have directly examined how much interaction occurs in tutoring and what its implications are for learning.

The contribution of interaction toward learning can be shown without specifying precisely what is the nature of an interaction. Two sets of empirical research have demonstrated that interaction fosters learning. One line of research that focuses on interaction is collaborative learning and peer problem solving (Brown & Palinscar, 1989; Collins et al., 1989). This line of empirical work reveals two general phenomena: first, that problem solving in pairs is more productive than working independently (Azmitia & Perlmutter, 1989; Skon, Johnson & Johnson, 1981; Tudge & Rogoff, 1989); second, the more successful dyads are characterized by a higher degree of communication among the partners (Cooper, Ayers-Lopez & Marquis, 1982; Foreman & Cazden, 1985). Although both of these findings are consistent with the interpretation that each of the partners are being more constructive (the S-hypothesis), nevertheless, they suggest that the interaction may somehow contribute toward learning. The Coleman study (Coleman, 1998), to-be-described next, shows even more clearly that both the opportunity for construction and interaction can contribute to learning.

Coleman (1998) tested the effectiveness of scaffolded explanation prompts in a group discussion format. She asked three-person groups to arrive at a consensually agreed upon solution to photosynthesis problems after the groups had two weeks of instruction about the topic. In the course of the discussion, she gave each member of the three person group

interchangeable roles, such as the “prompter,” the “explainer” and the “reader/writer.” The prompter had to choose an appropriate cue from a provided set of prompting cards. The reader/writer read the cue and wrote down a response, and the explainer generated explanations for the prompting cue and provided justifications. The roles rotated among the 3-member teams. The cues consisted of 8 provided prompts, ranging from a very general one such as “Can you explain this in your own words?” to justification prompt such as “Explain why you believe that your answer is correct or wrong,” to integration prompt such as “What did we learn in class about this particular topic?” Examining the individual students’ post-test scores, the intervention groups retained and comprehended more about photosynthesis than the control groups which did not participate in the intervention. When the groups’ consensual solution was rated on four conceptual levels from intuitive and incorrect to scientifically correct, the consensual explanations of the intervention groups tended to be more conceptually advanced than the control groups. Hence, the first result, namely that individuals in the intervention group retained and comprehended more about photosynthesis, can be attributed to a constructive hypothesis interpretation (in that each group member did have more opportunity to explain when they took the role of the “explainer” than the control group students). However, the second finding, of more advanced consensual solutions, would support the interaction hypothesis. Thus, it seems that interaction, even in this constrained way, did enhance learning above and beyond the advantage of being constructive.

A second line of research that points to the importance of interaction comes from studies of dialogues (McKendree, Stenning, Mayers, Lee & Cox, 1998). Learning seems to occur not only through participation in a dialogue, but some learning also occurs through observing or overhearing others participating in it (Clark, 1996; Cox, McKendree, Tobin, Lee & Mayers, in press; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). For example, Schober and Clark (1989) examined students’ performance on a task (such as constructing Tangram figures) when a “director” (like a tutor) tells the students how to construct each Tangram figure. One group of students participate in conversations with the “director” while learning to construct these figures, while another group merely overhears the dialogues of the participants. The results showed that participants performed better than the overhearers, suggesting that interaction facilitated performance.

The interaction hypothesis (I-hypothesis) essentially states that the effectiveness of tutoring arises from the advantage of interaction. However, what is an interactive dialogue and how can it be differentiated operationally from a constructive response? Although one could say that all dialogues are inherently joint acts (Clark, 1996), and therefore interactive, in the context of understanding learning, it may be important to differentiate between communicative acts that are “interactive” versus “non-interactive,” in terms of the content. From the perspective of the tutors, we define and code “interactive” acts to be those comments that elicit a response from the students, such as asking the students either content or comprehension gauging questions, or scaffolding them. “Non-interactive” comments would be those in which the tutors give lengthy explanations without giving the students a turn to respond, or tutors giving feedback followed immediately by an explanation, without interjecting a comprehension-gauging questions.

From the perspective of the students, the interactive nature of students’ responses need, moreover, to be differentiated from students’ constructive responses, in order to tease apart

Table 1  
Types of students’ responses to tutors’ comments

	Constructive	Non-constructive
Interactive	Elicited responses:	Acknowledgement responses
	Shallow & deep follow-ups	Continuers such as “uh-huh” and “okay”
	Responding to questions	*Gestures such as head nods
	Responding to scaffoldings	*Eye Gaze
Non-interactive	Self-initiated responses:	No responses:
	Initiate new topic	Ignore
	Self-explain	
	Read sentences	

\* Gestures and eye gaze would fall into this interactive but non-constructive category, but they were not coded in the tutoring protocols.

the independent contributions of being constructive from being interactive. Thus, we need to differentiate not only “interactive” versus “non-interactive” responses, but we also need to differentiate between responses that make a substantive contribution with respect to the content versus those that do not. Table 1 shows hypothetical types of responses that students can make to tutors’ comments, and which type of responses may be considered constructive and/or interactive. First, elicited responses would be a type of constructive and interactive ones in which the tutors make a move, such as asking a question, and the students respond to such elicitation. For example, if a tutor sets a goal or provides a procedural hint (for a problem solving dialogue), and a student responds by taking a step toward reaching that goal or elaborating upon that hint, then such a dialogue would be both constructive and interactive. Basically, a scaffolding type of dialogue would be both constructive and interactive, in the sense that guidance is provided by the tutors, and the students’ respond to that guidance with comments that follow-up on what the tutors said. Second, a constructive but non-interactive type of response would be self-initiated ones whereby the students are not following up to the tutors’ comments. Instead, the students *ignore* what the tutors say (in terms of the content), and initiate a response on their own (such as self-explaining), initiate a new topic of discussion, or simply read. Ignoring can be manifested also in the discourse style such as lengthy and extended monologues, without giving the tutors a turn to interject. Alternatively, these self-initiated comments can also be generated in the absence of tutors’ comments. Third, an interactive but non-constructive type of response might be mere acknowledgment type of comments, such as continuers “o.k.” or “uh-huh”; it may also include head nods, gestures, and eye gazes. That is, a student can be responsive in the sense of attentive eye gaze and appropriate turn taking (with comments such as “o.k.”) and yet be non-interactive in terms of the content of what the tutor says. Finally, a non-constructive and non-interactive type of responses would be ones whereby the students either ignore the tutors’ comments or simply give no responses. These four types of responses are referred to in Table 1 as elicited responses, self-initiated responses, acknowledgment responses, and no

responses. Thus, the I-hypothesis embodies two testable claims (again a behavioral one and a learning one):

- (1) One simple claim is that both tutors and students are maximally interactive. That is, tutors' comments are largely communicative acts that elicit responses from students; and students' communicative acts are responsive to tutors' elicitation.
- (2) A learning claim is that students' interactive construction (i.e., elicited responses) should foster learning more so than students' non-interactive (i.e., self-initiated) construction. Such a finding, if true, pinpoints the essence of the advantage of tutoring over and above merely being constructive (such as learning by self-explaining).

The following studies explore whether there is evidence in support of any or all of the hypotheses, and whether some hypotheses are more viable than others are. In addition, the studies have several methodological goals. First, they attempt to be more comprehensive in their analyses, in the sense of covering the entire corpus of the tutoring dialogues or protocols, rather than focusing on specific subsamples, such as either focusing on a specific tutoring tactic (such as giving feedback during errors), or focusing on the acquisition of specific rules or the removal of misconceptions. Second, the studies also used multiple tutors, whereas the majority of the studies in the literature used very few. Third, the tutoring activities were related to the learning outcomes, rather than being merely descriptive. Finally, these studies employed a conceptual domain rather than a procedural one (problem solving), for the interest of seeing if some of the existing findings are replicated here.

## **2. A study of naturalistic tutoring**

This study examined protocols collected in naturalistic tutoring sessions in which the students were learning a conceptual domain (the human circulatory system) from tutors who were told to tutor in whatever way they want.

### *2.1. Method*

#### *2.1.1. Participants*

Tutors were 11 college students (6 female nursing students, 1 male nursing student, and 4 male biology majors at the University of Pittsburgh) who had expertise in the domain of the human circulatory system, although they were unskilled tutors. That is, they had not had tutoring experiences or training per se. The students were 11 eighth-graders who were recruited from various schools in the Pittsburgh area (both private and public schools). They were recruited either by having the teachers pass out letters for parents to contact us and give consent, or through a community-tutoring center. The students were compensated for their time at a rate of \$6.00 per hour.

The students also had to provide their scores on the California Achievement Test, which will be used here as a gross approximation to their reading abilities.



### 2.1.2. Materials

A passage about the human circulatory system, consisting of 86 sentences (see Appendix A), was taken from a well-written and popular biology text (Towle, 1989). Each sentence of the passage was printed on a separate sheet of paper. The pretest consisted of two tasks: defining 21 terms of the circulatory system and drawing (and explaining while they drew) the blood-path on an outline of the human body. The post-test consisted of the same two tasks as in the pretest, plus answering 70 questions. (The 21 terms and the post-test questions are shown in Appendix B.)

The 70 questions were classified into 4 categories. The 17 questions in Category 1 could be answered by directly by referring to a specific sentence(s) given in the text. Category 1 questions can be considered to be text-explicit. For example, a student can answer question #6 *Which vein carries oxygenated blood?*, by reading the information contained in Sentence 76. The 14 questions in Category 2 required inferences and integration across text sentences. For example, to answer question #4, *Why is there an artery that carries deoxygenated blood?*, a student would have to integrate the information in sentence #69 (that pulmonary circulation is the movement of blood from the heart to the lungs and back to the heart), with sentence #70 (that blood from the right ventricle flows into the pulmonary artery and then to the lungs), and then reason that since blood is going there to get oxygen, it must be deoxygenated prior to getting to the lungs, whereas other arteries leave from the left ventricle, and blood in these arteries is oxygenated (see sentences #23, #24 and #30). Thus, students needed to integrate and reason from information explicitly stated in several sentences in the text. Category 2 questions can be considered to be text-implicit ones. Thus, although we treat Categories 1 and 2 questions as *shallow* ones in the analyses, it is obvious that Category 2 questions are far from trivial.

To answer Category 3 questions requires not only that students infer important new knowledge from the sentences, such as inferring the function of the atrium if only its structure and behavior are described, but some of them can be answered correctly only if the students have the correct double loop model (the last model shown in Fig. 1). These will be referred to as model implicit questions. For example, one cannot answer question #4 *Why is your right ventricle less muscular than your left ventricle?* by integrating information from several sentences in the text. First of all, in contrast to question #4 from Category 2, in which the information in the question is explicitly stated in the text (i.e., sentence #71 explicitly refers to the fact that the pulmonary artery is the only artery that carries deoxygenated blood, although it did not say why), there are no sentences that refer to the content of Category 3 type of questions. Thus, no sentences explicitly discuss the muscularity of the right and the left ventricle. In order to answer question #4 from Category 3 then, the student must have more or less the correct mental model, in which the blood from one side of the heart (the right side) goes to the lungs and the left side goes to the entire body. Thus, since the lungs are closer in proximity to the heart, the contraction would not have to be as great (sentence #19) as the left side (sentence #24). These questions are extremely difficult, in that they require deep understanding of the materials in the sense of having the correct mental model, not only of blood circulation, but also of diffusion.

Finally, the 11 questions in Category 4 were devised to test students' ability to apply their knowledge to everyday health-related issues, and these will be called health questions. These

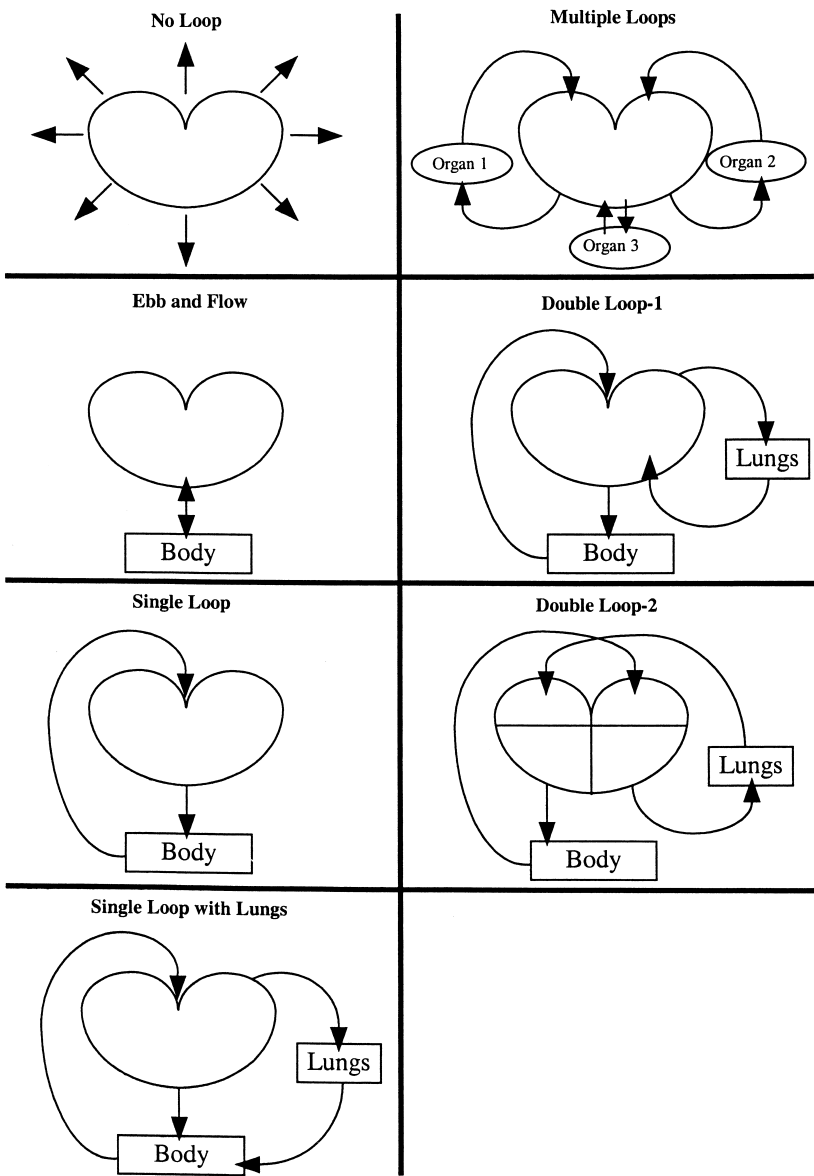


Fig. 1. Seven mental models.

questions often appear to have little relation to information presented in the text. For example, Question #9 asks why people feel dizzy when they suddenly stand up after sitting on the ground for awhile. To answer this question, they must reason that when they stand up, blood has farther to travel against gravity, so they might get less oxygen to the head. Thus, Categories 1–3 questions are text-based, in the sense that the information needed to answer them can be encoded, integrated, and inferred from the sentences provided in the text passage, whereas Category 4 health questions address issues not mentioned in the text

passage at all, and thus can be considered to be transfer questions rather than text-based. Categories 1–3 are analogous to the distinction provided by Pearson and Johnson (1978) between text explicit (Category 1), text implicit (Category 2), and script implicit (Category 3). In the analyses, Categories 1 and 2 will be combined and referred to as shallow questions, and Categories 3 and 4 will be combined and referred to as deep questions.

### *2.1.3. Design and procedure*

All the tutors were told to tutor their students “naturally,” in whatever manner they liked. They were encouraged to limit the tutoring to no more than two hours and no less than one-and-a-half hours. Thus, no format or constraints were provided, except that they were asked to include motivating comments. The tutors were told however, that the students had to learn the materials well enough to answer a set of post-tutoring questions. They were shown samples of questions that the students had to be able to answer.

The entire study consisted of three sessions given in sequence — the pretest session, the tutoring session, and the post-test session. In the pretest session, students were asked to define 21 terms and to draw and explain the path of the blood through the circulatory system on a paper that has an outline of the human body. After the pretest session, students participated in the tutoring session. Before the tutoring session began, the tutors and the tutees were both required to read the entire passage once. Each sentence of the passage was typed on a separate piece of paper, assembled in a three-ring binder. During the tutoring sessions, the tutors read each sentence aloud to the student about half of the time. They then engaged in tutorial dialogue with the students.<sup>2</sup> The post-test session was administered about one week later in order to assess long-term learning. In this session, students again defined the 21 terms and drew and explained the blood-path. In addition, they answered the 70 Categories 1–4 questions. These questions were not given in the pretest in case it influenced students’ thinking while learning in the tutoring sessions. The entire study took about 4 hr.<sup>3</sup>

In summary, the dependent measures of this study relevant to this paper consisted of the students’ pretest and post-test definitions and bloodpath drawings and explanations of the bloodpath drawing, and the post-test answers to Categories 1–4 questions. In addition, we also audiotaped the student-tutors dialogues during the tutoring sessions.

## *2.2. Analyses and results*

### *2.2.1. Overall learning, comparing pre-test and post-test*

There are two ways to assess learning: more shallowly at the piecemeal (or knowledge pieces) level, and more deeply (at the mental model) level. A knowledge piece corresponds to a proposition such as “*The aorta is an artery.*” In this sense, knowledge pieces are required to answer the shallower (Category 1 and 2) questions. A mental model analysis, on the other hand, captures the global coherent model that a student might have used to represent the circulatory system. The correctness of a student’s mental model determines how accurately they can answer the deeper (Categories 3 & 4) questions and draw the circulatory system.

*2.2.1a. Knowledge pieces analyses.* Because the definition test was administered during both the pretest and the post-test, the pre- to post-test gains were calculated. The definition

protocols gathered while each student defined each of the 21 terms during the pre- and post-tests were transcribed and scored. On the basis of the answers found in the pre- and post-test protocols, a scoring template of 96 knowledge pieces was created from the propositions that were mentioned by at least one student. Each knowledge piece corresponds to a proposition, such as “*the septum divides vertically.*” This method of constructing the scoring template means that there were many propositions that were mentioned in the text but not included in the template. This way of constructing the template has two implications. First, the template included only those propositions that the students could understand, in case the text was obscure. Second, this method also insures that the average scores of individual students were well below a ceiling level in their learning of the knowledge pieces. The mean proportion correct (0.13 at the pretest and 0.46 at the post-test) shows a gain of 32% and was significant at the 0.001 level,  $t(10)=5.63$ ,  $p < 0.001$ .

*2.2.1b. Mental model analyses.* Learning can also be assessed by depicting the amount of improvement in the students’ mental models from their drawings and explanations collected in the pretest and post-test. Mental models can be captured as described in Chi et al. (1994), in which we determined from their drawings and explanations which mental model students used to represent the human circulatory system. Fig. 1 shows 7 different mental models, the first 6 being flawed to different degrees. The 7 mental models can be ordered in their correctness from the most naive No Loop model to the most correct Double Loop-2 model. The Double Loop-1 model is essentially correct with respect to the overall circulatory path, but not with respect to the bloodpath internal to the heart. This ordering is based on our subjective analyses of each model’s conceptual approximation to the correct model.

There are two ways to look at the improvements in the students’ mental models. One simple way is to count how many of the 11 students had the correct Double Loop-2 model in the pretest (there were none), and how many had the Double Loop-2 model at the post-test (there were 8 out of 11 students). Another way to look at the improvement is to count the number of increments in the mental model progression from the least accurate to the most accurate. That is, if a student represented the circulatory system as an Ebb and Flow mental model at the pretest, and then as a Double Loop-1 model at the post-test, this would constitute an improvement of 4. Calculating improvement this way also allows for decrement. Overall, the students gained 22 increments, or 2.0 increments per student, in the mental model progression.

Thus, from these two analyses of the pretest and post-test gains (on gains in knowledge pieces from the definition test, and on improvements in mental models from the drawing and explaining test), students clearly learned from tutoring. If we assume that knowledge pieces measure shallow knowledge and mental models measure deep knowledge, then students have obviously learned both kinds of knowledge. Notice that we cannot say whether they learned more or less than a control group, because the goal of this study was not to compare tutoring effectiveness with some alternative method of instruction (such as classroom). Many studies have amply documented that superiority already. The purpose of reporting gains here is to make sure that the students were learning, and that the tutoring was effective.

### 2.2.2. *Question-answering at the post-test*

Category 1–4 questions were administered to students at the post-test only. Students' answers to questions in each category were transcribed and scored. Each answer was scored as either a 0 (if the students did not answer the question at all or had an incomplete answer), a 0.5 (if the answer was partially correct), and a 1.0 if the answer was completely correct. The scoring key was taken and adapted from the one used in Chi et al. (1994). Students answered 64% of Category 1, 45% of Category 2, 35% of Category 3, and 31% of Category 4 questions correctly (also shown later in Table 4). Notice that there is a trend in the systematic decrement in the proportion of correct answers as students progressed from the easiest explicit text-based questions (Category 1) to the hardest transfer-type health questions (Category 4). (Also see Chi et al., 1994 for the same trend even though the questions in these two studies are not identical.) This suggests that there is some validity in the way these questions were constructed and the sequence of difficulty they indexed.

### 2.2.3. *The tutoring dialogue protocols and general coding procedure*

It would be helpful to give a sense of what the tutoring sessions are like, since this is not a problem-solving task, as is commonly used in the tutoring literature. Prior to the beginning of the tutoring session, both tutors and students were required to read the entire passage on their own. When a tutoring session began, it typically proceeded in the following way. The tutor started by either explaining the sentence to the student or reading it aloud; then the tutoring dialogue followed. At some point, when the dialogue ended, they went on to the next sentence. The tutor usually determined when they should go on to the next sentence. (Appendix C gives samples of the tutoring protocols from Tutor #1 and Tutor #10, at sentences 7 and 85.)

The tutoring sessions lasted on average 1.80 hr, which generated 20 hr of transcribed protocols, filling 11 binders. There were three major comprehensive codings of the tutoring protocols (comprehensive in the sense that the entire corpus of the tutoring dialogues was examined). Each comprehensive coding used a different grain size and will be described below. Aside from these three comprehensive codings, there were numerous additional codings of subsamples of the protocols, the pretests and the post-tests, as well as quantitative counts of various sorts. These additional codings will be described and embedded throughout the paper.

Because of the sheer magnitude of these coding efforts, several of the codings were carried out by one individual (usually by one of the coauthors). A single coder was used under the following conditions. First, we relied on a single coder for codings that appeared to be straightforward and in which the coding categories could be well defined. Second, a single coder was used for codings that we have amply experienced with, based on previous codings (such as those done in Chi et al., 1994), since the text passage used in this study is almost identical to the one reported in Chi et al. (1994), with the exception that a dozen sentences about diffusion were added. Third, for coding categories that are ill-defined and new to us, we relied on two coders.

**2.2.3a. *Statement coding.*** The tutoring protocols collected during the tutoring sessions were first segmented into statements. One statement is equivalent to a single idea, sometimes



sentence length, but not necessarily. The double slashed lines below indicate the boundaries of statement segment. The statements below were uttered by two tutors, each enclosed in quotes:

“So, whenever you breathe out, so you are expelling the carbon dioxide.//

And whenever you are inhaling, you are bringing in the oxygen.//”

“So white, yeah, your white blood cells, if your count is real high it means that there is infection.//

So they have invaded whatever part of the body need assistance defending bacteria or whatever.//”

Each substantive statement (i.e., excluding acknowledgment type of comments such as “o.k.”) that the tutors made was classified by a single coder into 8 categories, and they are:

1. Giving explanations
2. Giving direct (either positive or negative) feedback, followed by a short corrective explanation if the feedback is negative, such as “No, when it went through the atrium, it went through the tricuspid valve” (when the student was about to say that blood flows through the “bi one”).
3. Reading text sentences aloud
4. Making self-monitoring comments (in which the tutors commented about their own instruction such as “I don’t know if this will help you,” or commented about the tutors’ own understanding of the materials, such as “I don’t know why they put [that line of text] in there, it just kind of confuses me.”)
5. Answering questions that the students asked
6. Asking content questions (such as “Which is the upper and which is the lower chamber of the heart?”)
7. Scaffolding with generic and content prompts
8. Asking comprehension-gauging questions (such as “Is this starting to stick?”)

We consider the first four categories to be non-interactive moves (as defined earlier, see Table 1). For example, giving explanations is considered to be a non-interactive move because for the majority of the times, multiple explanation statements were given before students were given a chance to take a turn with a response. And if the multiple statements were followed by a question, then basically the tutors’ explanation statements were followed by a tutor question-asking statement, thereby the explanation statements themselves did not explicitly elicit a response from the student. The same is true for feedback moves. After giving a yes/no plus a short corrective feedback, the tutors may launch into an extended explanation without giving students a *turn* to respond. Because the coding is at the level of a statement and not a turn, it is possible for the tutors to make many statements without getting any responses from the students. In contrast, the last four categories were moves that explicitly responded to or elicited a response from the students, thus we consider them to be interactive. However, the division of these 8 categories into interactive and non-interactive ones is a gross characterizations of interaction. More specific characterization is undertaken in the next (interaction) analysis.

The first six categories are self-explanatory, but clarifications and examples are needed for the last two categories, scaffolding and comprehension gauging questions. Earlier, we had gleaned from the human tutoring and developmental literature several kinds of (nonfeedback type of) guiding activities that can be considered to be scaffolding. From such literature, we have culled 15 types of moves that can be operationalized for coding (as Category 7 scaffolding statements) including tactics such as: (1) pumping for “what else” (Graesser et al., 1995), which is comparable to what we have labeled as generic or content-free prompting (Chi, de Leeuw, Chiu, & LaVancher, 1994), (2) hinting (“So, it’s kind of leaving out the lungs there?”), (3) fill-in-the-blank kinds of requests (such as “OK, and it goes from the atrium to the \_\_\_\_\_ pause”), (4) highlighting critical features (Stone, 1993), (5) decomposing the task, (6) executing parts of the skill (Rogoff, 1990), (7) providing physical props or cue cards (Scardamalia, Bereiter, & Steibachm, 1984), (8) describing the problem so as to orient the students to the important features (McArthur, Stasz, & Zmuidzinas, 1990), (9) comparing the current problem with a previously solved problem (McArthur et al., 1990), (10) maintaining goal orientation or reminding the student of some aspect of the task (Woods et al., 1976), (11) completing the students’ reasoning step or “splicing in” (or jumping in and providing) the correct answer when the student commits an error, without acknowledging that an error has been made (Graesser et al., 1995), (12) initiating the beginning of a reasoning step or a task (such as “Name every part on there.”), (13) asking a leading question (such as “And where do you think it goes?”), (14) redirecting the student, or (15) providing an example. In general, a scaffolding move is a kind of *guided prompting* that pushes the student a little further along the same line of thinking, rather than telling the student some new information, giving direct feedback on a student’s response, or raising a new question or a new issue that is unrelated to the student’s reasoning (as an initiating question would in Step 1 of the tutoring frame). The important point to note is that scaffolding involves *cooperative execution* or *coordination* by the tutor and the student (or the adult and child) in a way that allows the student to take an increasingly larger burden in performing the skill. Thus, scaffolding is clearly a joint activity in which the tutor plays some role, and often sets up a role for the tutee to play.

A comprehension gauging question (Category 8) is one in which the tutors ask the students whether they understand, such as “Did you understand that?,” “Are you unclear on anything?,” “Does that kind of make sense?,” or “Do you have any questions so far?.” Thus, it is an attempt to prod the students to reflect upon their own comprehension. Comprehension gauging questions cannot serve the function of assessment, since students’ responses would be a kind of self-evaluation, which may be accurate or inaccurate.

As indicated by the 5-step tutoring frame, since tutoring dialogues were predominantly controlled by the tutors, there were far fewer kinds of student statements. The students’ (excluding “o.k.” type of) statements were classified into six categories and they are:

9. Reading sentences aloud
10. Spontaneously giving unprompted self-explanations (although they were not coded to see whether or not they contained inferences beyond the text sentences, as was done in Chi et al., 1994)
11. Asking questions

12. Answering questions
13. Responding to the tutors' scaffolding prompts
14. Reflecting

Only the last category needs some explication. Reflection consists of comprehension monitoring statements that were made either in response to tutors' comprehension gauging questions (CGQs), or some other tutoring moves. For example, a tutor's CGQ might be "Did you feel like you understood it at all?" and the student can respond by saying "Um, . . . I understood a lot of it, but not all of it." In addition, reflection also includes self-initiated comprehension monitoring statements. Students sometimes generate these monitoring statements in the midst of a tutor's explanation, for example, without the tutor having asked a CGQ. Both of these (elicited and self-initiated) kinds of comprehension-monitoring responses are considered to be a kind of reflecting statement.

We consider the first three categories of comments (Categories 9, 10, 11) to be more-or-less self-initiated and non-interactive, whereas Categories 12 and 13 were explicitly elicited by the tutors. We say "more-or-less" because we could not determine with certainty what exactly initiated the students' actions of the first three categories.

*2.2.3b. Interaction coding.* Although the statement coding grossly characterized the interactive nature versus non-interactive nature of the dialogues, this second coding attended specifically to the kind of interactive responses students provided. We examined adjacent pairs of comments in which the first contribution was one of four tutor moves: the tutors' explanation, feedback, scaffolding, and asking CGQs. We then coded the kind of contribution the student made in the next turn. Thus, in the interaction coding, even nonsubstantive comments made by the students such as "uh huh" or "O.K." were coded, whereas they were not counted in the statement coding. Following and extending Clark and Schaeffer's (1989) categories, the students' responses were coded either as (1) a *continuer* with comments such as "uh huh," or "okay," (2) a shallow *follow-up* (Stenstrom, 1984) which could be an elaborative paraphrase, (3) a deep *follow-up*, which is an elaborative inference that extends what the tutor said, (4) a reflecting response, such as "I don't understand.," or (5) a response that initiates a new topic (in the sense that it did not answer or expand on the same topic that the tutor initiated). Two examples of shallow and deep follow-ups are shown below, along with the text sentences:

Text sentence #1: *Human life depends on the distribution of oxygen, hormones, and nutrients to cells in all parts of the body and on the removal of carbon dioxide and other wastes.*

Tutor: "Basically, what we are talking about is the circulatory system is an exchange of materials."

Student: (shallow follow-up) "You take out the waste and you put in the nutrients."

Text sentence #16: *Each of the valves consists of flaps of tissue that open as blood is pumped out of the ventricles.*

Tutor: "OK. So opening and closing, what would that do?"

- Student: (shallow follow-up) “It would allow the blood to enter like from the atrium without it falling straight through.”
- Text sentence #16: *Each of the valves consists of flaps of tissue that open as blood is pumped out of the ventricles.*
- Tutor: “blood actually flows out through there.”
- Student: (deep follow-up) “This contracts like a balloon and forces this venous blood up here.”
- Text sentence #43: *At first the molecules of sugar are more concentrated in and around the sugar cube and less concentrated in the water farther from the cube.*
- Tutor: “This cube of sugar is disintegrating, breaking apart, expanding into all spaces . . .”
- Student: (deep follow-up) “Until, until equilibrium is accomplished.”

In this analysis, shallow follow-up, deep follow-up, reflecting, and initiating a new topic, are considered to be constructive responses (as defined in Table 1), whereas a continuer is a kind of non-constructive response. Note that both shallow and deep follow-ups are comparable to the kind of self-explanation inferences that were coded in Chi et al. (1994).

This interaction coding differs from the statement coding in several ways. First, it is coding at a slightly larger grain size, at the level of the turn. Take tutor explanations as an example. Tutor explanations in the statement analyses were considered to be a non-interactive move, in that many statements were made within the same tutor turn. In this interaction coding, each unit of analysis is the turn. Thus, once the tutor took a turn in explaining, we coded the kind of responses that the student gave. For example, in Appendix C, for tutor #1 in Sentence 7, two of the three tutor explanations were rather lengthy, so they contain multiple statements. But in this analysis, each lengthy explanation is considered a unit, and the kind of responses students gave was coded when the students took the next turn, after the tutors completed a unit of explanation.

Moreover, in the statement coding, we were capturing the *format* of students’ (and tutors’) comments, for example, in terms of whether the students’ statement was a response to a scaffolding prompt or a CGQ. For the interaction coding, this analysis characterized the *content* of the students’ response, irrespective of whether a given response was an answer, a scaffolding response, or a question. (See Chi, 1997, for a distinction between format and content codings.) For example, if a tutor scaffolded, a student could respond with a continuer, a shallow or deep follow-up, a reflection, or the student could initiate a new topic, based on a content analysis. These 5 types of responses would have been all lumped into one category (as students’ responses to scaffolding) in the statement coding. Finally, the interaction coding was undertaken by 2 coders. The two coders initially conferred upon a category label of each response students gave to tutors’ explanations, feedback, scaffoldings, or comprehension gauging questions, until mutual agreement was reached. Then both coders independently coded the protocol of one student. Both coders resolved any discrepancies between them. The remaining protocols were then coded by one of the coders.

**2.2.3c. Episode coding.** A third coding was undertaken at even a larger grain size, focusing explicitly on the nature of scaffolding dialogues. Coding individual statements or adjacent

pairs of contributions might have overlooked the essence of scaffolding, whose benefit is considered to arise in the tutoring context from the sequence of 5–10 turns interchange, according to Step 4 of the tutoring frame. We operationalize a scaffolding episode to be an extended dialogue that is initiated by the tutors' scaffolding moves, and continue on the same concept or topic for several turns, presumably until the tutors are satisfied that the students have mastered the concept. Thus, an extended scaffolding episode may be ones that encourage deeper understanding. Appendix D provides two samples of scaffolding episodes.

To identify these episodes, the entire corpus of tutoring dialogues was recoded again. This time it was done by two coders, with an inter-rater agreement of 75%. The disagreement centered primarily on the boundary of an episode and not on whether or not an identified segment of dialogue is a scaffolding episode. These boundary discrepancies were resolved by discussion.

#### *2.2.4. Is there evidence to support any of the three hypotheses?*

For each of the hypotheses, we had postulated a couple of testable claims, one about the nature of the behavior and the other about learning. For the T-hypothesis, we should see that the tutors took control of the tutoring process and that some tutoring tactics do correlate with learning. It is important to find evidence in support of the T-hypothesis, consistent with the literature, in order to show that our study and results are not flawed in some ways. In order to find evidence in support of the S-hypothesis, we should see results that show that students are being more constructive in a tutoring context (as opposed to a classroom context), and that some of the students' moves also correlate with learning. Finally, for the I-hypothesis to have any validity, we should see evidence that tutoring is dominated by interactive dialogues, and perhaps that interactive (or elicited) responses correlate with learning more so than non-interactive (or self-initiated) responses. We begin by describing the results of all the analyses of the coded data and specify which hypotheses they support. Then the evidence in support of each hypothesis is summarized.

*2.2.4a. Quantity and type of statements from the statement coding.* On the basis of the statement coding, an average total of 827 statements were made by the tutors and the students in the tutoring protocols. Of these, each student said on average 206 statements, which are about 25% of the tutor-student dialogues. Although we cannot directly compare the amount of student statements to those taking place in a traditional classroom, a simple calculation shows that if each student in a classroom contributes 206 statements in the span of 1.8 hr of tutoring (or about 1.9 statements per minute), this would mean that in a class size of 30 students, a total of 3,420 student statements would be articulated per hour! This is clearly not the scenario of a typical classroom. Not only is it the case that the teacher in a traditional classroom dominates the dialogues, but moreover, 3,240 statements cannot possibly be articulated by the students in an hour-long class. Thus, it's easy to see that students are having a greater opportunity to be more constructive in a tutoring situation, compared to a hypothetical classroom scenario.

Although we do not have classroom data as a control, a better comparison we can make in order to claim that students are being more constructive in a tutoring context is to look at the number of questions students ask. Each student asked, on average, 14 questions in the



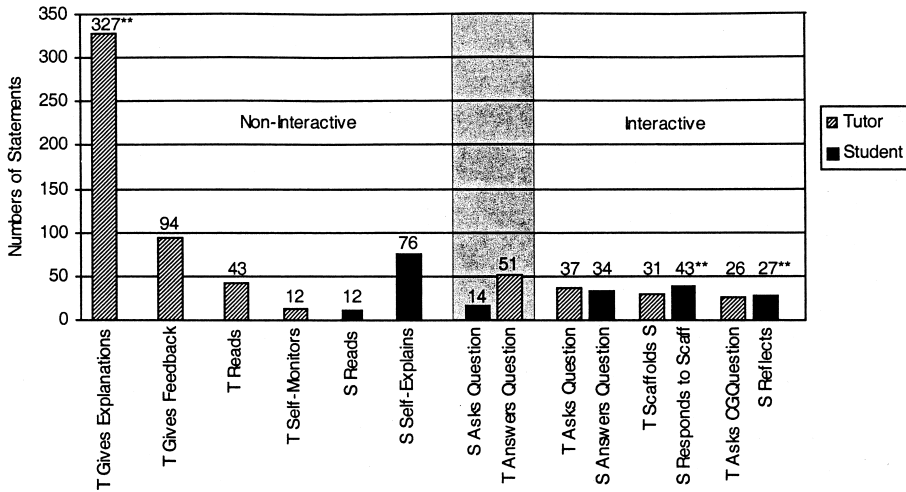


Fig. 2. Frequency of each type of tutor and student statements (Study 1).

span of 1.8 hr of tutoring, which amounts to around 7.8 questions per hour. (Notice that this number is almost identical to Graesser & Person’s 1994 finding of students asking 8 questions per hour in their tutoring sessions.) In contrast, according to Graesser and Person (1994), each student in a classroom asks on average 0.11 questions per hour, and an entire classroom asks only 3.0 questions per hour. Thus, using the frequency of question asking as an index of constructive opportunities, there is no doubt that students are much more constructive in a tutoring context. Thus, using the number of overall statements students made and the number of questions they asked, the data clearly support the first (behavioral) claim of the S-hypothesis.

Fig. 2 shows the distribution of the eight types of tutor statements (hatched bars) and the six types of student statements (solid bars). The tutors’ non-interactive moves as well as the students’ self-initiated responses, are graphed individually on the left as single bars. Those categories of tutor moves that clearly elicited a response from the students (i.e., were interactive) are graphed on the right as pairs. The only category to which the students initiated an action and the tutors responded was students asking and tutors answering questions. This pair is graphed in the middle with shaded background.

The tutors made a significantly greater number of statements overall than the students (621 vs. 206,  $t(20)=5.06$ ,  $p < .001$ ). Moreover, they also controlled the tutoring situation, which can be captured in the following ways. First, the tutors frequently spoke first. For the majority of the times the tutors spoke first (54% of the times), whereas the students spoke first only 12% of the times. (The percentage does not add up to 100% because we ignored reading statements). Second, the tutors took many more turns at speaking than students did (256 turns on average per tutor vs. 224 turns for the students). Taking more turns mean that the tutors usually initiated and terminated the conversational dialogue about each sentence. Thus, tutors appeared to dominate and control the tutoring sessions, supporting the first (behavioral) claim of the T-hypothesis.

Excluding answering students’ questions, tutors disproportionately (476 out of 621 state-

ments) uttered moves that probably did not elicit responses from the students. That is, 77% of the tutors' moves were non-interactive ones (those shown on the left side of Fig. 2). Most noticeably, the proportion of tutors' statements was overwhelmingly dominated by explanations (327 statements or 53% of the tutor statements were explanations). Only 15% of the tutors' statement (or 94 statements, summing across tutors asking questions, scaffolding, and asking CGQs) attempted to elicit responses from the students (104 of them). This means that the tutors were not maximally interactive by any means. However, over half of the students' comments were interactive in being either responsive to tutors' elicitations (104 comments) or asking tutors questions (14), versus 89 non-interactive ones, in which the students read on their own, or self-explained.

Notice that the two numbers, 94 tutor elicitation comments and 104 student responses to those elicitations, are not identical because this analysis is a coding of the number of statements articulated, so that one could respond with more than one statement, and moreover, these numbers reflect averages across subjects. Nevertheless, the more-or-less comparable number of statements (94 for tutors and 104 for students) suggests that students never ignored the tutors' elicitations. Thus, even though eliciting moves did not comprise a large proportion of the tutors' moves, whenever tutors did invite students to respond, the students were 100% responsive. Overall, the first claim of the I-hypothesis is only partially supported, in that only the students were somewhat interactive; the tutors were barely interactive.

In sum, this analysis of the quantity and types of statements provides support for the behavioral claims for two of the three hypotheses. The data support the T-hypothesis in showing that the tutors did, consistent with those findings in the literature, dominate the dialogues in that they spoke much more frequently than the students, and they controlled the situation by speaking first and taking more turns. Nevertheless, the data also support the S-hypothesis, in that the students asked more questions (as compared to classroom data reported in the literature). Finally, there is some support for the I-hypothesis in that over half of the students' comments were responses to tutors' elicitation; whereas only a small proportion (15%) of the tutors comments were interactive ones.

*2.2.4b. Correlation of statements with student learning.* In order to know whether any tutor and student moves correlated with learning, a step-wise regression analysis was undertaken, using the frequency of tutor and student statements as the variables entered. Students' pretest and reading scores were always merged and entered as the first variable. The analysis then selected the next variable that has the largest correlation with learning, as measured by the post-test scores of either the mean of Categories 1 and 2 (shallow learning) or the mean of Categories 3 and 4 (deep learning). The regression analysis terminated when the significance level is higher than 0.10. Aside from the contribution of pretest and reading ability scores toward learning, Table 2 shows the additional amount of variance each variable accounted for ( $R^2$ ) and whether the change in  $R^2$  is significant or not ( $F$  and  $p$ ) for shallow (the upper half of Table 2) and deep (the lower half) learning.

The first unsurprising result to note from this step-wise regression analysis is that, consistent with the literature (e.g., Adams, Bell, & Perfetti, 1995), prior knowledge (as assessed in the pretest) and reading ability accounted for more than half of the variance in learning, and more so for deep (0.705) than shallow (0.558) knowledge. This means that if

Table 2  
Step-wise regression analyses for shallow and deep learning

	R <sup>2</sup>	R <sup>2</sup> change	F	p
Shallow learning				
Pre-test and reading	.558	.558	5.04	.038
S responses to scaffolding	.837	.280	12.02	.010
T explanations	.952	.115	14.47	.009
Deep learning				
Pre-test and reading	.705	.705	9.55	.008
S reflective comments	.965	.261	52.68	.000

any other variables accounted for learning, it does so above and beyond the effect of prior knowledge and reading ability. The second result to note is that two of the three variables that correlated with learning were student variables: students' responses to scaffolding and students' reflective comments, suggesting that students' responses seemed more important overall than the tutors' moves. The third important finding is that tutors' explanations and the students' responses to scaffolding correlated only with shallow learning. Finally and most interestingly, the only variable that correlated with deep learning was students' reflection. (Those variables or categories that correlated with learning are also indicated in Fig. 2 by asterisks.)

The second claim of the interaction hypothesis states that if interaction is important, then students' elicited responses should be more important for learning than students' self-initiated responses. This can be inferred from the regression analysis of Table 2, showing that students' elicited (or scaffolded) responses correlated with learning, whereas students' self-initiated responses (such as self-explanations, asking questions) did not (therefore they are not shown in Table 2). We even carried out a regression analysis in which we manually entered students' self-explanations as the second variable (after pretest and reading ability), and found that it did not account for extra variance. Thus, this result supports the second claim of the I- hypothesis, that interaction, in the form of giving scaffolded responses, accounts for some learning above and beyond self-initiated (therefore non-interactive) construction, such as by self-explaining. (Notice that the self-explanation statements coded here should more accurately be called unsolicited or unprompted explanations. They are not self-explanation inferences, as used in Chi et al. (1994), so that the lack of correlation between these unprompted explanations and learning does not contradict our earlier result. Nevertheless, these explanations were self-initiated and not elicited.)

Overall, the main findings of the step-wise regression analysis indicates that the learning claims of all three hypotheses are supported. First, one kind of tutor moves (giving explanations) did correlate with shallow learning, thus supporting the second learning claim of the T-hypothesis. Second, two kinds of student moves (responding to scaffolding prompts and reflection) correlated with either shallow or deep learning, thus supporting the second learning claim of the S-hypothesis. In fact, given that two of the students' moves, as compared to only one of the tutors' moves, correlated with learning, this suggests that students' construction was more important than tutors' explanations. Finally, the results also support the learning claim of the I-hypothesis, in that the students' elicited construction (such

Table 3  
Step-wise regression analyses for shallow and deep learning

	Shallow (Cat. 1 & 2)				Deep (Cat. 3 & 4)			
	R <sup>2</sup>	R <sup>2</sup> change	F	p	R <sup>2</sup>	R <sup>2</sup> change	F	p
Model 1								
Pre-test and reading	.558	.558	5.041	.038	.705	.705	9.551	.008
S respond to scaffold	.863	.305	15.541	.006	.764	.060	1.774	
T scaffold	.863	.001	.030		.769	.005	.121	
Model 2								
Pretest and reading	.558	.558	5.041	.038	.705	.705	9.551	.008
T scaff	.758	.201	5.815	.047	.727	.022	.564	
S respond to scaff	.863	.105	4.607	.075	.769	.042	1.099	
Model 3								
Pretest and reading	.558	.558	5.041	.038	.705	.705	9.551	.008
S reflect	.777	.220	6.914	.034	.965	.261	52.680	.000
T ask CGQ	.788	.011	.299		.966	.000	.033	
Model 4								
Pretest and reading	.558	.558	5.041	.038	.705	.705	9.551	.008
T ask CGQ	.965	.176	4.625	.069	.858	.154	7.586	.028
S reflect	.966	.054	1.540		.966	.107	18.686	.005

as responding to scaffoldings) seemed more effective for learning than students' self-initiated construction (such as self-explaining, since self-explaining did not correlate with learning).

*2.2.4c. Manually entered step-wise regression analyses.* It is difficult to give total credit to the S-hypothesis merely by pointing out that students' responses, in the form of responding to scaffoldings and reflection, correlated with learning because one could argue that unless the tutors provided the scaffoldings, students would not have had the opportunity to respond to the scaffoldings. However, the variables entered in the regression analyses were not the frequencies of tutor and student moves per se, but the number of statements of each kind of moves. Therefore, there is independence in that a student can respond with a few or a multiple number of statements in response to tutors' scaffoldings, which in turn can also be brief or extensive. Thus, in order to assess whether students' responses to scaffoldings and reflection were more important than tutors' elicitation (scaffoldings and asking comprehension gauging questions), since there were significant correlations with learning for the former but not the latter, it might be informative to directly pit one against the other. To answer such a question, we manually entered the variables of the number of statements of tutors' scaffolding, tutors' CGQs, students' responses to scaffoldings, and students' reflection (which included both responses to tutors' CGQs and self-initiated ones). Table 3 shows the R<sup>2</sup>, the change in R<sup>2</sup>, the F values for the change, and the p values. For Models 1 and 2, the tutors' scaffolding and the students' responses to scaffolding were reversed (after entering pretest and reading scores as the first variable). Models 3 and 4 entered and reversed the variables of students' reflection and tutors asking CGQs.

Two important findings emerge. First, if we manually entered the tutors' moves first (after pretest and reading scores) as in Models 2 and 4, then the tutors' moves do correlate with learning: scaffoldings with shallow learning (Model 2); and asking CGQs with shallow

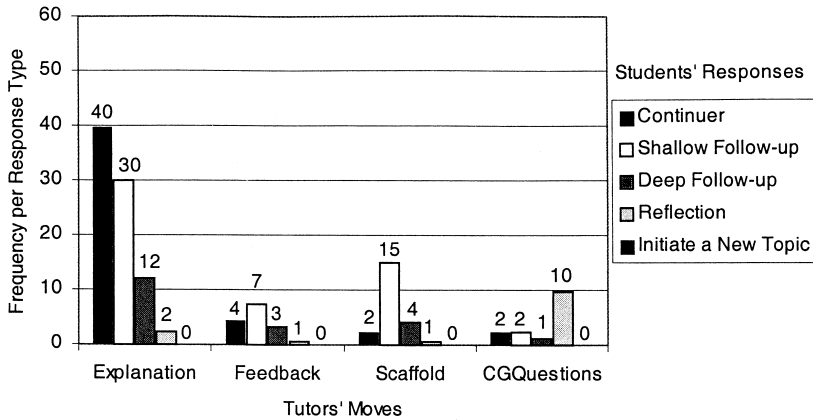


Fig. 3. Frequency of students' responses as a function of tutors' moves in the tutoring condition (Study 1).

(marginally) and deep learning (Model 4). However, the students' responses continued to significantly account for additional variance. On the other hand, when the students' moves were entered first (Models 1 & 3), then the tutors' moves were no longer significant. For example, students' responses to scaffolding seemed to account for more variance in shallow learning than tutors' scaffoldings (for shallow questions, the  $R^2$  change for Model 1 is 0.001, not significant, whereas the  $R^2$  change for Model 2 is 0.105, marginally significant at the 0.07 level). Second, looking at models 3 and 4, students' reflection again seemed to account for more variance than tutors asking CGQs, for deep learning. This is evident if student's reflection is entered before tutors asking CGQs (Model 3), then tutors asking CGQs no longer accounted for any additional variance. When students' reflection was entered after tutors asking comprehension-gauging questions (Model 4), students' reflection continued to account for additional variance in deep learning. Thus, in looking at the independent contribution of the frequency of tutors' and students' statements (for scaffolding and reflection), it seems that in both cases, the students' responses to scaffolding and reflection were more important than the tutors' scaffolding and asking the students to reflect. Both of these results show unambiguously that the students' active responses seemed to play a more important role in enhancing learning than the tutors making these eliciting moves, again suggesting that the S-hypothesis might be more viable than the T-hypothesis.

**2.2.4d. Analysis of interaction.** The statement coding was not sensitive to the actual linking of specific individual tutor elicitations and individual student responses. In other words, the 31 scaffoldings and the 43 scaffolding responses (as shown in Fig. 2) were averages across subjects. The interaction coding, on the other hand, captures the exact type of response a student gave to each kind of tutor elicitation, at the next turn. On the basis of the interaction coding, Fig. 3 plots the absolute frequency of each type of student responses (continuer, shallow follow-up, deep follow-up, reflection, initiating a new topic) as a function of specific tutor moves (giving explanations, giving feedback, scaffolding, and asking comprehension gauging questions). We consider a continuer as a kind of acknowledgment that is a non-constructive response (see Table 1 again), whereas the other four responses (shallow and



deep follow-up, reflection and initiate a new topic) are *constructive* in that they are substantial contributions.

Several interesting findings emerge that illuminate the correlational results of the statement coding. First, tutor scaffoldings (the 3rd set of columns in Fig. 3) definitely elicited shallower follow-up than deep follow-up (15 vs. 4), as one would expect from the operational definition of scaffolding. This result also explains why students' responses to scaffolding correlated only with shallow learning (Table 2). Second, likewise, among students' four types of constructive responses to tutor explanations, shallow follow-ups were the most dominant, therefore it again makes sense that tutors' explanations only correlated with shallow learning (Table 2). Third, both proportionately and absolutely, explanation is the only category of tutor moves that elicited more continuers than any other kinds of constructive responses. This suggests that relatively, explanations, as a category of tutor moves, were not as effective at eliciting constructive responses from students as scaffolding, CGQ and feedback, in decreasing order. This may explain why students learn minimally from explanations (given the large number of them), both in this study, as well as other studies cited earlier (VanLehn et al., in press; Webb, 1989). Fourth, feedback elicited the fewest total number of constructive responses (a total of 11). This might explain why tutor feedback did not correlate with learning at all, because the feedback in this study did not elicit much constructive responses from the students. This result again contrasts the importance of a S-hypothesis versus a T-hypothesis, in that it is not the tutors' moves per se that is important, but whether or not they elicited *constructive* responses from the students. Thus, overall, it seems that whether students learned at all, or whether they learned shallowly or deeply (as indicated by the correlation analyses of the statement coding), depended on the amount and kind of responses they gave, which in turn depended on the kind of moves tutors undertook to elicit these responses. Some tutor moves (such as scaffolding) elicited more shallow follow-ups than continuers. Other tutor moves (such as feedback) elicited very few constructive responses overall. Thus, there is clearly an interaction between the tutors' moves and the students' responses, supporting the I-hypothesis.

A non-interactive dialogue, as we defined earlier in Table 1, would be one in which the students ignored the tutors' comments, and either gave no responses to the tutors' comments or gave self-initiated responses. Fig. 3 shows that the dialogues, coded at the level of the turn and in response to these four specific tutor moves, were always interactive since the students never initiated a new topic in their comment. Thus, this analysis is consistent with the result of the prior statement coding analysis that also showed that students were maximally interactive in that they always responded to the tutors' eliciting comments. (That is, the number of tutor eliciting statements, 94, were comparable to the number of student's response statements to those elicitations, 104). An alternative way to think about this is that tutors do in fact lead the dialogue in choosing and initiating the topic of the queries, explanations, and scaffoldings.

Although the interaction analysis clearly showed a dependency between the students' moves and the tutors' moves, it also revealed a puzzling lack of systematicity. That is, even though there is a dominant type of responses for any given type of tutor moves (as shown in Fig. 3), such as acknowledging with a continuer to explanations or giving a shallow follow-up to scaffolding, there is nevertheless a distribution of different types of student

responses. This means that the interaction between the tutors and the students depended not only on the kind of moves the tutors made (i.e., the format of the moves, whether it's a scaffolding or a CGQ, for example), rather, the distribution suggests that some other factor, perhaps the content of the moves, must also play an important role. The lack of systematicity in the kind of responses students consistently provide is reminiscent of the lack of systematicity in the kind of moves tutors gave in response to a specific student move (VanLehn et al., in press). This suggests that the specific move chosen, by either the tutor or the student, depends on some aspect of the interaction that we have yet to capture, such as the content of the dialogues. Thus, a tutor-centered approach of identifying which tutoring tactics are useful for maximizing learning would be misinformed unless the analysis is pursued in conjunction with an understanding of why a particular student response is elicited in that context.

Overall, this interaction analysis not only provides insights into why certain moves by the tutors and the students correlated only with shallow (but not deep) learning, this analysis (Fig. 3) further contrasts the importance of students' constructive responses from the tutors' moves. That is, the amount of students' constructive responses seems to determine learning more so than the appropriateness of a particular tutor move. On the other hand, some tutor moves are definitely "better" than others at eliciting a constructive response are. (For example, explanations can be said to be *not* a good tutoring move because proportionately, they are more likely to elicit a more passive continuer-type of responses, whereas scaffolding is more likely to elicit constructive responses, such as shallow follow-ups.) However, no tutor move elicited consistently a particular type of student responses; the distribution of different types of student responses to a single tutor move suggests that the responses depend on some other aspect of the interaction (such as the content of the dialogues) that we have not explored. The interaction also appears to be asymmetrical, as one would expect, in that the students rarely (never, in fact) responded by initiating a new topic of discourse, in the context of these four tutor moves.

*2.2.4e. Analysis of scaffolding episodes.* Although the incidence of scaffolding by the tutors accounted for only 5% (31 out of 621 statement) of the tutors' activities, students' responses to it nevertheless correlated significantly with shallow (but not deep) learning. Why did responses to scaffolding not correlate with deep learning? Analysis of the statement coding showed that scaffolding responses correlated with shallow learning (Table 2) perhaps because students' follow-up responses to scaffoldings are mostly shallow, as shown in the interaction analysis of Fig. 3. However, is this because the tutors' scaffoldings were also shallow, or did the tutors scaffold deeply?

In coding, deciding whether a tutor's scaffolding prompt is considered shallow or deep can be interpreted more accurately in the context of a scaffolding episode, which we have defined to be an extended multiturns scaffolding dialogue on the same topic. From such episode coding, a total of 52 scaffolding episodes were identified across the entire corpus of protocols. For each of the 52 scaffolding episodes identified, we judged whether it was a shallow or a deep episode, based largely on the kind of scaffolding prompts the tutors gave. Appendix D provides examples of shallow and deep scaffolding episodes. Shallow episodes were defined to be those that contained primarily shallow prompts, such as prompts that

required (1) a single integration, such as relating the topic to one other definition; (2) remembering a word or a definition; (3) asking where something is in relation to another body part; (4) restating what the students meant in different words; (5) relating the topic back to something the student is already familiar with. The shallow episode example in Appendix D contains scaffoldings of type 3, for example. Deep scaffolding episodes contain scaffoldings that prompt for (1) multiple integration; (2) inferences such as “Why does the heart need to pump so hard?,” (3) implications such as “Why is that important?,” and (4) asking for how something works or the causes associated with a certain mechanism. The deep episode example in Appendix D contains scaffoldings of type 1, for example. The result showed that the majority (80.8%) of the scaffolding episodes were shallow; therefore it is not surprising that the majority of students’ follow-ups were shallow (in the interaction analysis, see Fig. 3).

Notice the difference between the episode coding and the interaction coding. In interaction coding, we coded *students’ responses* to scaffolding as either shallow or deep follow-up. In episode coding, we basically determined whether an episode is shallow or deep on the basis of the *tutors’ scaffoldings*. But both analyses show that scaffolding prompts and scaffolding responses tend to be more shallow than deep. By shallow, we mean probing that ask students to go a little further on what the student has just said, rather than asking students to integrate several comments that they had made, or to compare the current sentence with previous sentences. Students responded to these shallow prompts with shallow continuation in their reasoning. They almost never initiated a new direction or a new topic of inquiry, for example. Consequently, because scaffolding tends to prompt for and students tend to follow up with shallow reasoning, it follows that responses to scaffolding correlated only with shallow learning. This may be why there is such an appeal for a Socratic kind of tutoring, in which the tutors pose deep counterfactual questions, and challenge the students in a way that allows the students to discover their own misconceptions (Collins & Stevens, 1982). Unfortunately, there is ample evidence to suggest that everyday tutors do not undertake sophisticated Socratic type of reasoning (Chi, Siler, & Jeong, in preparation; Graesser et al. 1995).

#### 2.2.5. Summary of evidence in support of each hypothesis

Altogether, the preceding analyses gave evidence in support for the behavioral aspect of two of the three hypotheses, and the learning aspect of all three hypotheses. Here we summarize them briefly.

The evidence in support of the tutor-centered hypothesis is:

- (1) Tutors did seem to control and dominate the tutoring sessions, in the way they initiated the conversation by speaking first, in the way they took more turns, and in the overwhelmingly large number of statements they made.
- (2) One tutoring move, giving explanation, did correlate significantly with students’ learning of shallow knowledge.

The evidence in support of the student-centered hypothesis is:

- (1) Students did have a much greater opportunity to be actively constructive in the number of questions they asked, as compared to a conventional classroom.

- (2) Moreover, their constructive responses (giving scaffolded responses and reflection) did correlate with either shallow or deep learning (whereas the tutors' moves to elicit these responses, scaffoldings and asking CGQs, did not).

The evidence in support of the interactive hypothesis is:

- (1) On the basis of the statement coding analysis, the results suggest that tutors were not very interactive in that only 15% of their comments were targeted at eliciting student responses. On the other hand, over half of the students' comments were interactive, either in responding to tutors' elicitation or in asking questions to tutors, suggesting that students were more interactive than the tutors were. Similarly, results from the interaction analysis further confirmed that the students were always interactive, in that they never initiated a new topic in their comments in response to four types of tutor moves. Thus, overall, the behavioral aspect of the I-hypothesis is supported only for the students.
- (2) The elicited responses (from scaffolding) correlated with learning whereas the unsolicited comments (coded as self-explanations) did not, suggesting that elicited responses (from interaction) are more important for learning than self-initiated statements.

#### 2.2.6. Discussion

The evidence provided in this study clearly suggests that all three hypotheses may be responsible for explaining tutoring effectiveness. Although we cannot make any causal claims between the tutors' and the students' moves with learning (because the analyses were correlational), we can assess the relative contributions of each hypothesis. This was done throughout the analyses in several ways, such as by contrasting the contribution of tutor moves with student moves. For example, two kinds of student moves (responding to scaffolding and reflection) correlated with learning whereas only one kind of tutor moves (explanation) did. Moreover, the tutor moves (scaffolding and asking CGQs) that elicited these students' responses did not correlate with learning. And if we directly pit these hypotheses by manually entering these moves in specific order into step-wise regression analyses, then the students' moves accounted for additional variance even if they were entered after the tutors' moves to elicit these responses, whereas the reverse was not true. Finally, whether or not a type of tutor moves correlated with learning seemed to depend on the total amount of constructive responses the students generated.

If we examine the results across different coding analyses, additional subtle contrasts appear. For example, the fact that feedback did not correlate with learning (from the statement coding) may be explained by the fact that very few constructive responses (from the interaction coding) were elicited by these feedback, probably because for a conceptual domain, an appropriate response to feedback is not always obvious, in contrast to a procedural domain. These results suggest that the students' responses (S-hypothesis) seems to determine learning more so than the tutors' moves (T-hypothesis). Another subtle contrast is the finding that students responded predominantly shallowly (from the interaction coding) primarily because tutors scaffolded them shallowly (from the episode coding), also suggesting that the students' responses depended on the tutors' moves, so that it's the interaction that's important, not merely the kind of moves tutors or students undertook per se.

Thus, in sum, even if we cannot claim with certainty the importance of the alternative S- and I-hypotheses in accounting for the effectiveness of *learning* from tutoring, their relative importance, in comparison to the T-hypothesis, gives us confidence that these alternative hypotheses are viable explanations. In the next two sections, we explore further why tutors' explanations correlated only with shallow learning, and whether elicited reflection is beneficial to learning as compared to self-initiated reflection.

*2.2.6a. Learning from hearing tutors' explanations.* As one can see from Fig. 2, the prominent activity undertaken by the tutor is giving extensive explanations: over half of the tutors' talk consisted of giving explanations.<sup>4</sup> There are two intriguing questions. First, why did tutor explanations correlate with learning at all when other studies have shown that students do not seem to learn from either tutors' explanations (VanLehn et al., in press) or from listening to a partner's explanations in a collaborative context (Webb, 1989)? Second, why did tutor explanations correlate with shallow learning only?

An explanation of text materials in the context of this tutoring task can be construed to be "helpful" or "useful" for learning in six ways. First, an explanation can rephrase the text sentences well, perhaps by linking the content of the current sentence to prior knowledge introduced earlier in the text, as well as integrating it with information to-be-introduced in the upcoming sentences. Second, an explanation can enhance learning if it contained extraneous information that can fill gaps in the text sentences, or provide a more complete and coherent understanding. Third, an explanation can reinforce learning if it repeats the information in the text. Fourth, an explanation can promote learning if it is rephrased in more understandable everyday language (such as in a more conversational style using nontechnical jargon), rather than in the expository style of a written text. Fifth, tutor explanations can enhance learning if they elicited constructive responses from the students. Finally, from an instructional point of view, an explanation is considered a good one if it is given in response to students' direct queries or when students signal that they don't understand (Leinhardt, in press). Presumably, we would like "good instructional explanations" to be targeted at the students' confusion, lack of understanding, and misunderstanding.

We propose that the students learned from listening to the tutors' explanations for the first five characteristics listed in the preceding paragraph, and not because these explanations were "good instructional explanations" (i.e., being tailored and responsive to students' misunderstanding and confusion). We reject this sixth characteristic for several reasons. First, this question has been tentatively answered in the negative in the microgenetic analyses mentioned earlier (Chi, 1996; VanLehn et al., in press). In the Chi analysis, it was shown that the tutor ignored the student's confusion, and the tutor's explanation focused on pursuing explication of knowledge that the tutor thought the student needed to know. In general, the tutor seemed to be unresponsive to the student's exhibited confusion.

Second, we can show that the tutors could not have been responsive to the students' need for explanations since these explanations were didactic in style. To illustrate what the dialogues look like when tutors gave so many explanations, Appendix C shows samples of Tutor #1 and Tutor #10's tutoring protocols, at Sentence 7 and Sentence 85. These explanations were usually generated in the context of multiple statements all occurring on a single turn (referred to as a "unit" of explanation) and focused on the delivery of didactic content.



This means that the dialogues were not interactive. The students did not participate in the explanation, to either show that they understood or to interject queries. This didactic style is very similar to the knowledge display style common to traditional classroom instructional dialogues. A good proportion of instructional dialogues contain “presentation,” which refers to the “teachers’ uninterrupted explanation of new or recently learned materials” (Leinhardt & Greeno, 1986, p. 83). For example, Webb (1983) found that teachers took 46–61% of the discussion time of groups they led. Similarly, in small-group medical problem-based learning context, led either by experts or novices, the experts have a greater tendency to take over the discussion than the novices (Silver & Wilkerson, 1991). Thus, a didactic or knowledge-display style implies that these explanations were not intended to be responsive to the students’ confusion and queries, thus they may not enhance students’ learning.

Third, these explanations could not have been responsive to students’ need for explanations because they were frequently unsolicited by the students. As shown in the examples in Appendix C, tutors often jumped right in and explained before students even showed any signs of wanting or needing explanations. This eagerness can be verified by coding the frequency with which tutors initiated an explanation right after a sentence is presented (upon the turning of a page, since only one sentence is presented on each page). On average, for 35% of the sentences, tutors initiated an explanation “unit” (consisting of multiple explanation statements given in a single turn), instead of some alternative moves, such as self-monitoring, asking questions, scaffolding, asking CGQs, and so forth. Another way to show that these explanations were largely unsolicited is to note the disproportionate number of explanation statements given (327) in contrast to the number of questions students asked (14; see Fig. 2 again).

Fourth, in a separate paper, we have direct evidence to suggest that tutors do not have an accurate understanding of students’ misunderstanding (Chi, Siler & Jeong, in preparation; they only have some sense of what students do understand). Thus, these four reasons, together, reject the conjecture that students might have learned from tutors’ explanations because the explanations were adaptive and responsive to students’ lack of understanding and misunderstanding.

Instead, we propose that in this study, students learned not because the tutor explanations were responsive to students’ misunderstanding, but because the explanations could have been “useful” or “helpful” in the five other artifactual ways stated earlier: That is, the tutor explanations might have provided integration, provided new information not contained in the text, repeated the information in the text, restated the text sentences in everyday language, and elicited constructive responses from the students. To support the first conjecture, we randomly selected two tutors and coded the content of their explanations in terms of providing analogies, linking by anticipating upcoming information, linking to previous explanations and sentences, supplementing the text explanations with drawings, and so forth. For these two tutors, about 40% of their explanations either anticipated and linked forthcoming information or referred to prior explanations and sentences. Thus, these explanations do provide an integrative function.

To test the second conjecture, we coded each explanation statement from all 11 tutors (at the knowledge pieces level) to see whether the content knowledge was redundant (i.e., elaborations and repetitions) of information in the text passage or extraneous. For example,



the tutor might go into a lengthy discussion about high blood pressure and high cholesterol. The results show that 46% of the tutors' explanation statements provided extra information that is not contained in the text passage. This extra information often pertains to diseases and health issues, thus relevant to answering the far transfer Category 4 health questions.

Another way to think about this result is that almost half of the tutors' 327 explanations (or 150 explanation statements) contain new information, while the other half (or 177 explanation statements) was re-presentation of the same information. For a text passage length of 86 sentences, this means that the students also heard the same information two extra times, supporting the third conjecture.

The fourth conjecture is that the explanations were rephrased or couched in everyday language (i.e., using everyday lexicon and a conversational style) that is easier for students to understand. For example, after reading sentence #46, *When the concentration of the molecules of a substance is the same throughout a space, a state called equilibrium exists*, one of the tutors explains:

"Okay. So something being equal keeping you balanced, okay, the equal proportion of fluid on this side and this side, you're now balanced which keeps you centered and upright."

Basically, the tutor substituted everyday familiar lexicons such as "balanced," "equal proportion" for difficult scientific jargon such as "concentration" and "equilibrium." Or a tutor might explain a sentence in simpler conversational style, using shorter phrases. For example, after reading sentence #24 *Strong contractions of the muscles of the left ventricle force the blood through the semilunar valve into a large blood vessel and then throughout the body*, a tutor explains:

"So it goes out and spreads through the system."

Fifth, despite the fact that a disproportionately large number of tutor explanations elicited a non-constructive response (or 40 continuers, see Fig. 3), nevertheless, a slightly larger number of responses were constructive (a total of 44 shallow, deep follow-ups and reflection). Such constructive responses obviously could induce learning.

Thus, overall, it seems that students might have learned from hearing these explanations because the explanations provided some integration, contained extraneous relevant information, repeated the information, and were rephrased in everyday language. Moreover, the explanations did elicit constructive responses from the students. It appears not to be the case that the explanations were carefully crafted and tailored to the students' confusion or misunderstanding (i.e., they were not necessarily "good instructional explanations").

How is it that students seemed to have learned only shallow knowledge? There are two possible reasons. First, the preceding arguments proposing that tutors' explanations were not "good instructional explanations" might be one reason why students did not learn deeply. A second reason is the results uncovered from the interaction analyses. In Fig. 3, we had shown that tutor explanations do elicit constructive responses from the students. Unfortunately, the majority of them were shallow rather than deep follow-ups, thus resulting in shallow learning.

In sum, students' shallow learning correlated with tutor explanations probably not

because the explanations were well-crafted and responsive to students' misunderstanding and confusion, but rather, because the explanations (a) integrated the information, (b) contained new information not presented in the text, (c) repeated the old information two extra times, (d) presented the information in everyday language in a nonexpository style and moreover, (e) these explanations did elicit constructive responses from the students about half of the times. We fully recognize that the argument presented in this section treats learning as if it was caused by tutors' explanations when in fact the data were merely correlational. We are simply entertaining these ideas as plausible accounts to be explored in future studies.

*2.2.6b. Reflection.* The strongest and most surprising finding from the regression analyses is that students' reflection correlated with deep learning. These reflections occur both as a responses to tutors' comprehension gauging questions (as well as by other types of tutor moves, as shown in Fig. 3), and self-initiated. There is a well-established literature showing that more frequent, spontaneous, and self-initiated monitoring of one's own comprehension correlates with improved comprehension and learning (Brown & Palinscar, 1989; Collins et al., 1989; Palinscar & Brown, 1984). However, there is no evidence in the literature to suggest that elicited comprehension monitoring enhances learning. For example, although Graesser et al. (1995) noted that 35% of their tutors' questions were comprehension gauging ones such as "Are you following?" (corresponding to Step 5 of the tutoring frame), they did not address the impact of this tutorial move on learning.

Unfortunately, since our coding conflated elicited and spontaneously self-initiated reflection, we could not tell whether elicited comprehension-monitoring responses correlate with learning. This is important to know because if this were true, then one can easily implement comprehension gauging questions either in an intelligent tutoring system, or train teachers to ask such questions, such as "Are you following?," in order to enhance learning. Therefore, it is important for us to disambiguate our results in the context of the findings in the literature. Accordingly, we separated out those reflection statements that were spontaneously self-generated from those that were elicited (i.e., responses to tutors' comprehension gauging questions). We correlated these variables with Categories 1–4 questions (controlling for pretest and reading scores). It turns out that the results strongly and unambiguously confirm the findings in the literature, in that only the self-initiated monitoring statements correlated with learning, whereas the elicited responses to comprehension gauging questions did not. In fact, the self-initiated monitoring comments correlated significantly (at the 0.01 level or better) with Categories 2–4 (with correlations of 0.81, 0.79, 0.78 respectively) but not with Category 1 questions, suggesting that comprehension monitoring is more important for the more difficult materials. This result also offers more support for the S-hypothesis rather than the T-hypothesis in that it is the students' constructive responses, in terms of self-initiated comprehension monitoring, that seem to correlate with learning more so than tutor-elicited comprehension monitoring responses. Moreover, this result is also consistent with VanLehn et al.'s (in press) results, which showed that students were more likely to learn a rule if they made an error and *recognized that they made an error*. Similarly, Chi (2000) also suggested that students were more likely to repair their own flawed mental models if they recognized

that their mental models conflicted with the text model. Such recognition could have arisen from comprehension monitoring. Unfortunately, it remains unclear how we can encourage students to reflect in a way that is beneficial to learning.

### **3. A study of interactive tutoring**

There are three important results from Study 1 that need to be addressed further. The first is that there is definitely support for an interactive hypothesis in explaining the effectiveness of tutoring, suggesting that perhaps a more interactive style of tutoring may in fact be even more effective for learning. The second result is that even though the tutors' explanations dominated the tutoring sessions, only shallow learning resulted from listening to these explanations. Thus, tutor explanations may have limited value, especially in light of our argument and evidence showing that the explanations seemed not to be sensitive to students' need for explanations. The third result is that there were far fewer deep than shallow follow-ups for all of the tutors' moves. This suggests that presumably, if we can get more deep follow-up type of responses, more learning of deep knowledge will ensue. The implications of these three results converge upon the rationale for the design of this second study. That is, ideally, we should be able to make the tutoring sessions more interactive with less didactic explanations, and hopefully this may result in more deep learning.

One way to make the tutoring dialogues more interactive from the tutors' perspective (recall that only 15% of the tutors' comments were interactive) is to suppress tutors' didactic style of tutoring, which is to give lots of explanations and feedback. If tutors' explanations and feedback did play an important role in students' learning, then without tutors' explanations and feedback, students' learning should be considerably depressed, for both shallow and deep knowledge, since tutor explanations and feedback did elicit the majority of both shallow and deep follow-up responses (a total of 52 vs. 21 for scaffolding and CGQs, see Fig. 3). Thus, the goal of this second study was to see if we can increase the frequency of interactive dialogues by restricting the kind of tutoring activities they may undertake. The procedure to do so was to supply tutors with content-free prompts that invite dialogues and constructive responses.

#### *3.1. Method*

##### *3.1.1. Participants*

The same 11 tutors were asked to return a week later for a second session of tutoring. Each of them tutored a different student, of the same age as Study 1 (8th graders), with a set of constraints, to be described below.

##### *3.1.2. Procedure for tutoring*

The instruction given to the tutors prior to the tutoring session was to suppress giving explanations, feedback and other extra information to the students, but instead, to invite dialogues from the students with prompts. A list of prompts (listed in Appendix E) was

provided to them as *examples* of the kind of comments they may make in the tutoring sessions. They were not restricted to saying these prompts in a verbatim way.

These prompts were designed with several goals and constraints in mind. First, they were open-ended, which seemed to be more conducive to inviting responses from the students, such as “What’s going on here?,” “Anything else to say about it?,” “Could you explain or put this in your own words?,” “What do you think?.” Second, they were content-free. This constraint pertains to a pragmatic reason, namely that they can be easily automated (Hausmann & Chi, in preparation). Finally, even though they were content-free, some of these prompts were analogous to deep scaffolding prompts, such as “Do you have any ideas/thoughts on why that might be the case? or ”Could you connect what you just read with what you have read before?.”

Besides giving them a list of prompts as guidance for their tutoring comments (which they could refer to during the actual prompting session), tutors were also given training on prompting, which consisted of two steps. First, they read a description on what prompting is and what it is not, that it is effective, and hypotheses for why it is effective. Second, they read several excerpts taken from a pilot study involving tutors prompting students, to get a sense of what a prompting dialogue looks like.

## 3.2. Results

### 3.2.1. Was the instructional manipulation effective?

In order to answer this question, we need to compare the tutoring processes of this study with the more didactic style of tutoring in Study 1. The tutoring session lasted on average 1.81 hr, which was almost identical to the amount of time taken in the didactic type of tutoring in Study 1 (which was 1.80 hr). These sessions generated another 11 binders of transcribed protocols, and these protocols were segmented and coded in the same three ways as the first study: statement coding, interaction coding, and episode coding.

Were the tutors successful at becoming more interactive and less didactic? Before presenting explicit evidence to support the effectiveness of the instructional manipulation, a look at the actual protocols gives a sense of the difference (Appendix F and C present the tutoring dialogues for the same tutors at the same sentences under the two conditions, prompting and tutoring, respectively). One can see that there were fewer long didactic explanations given by the tutors, and the dialogues seem much more interactive. Below, we present explicit evidence to show that the instructional manipulation was successful, when comparing their performance in the more didactic style of tutoring (Study 1) with the more interactive style of prompting (Study 2).

First, if tutors have learned to become more interactional and less didactic, then presumably they would not control the tutoring dialogue quite as much in prompting (Study 2) as in tutoring (Study 1). One way to capture this control is in terms of whom initiated the dialogue first. Fig. 4 shows the frequency with which either the tutor or the student spoke first. The results clearly show that not only did the tutors speak first much less often in the prompting condition than in the tutoring condition, but the students actually spoke first more often than the tutors in the prompting condition. (The interaction is significant at the 0.0001 level,  $t(20)=4.782$ .)

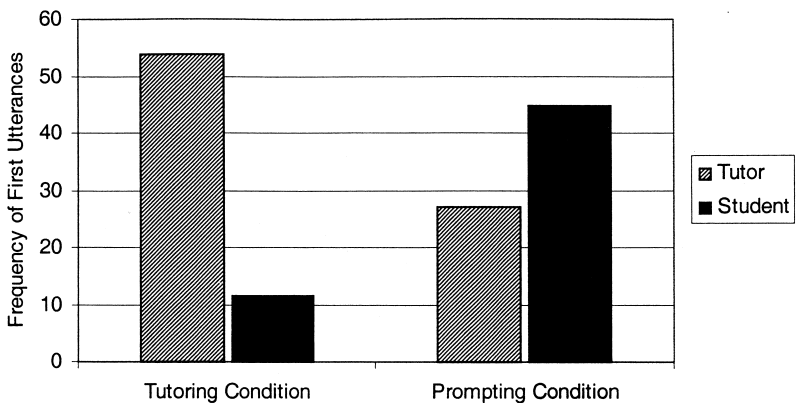


Fig. 4. Frequency of the first utterance per sentence.

Second, in addition to not speaking first, if tutors became more interactive and less didactic, then they also should have spoken less. Fig. 5 shows the number of statements tutors and students made. In the tutoring condition, the tutors made a total of 621 statements on average, whereas the same tutors in the prompting condition made only 178 statements on average. The reduction is highly significant,  $t(20)=5.59, p < 0.001$ . On the basis of this difference alone, one can surmise that tutors in the prompting condition did take our instruction seriously, and they suppressed explanations, feedback, and so forth. In contrast, and not surprisingly, the students in the tutoring condition spoke significantly less ( $M=206$ ) than they did in the prompting condition ( $M=375$ ),  $t(20)=3.47, p < 0.01$ .

Third, if the tutors were being less didactic in this study, then presumably they should take fewer turns and say less for each turn that they took. Fig. 6A shows the number of turns tutors and students took in tutoring and prompting. Clearly, tutors took a greater number of turns in tutoring, which means that they generally initiated and terminated the dialogues for each sentence. However, in prompting, the more equivalent number of turns taken by the tutors

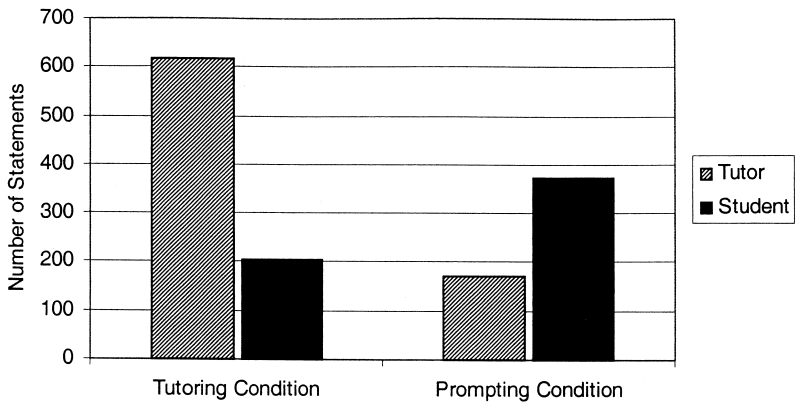


Fig. 5. Mean number of statements per tutoring session.

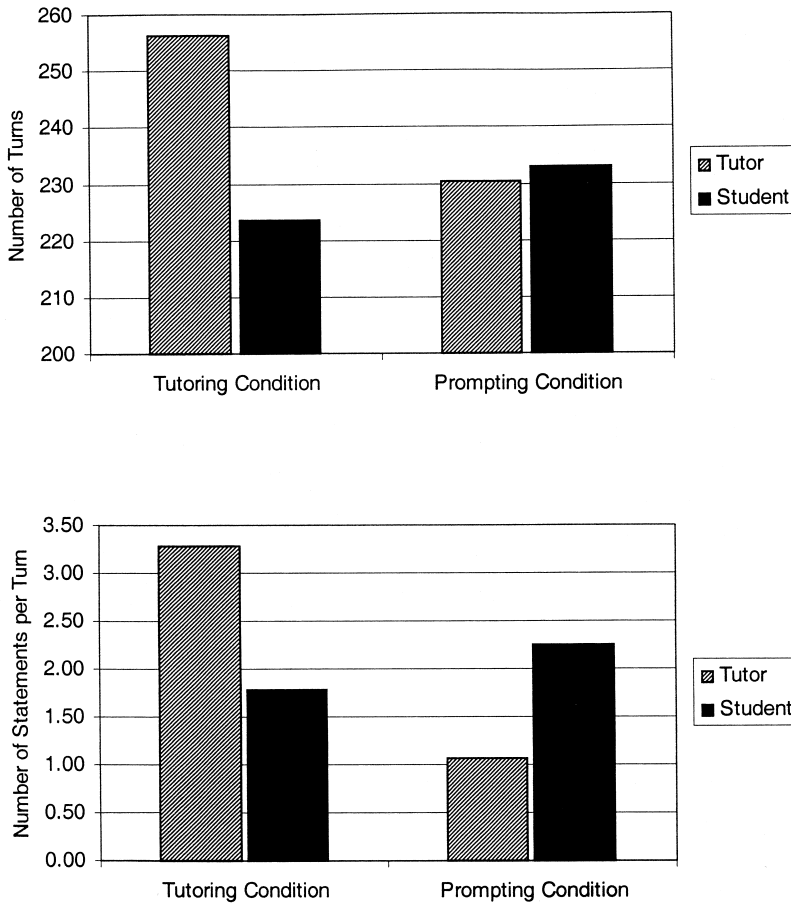


Fig. 6A. Mean number of turns per tutoring session. B. Mean number of statements per turn.

and the students indicate a more interactive style of dialogues. Fig. 6B shows the number of statements tutors and students took for each turn. Again, it is clear that the tutors took fewer turns and spoke fewer statements per turn in prompting (1.07) than in tutoring (3.29 statements per turn,  $t(1,20)=2.99$ ,  $p < .01$ .)

In sum, across these four analyses, there is no question that the instructional manipulation was successful in getting the tutors to be more interactive and less didactic. But did the tutors actually suppress giving explanations and feedback, and undertake more prompting? Fig. 7 shows the distribution of tutor and student actions from the statement coding. The differences in the frequency of tutor moves between tutoring and prompting were significant (either marginally, at the 0.05, or the 0.01 level) for all categories. For example, as can be seen by comparing Figs. 2 and 7, tutors' explanations dropped from 327 in tutoring (Fig. 2) to 7 in prompting (Fig. 7); similarly, tutors' feedback dropped from 94 in tutoring to 44 in prompting. Thus, the manipulation was clearly successful at suppressing tutors' explanations and feedback. In contrast, responsive to our instruction, the tutors were more interactive by undertaking many more scaffolding comments in prompting (56%, or 100 out of 178



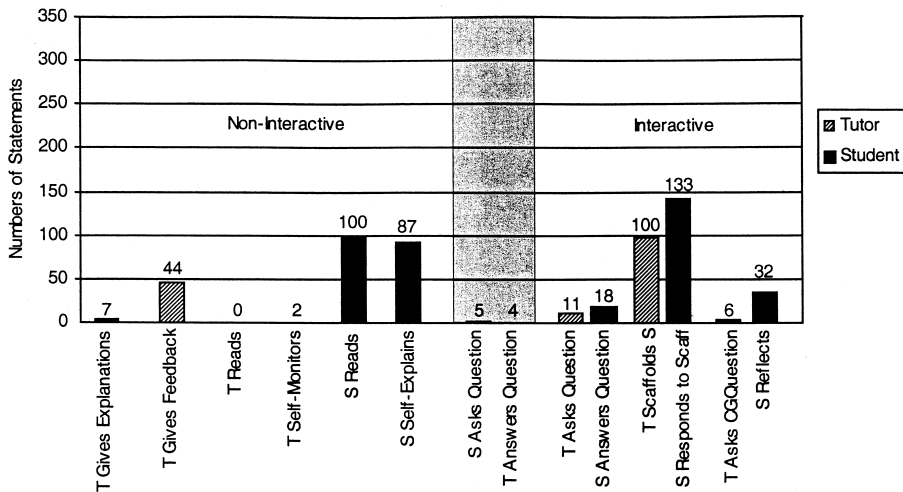


Fig. 7. Frequency of each type of tutor and student statements (Study 2).

statements, see Fig. 7), than in tutoring (5%, or 31 out of 621) in tutoring. Thus the number of scaffoldings increased substantially from 31 in tutoring to 100 in prompting.

An interesting result to note is that when the tutors prompted the students in Study 2, they did not always use content-free prompts that we had provided. Instead they used scaffolding prompts (meaning that they were no longer content-free). As listed earlier, scaffolding prompts consist of a variety of moves, such as pumping for more, hinting, requesting to fill-in-a-blank, and asking a leading question. Although we constructed our prompts to be content-free in the sense that they *did not* contain content words, tutors delivered these prompts with content, such as “And what would you expect to be passed back and forth?” as highlighted in the prompting dialogues presented in Appendix D. More examples can be seen in Appendix F, which shows the dialogues carried out by the same tutor at the same sentence in prompting. For both Tutor #1 and #10, they replaced content-free prompts with content words, such as “What do you, what do you think a septum is?” for Tutor #1, and “What does blood do?” for Tutor #10, and these are highlighted. But they also use content-free prompts at times, such as “What do those two sentences mean to you?” (Tutor #1, Appendix F). We coded the frequency of each kind of prompts. The result show that tutors delivered scaffolding prompts much more often (73.18% of the times) than content-free prompts (26.54% of the times). This indicates that scaffolding is a relatively easy and natural tactic to implement, and it most likely is driven by the content knowledge of the local interactions. That is, tutors scaffold students on the basis of what the students have just said in order to make the students’ responses more complete, more elaborate, or more correct.

### 3.2.2. Overall learning

Having established that the prompting sessions successfully suppressed tutors’ explanations and feedback, was the students’ learning also depressed? Using the same analyses as carried out in the first study, students’ learning can be assessed by the gains from the pretest to the post-test in terms of both the knowledge pieces expressed in the definition of terms,

Table 4  
Proportion of correct answers on post-test questions

Categories	1	2	3	4
Tutored students (Study 1)	64	45	35	31
Prompted students (Study 2)	65	46	33	41

as well as the mental model improvement. The mean proportion of knowledge pieces expressed in the pretest was 22%, and at the post-test it was 45%, with a gain of 23%, which is significant at the 0.001 level,  $t(10)=7.86$ .

With respect to the improvement in mental models, as in Study 1, there were no correct Double Loop-2 models represented at the pretest, and 7/11 students represented the circulatory system with the correct Double Loop-2 model (as compared with 8/11 in Study 1). Counting the number of increments in the mental model progression gives an improvement of 2.18 (as compared with 2.0 in Study 1). Basically, as in the tutoring condition, there was significant learning.

3.3. Conclusion

By suppressing tutors’ explanations and feedback, the tutoring dialogues became much more interactive. Moreover, students did learn significantly from the pretest to the post-test in this less didactic and more interactive style of tutoring. In order to address how well the students learned in comparison to a more didactic (naturalistic) style of tutoring, the next section compares the performance of the students in tutoring versus prompting.

4. Comparison of study 1 and 2 results

4.1. Question-answering at the post-test

Knowing that students in the prompting condition did learn significantly, the question is how well did they learn when the tutors gave essentially no explanations and very little feedback? If the students in prompting did less well than students in tutoring, then one might conclude that tutor explanations were central in fostering students’ learning.

To answer this question, we compared the proportion of correct answers for the tutored (Study 1) and the prompted (Study 2) students. Table 4 shows the mean proportion correct for each category of questions for the tutored and the prompted students. There were no significant differences among the means (controlling for pretest and reading scores). Thus, eliminating tutor explanations and reducing tutor feedback did not hinder the prompted students’ learning of either the shallow or the deep knowledge. (In fact, the prompted students even knew more answers to the health-related Category 4 transfer questions than the tutored students, 41% vs. 31%, which is surprising given that a great deal of the tutors’ explanations in Study 1 contained extraneous health-related information). A similar study contrasting didactic tutoring and Socratic tutoring (analogous to our tutoring and prompting

conditions, respectively) by a single tutor, also found no significant differences in learning concepts in basic electricity and electronics (Rose, Moore, VanLehn & Allbritton, 2000).

Perhaps the way to see how effectively students can learn without the benefit of tutors' explanations is to get a sense of exactly how helpful tutors' explanations were. We had documented earlier that 46% of the tutors' statements contained relevant new or extra information (whereas 54% contained redundant information in the forms of elaborations and repetitions). However, it is difficult to assess precisely how helpful such extra information is, since different pieces of it can be combined and used to answer the post-test questions. However, one approximate way to determine the effectiveness of this extra information is to identify whether the new information actually gave away the answers to any of the questions directly. Thus we recoded the tutors' explanations from Study 1 to see whether they explicitly stated the answers to any of the questions in the post-test. We found that for 9 of the 11 tutored students, they did receive explicit answers to 32 questions, totaling to a possible score of 169 points. However, even though the answers were given explicitly, the tutored students only scored a total of 127 points on these questions. In contrast, the prompted students (of the same 9 tutors who gave away these answers) scored a total of 75 points on these same questions. This means that tutored students were able to learn or recall 75% (127/169) of what the tutors explained to them, whereas the prompted students were able to construct on their own, only 44% (75/169) of the answers. This also means that in order for the tutored and the prompted students to get similar averaged post-test scores across the four categories, the prompted students must have constructed more answers correctly on those questions for which the tutors did not give away the answers explicitly, suggesting that construction may be more effective than hearing tutor explanations (supporting the S-hypothesis). In sum, although the benefit of hearing tutor explanations seems substantial, in that the tutored students have learned or recalled 75% of the explicit answers in the explanations, this result also suggests that tutored students may *not* have fared as well as the prompted students, had their tutors restricted the content of their explanations to nonanswers and redundant information only.

Although there was no control group in which the students merely read the text by themselves, it should be noted that such a control group was used in the Chi et al. (1994) study, whereby a prompted group (analogous to the prompted group here, except a single tutor prompted all 14 students, and the prompts were largely content-free rather than scaffolding prompts) was contrasted with a control group in which the 10 students read the text twice by themselves. In that study, using an almost identical text passage as the one used in this study, the prompted students answered significantly more questions correctly on the post-test than the control group: The difference is equivalent to a 0.37 sigma effect, as measured by a standard deviation unit (see Bloom, 1984; Cohen et al., 1982). Moreover, the differences in the scores between the prompted group and the control group were more pronounced for the harder inference and health questions (Categories 3 & 4), suggesting that they achieved deeper understanding. Reasoning by transitive inference, one might assume that the scores from the prompting and tutoring conditions here would be significantly higher than those from a control group, had one been tested. Because these kinds of studies that use protocols are extremely labor-intensive to code, we decided not to test a control group for this study. Unfortunately, a direct comparison cannot be made between the prompting groups of the current study and the 1994 study because the two studies did not use identical test questions, nor were the text passages exactly identical.

#### 4.1.1. How did the students learn equally effectively in the tutoring and the prompting conditions?

There are basically three conjectures for how students learned without tutor explanations. First, students seem to be more constructive in prompting than in tutoring, thus suggesting that it is these constructive activities that promoted learning in both conditions. Second, a more interactive dialogue pattern gave students greater opportunities to master an idea/concept through multiturn scaffolding episodes. Finally, students also took greater responsibility for learning. The evidence for each of these conjectures is presented next.

*4.1.1a. Students were more constructive.* Looking at Fig. 5, a descriptive answer to the question of how students learned equivalently in the two conditions is to say that students' constructive responses must have compensated for the absence of tutors' explanations. Although students' responses (in terms of the total number of statements made) went up significantly from 206 to 375, only two categories of moves increased significantly. These were the number of scaffolded responses (from 43 to 133) and the number of reading statements (from 12 to 100, compare Figs. 2 and 7). However, this simple answer would predict that the frequency of students' moves, such as their responses to scaffoldings and reading statements, should correlate with learning, but they did not. In fact, a step-wise regression analysis showed that the students' post-test performance did not correlate with the frequency of *any* moves, whether undertaken by the tutors or the students (controlling again for the pretest and reading ability scores). This means that the frequency of any of the activities tutors and students undertook (when considered alone) in this prompting condition did NOT effect how well the students learned. Curiously then, how did the students learn, given that they learned as well in this suppressed tutoring condition?

The most salient difference between tutoring and prompting is the amount of interaction. Yet, several quantitative analyses of the frequency of interaction were undertaken, and none of them correlated with learning. These quantitative analyses consisted of correlating learning outcomes with: (1) the frequency of turns, (2) the frequency of scaffolding episodes, and (3) the weighted frequency of interaction (i.e., we counted and weighted the proportion of sentences that contained more or less interaction in terms of the number of turns, and whether or not they contained scaffolding episodes.) The only logical conclusion from the above analyses is that learning in prompting was not correlated with the frequency of either the *format* of the moves (i.e., whether it was a scaffolding response or a question-asking move), or the amount of interactions. As alluded to earlier, there may be something about the nature of interactive dialogues that we have not captured. This suggests that a *content* analysis might be more informative, as shown by the interaction analysis in Fig. 8.

If we compare Figs. 3 and 8, it turns out that there is a slight (but nonsignificant) increase in the total number of shallow follow-ups and deep follow-ups between the two conditions. There were 54 shallow follow-ups in tutoring and 67 in prompting, and 19 deep follow-ups in tutoring and 29 in prompting. The increases, of 24% and 53% respectively, suggest that the amount of content-relevant constructive responses students undertook (regardless of which tutoring moves elicited such construction), might compensate for the lack of tutor explanations in the prompting condition, again supporting the S-hypothesis. The number of non-constructive responses (i.e., continuers) dropped from a total of 48 in tutoring to 8 in

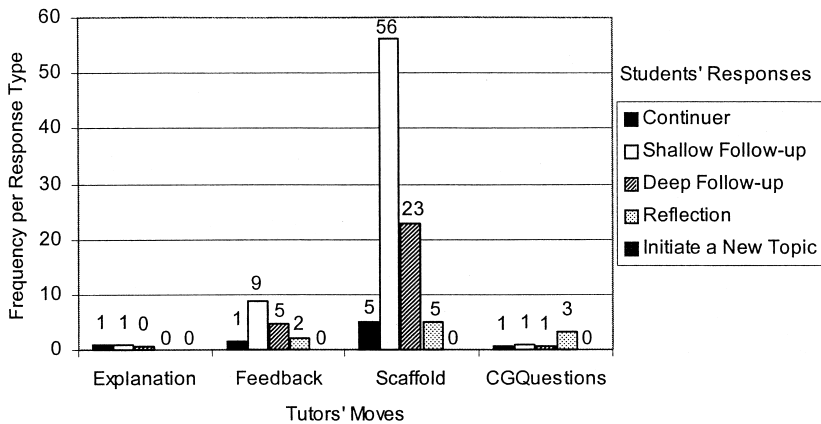


Fig. 8. Frequency of students' responses as a function of tutors' moves in the prompting condition (Study 2).

prompting. The relatively small number of continuers in prompting suggests that they could not have contributed much to learning, further affirming our interpretation that interaction per se is not sufficient for learning, but the interactions must contain substantive contributions from the students.

*4.1.1b. Greater and deeper scaffolding episodes.* An episode of scaffolding, as defined previously, is a dialogue that consists of multiple turns in which the tutors scaffold the students on the same concept or idea. Thus, a more interactive dialogue pattern, as occurred in prompting, may have produced more extended episodes of scaffoldings. To see whether this is true, analogous to the analysis carried out in Study 1, scaffolding episodes were identified by two coders for the prompting condition. A total of 98 scaffolding episodes were identified across the 11 students, which is almost twice as many as the number in tutoring (52). Thus, in the prompting condition, not only did the number of scaffolding prompts increased, but the number of scaffolding episodes also increased. Moreover, there was a significant increase (from the tutoring to the prompting condition) in the proportion of scaffolding episodes that were deep (from 0.91 to 3.91 per student,  $t(20)=2.52$ ,  $p < .05$ ), but not for the shallow (from 3.91 to 5.27). (See Fig. 9.) This suggests that an interactive style of dialogue may naturally allow tutors to pursue more extended scaffolding episodes because the students' greater participation in the dialogues allow the tutors to assess what they do not understand. Thus, knowing what students have not understood may have encouraged the tutors to keep scaffolding until the students have mastered the concept (resulting in a greater amount of and deeper scaffolding episodes in prompting).

*4.1.1c. Students took greater responsibility for learning.* Suppressing tutors' explanations and feedback also meant that the control of learning might shift from the tutors to the students. This can be seen in the significant increase in the frequency with which students read in prompting (100 times, compare Fig. 2 and 7) versus in tutoring (on average 13 times). One outcome of this greater reading activity in the prompting but not the tutoring condition, is that the effectiveness of the reading was determined by the students' reading or compre-

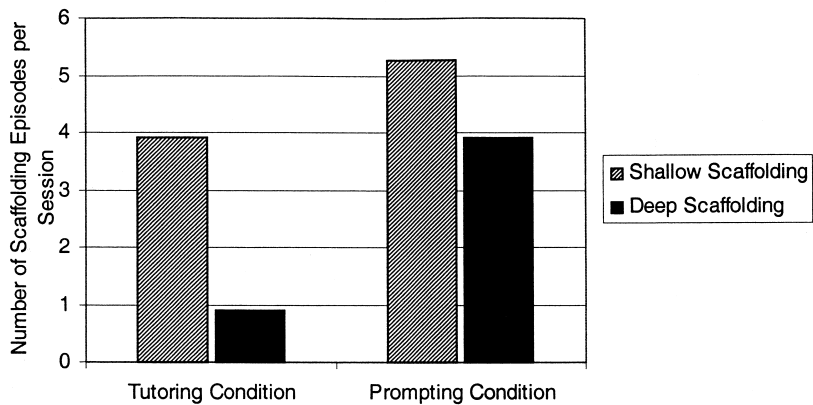


Fig. 9. Mean number of episodes per tutoring session.

hension ability. This can be seen by the significant correlation between their reading ability and learning. Table 5 shows that there was a highly significant correlation between reading ability with *every* category of the post-test measures in the prompting condition but *not for any* of the post-test measures in the tutoring condition. (The correlation values and the significance values are shown in Table 5, controlling for pretest). Thus, the presence of a strong positive correlation of reading ability with learning in the prompting but not in the tutoring condition, suggests that the students took more effort to learn in the prompting condition. (Notice that the frequency of reading itself does not correlate with learning in either condition. This is because the degree to which students learned did not depend on how frequently they read the sentences, but rather, on how much they can get out of reading, which is limited by their comprehension ability.)

In sum, we can tentatively construct the following interpretation for how students learned in prompting (without the benefit of tutor explanations). Basically, students in prompting were more constructive overall since the tutors prompted and scaffolded them more often than in tutoring. The students’ greater opportunities to be more constructive overall means that they displayed more of what they knew and didn’t know to the tutors. This display in turn allows the tutors to evaluate more accurately what the students know and don’t know. This evaluation in turn allows the tutors to pursue extended scaffolding episodes on concepts that the students did not know until the students have mastered them. Finally, the students

Table 5  
Correlations of reading ability and post-test questions (controlling for pre-test scores)

		Cat. 1	Cat. 2	Cat. 3	Cat. 4 (HQ)
Tutoring	Reading scores	r = -0.0488	0.3115	-0.0357	-0.0156
		p = .893	p = .381	p = .922	p = .966
Prompting	Reading scores	r = 0.6494	0.8606	0.8475	0.6535
		p = .042	p = .001	p = .002	p = .040



also put more effort into and took greater control of learning, as reflected in their greater frequency of reading. In contrast, students in the tutoring condition learned from a combination of being constructive and from hearing explanations that contained extra information, were presented in everyday language, were repetitive, and so forth, but probably not because tutors were necessarily more adaptive in the tutoring condition. Thus, the processes of learning differed in the two conditions.

## **5. General discussion**

What have we learned about human tutoring from these two studies on naturalistic tutoring and suppressed tutoring? The prevailing notion is that tutoring effectiveness derives from the correct and appropriate application of pedagogical skills that are specific to tutoring. Our goal for these studies was an attempt to highlight, in addition, the possible roles of students' construction in learning from tutoring as well as the effect of interaction on learning. Thus we propose that tutoring effectiveness can arise not merely from tutors' pedagogical skills, but also from students' construction, especially substantive construction that results from interactions of the tutors and the students.

Our analyses in Study 1 explicitly sought evidence in support of all three hypotheses, which required that we coded and analyzed the data from each of the three perspectives: the tutor-centered one, the student-centered one and an interactive one. The results of Study 1 basically uncovered evidence in support of all three hypotheses. Thus, even though we replicated the general findings in the tutoring literature, in showing that tutors did dominate and led the tutoring dialogues, nevertheless, we were able to uncover the subtle contributions of students' construction as well as the tutor-student interactions toward the effectiveness of learning from tutoring. To further assess the value of interactions, we maximized opportunities for interactions in Study 2 by asking tutors to suppress explanations and feedback, and instead prompt the students for construction. The results showed that students learned just as well, without the benefit of hearing any tutor explanations and feedback.

Although the prompted students did not learn more in the more interactive style of tutoring (although they might have, had the tutored students not receive so much extra information in the tutors' explanations), we might consider in what ways a more interactive style of tutoring (as in Study 2) can be beneficial. There are three possible benefits that need to be further explored. First, an interactive style seems to be more motivating, which can lead to a greater enjoyment of learning. Second, there is some suggestion that tutors are more accurate in their assessment of students' understanding when they are prompting and scaffolding the students, rather than giving explanations (see Chi, Siler & Jeong, in preparation). This can be explained by the fact that tutors may have more cognitive resources left for such an assessment, as well as the fact that students, by talking more, are displaying more of their understanding for the tutors to evaluate. Third, there is a hint that the students in the interactional style of tutoring (Study 2) did seem to be able to transfer their knowledge better than the students in the didactic style of tutoring. For example, the fact that the prompted students scored higher for Category 4 health questions is impressive given that many of the tutors' explanations in the tutoring condition contained extraneous but relevant health-related

information, which can be used to answer the health questions. Without these extra information, the students in prompting might have performed significantly better.

We close with some remarks about what the results of these studies implicate about the development of an ITS. The results of both studies clearly show that students' substantive construction from interaction is important for learning, suggesting that an ITS ought to implement ways to elicit students' constructive responses. One way to elicit construction is through scaffolding. In Study 2, it appears that tutors were extremely successful with scaffolding, even though it seems to be a complex skill and no instruction was provided on how it is done. However, it is daunting to think about implementing a scaffolding type of prompting in an ITS because scaffolding requires an understanding of what the student uttered, whether it was locally correct, incorrect, or incomplete, or somehow globally flawed (Chi, 2000). Such understanding may be all that is needed to determine what kind of scaffolding prompts to give and when to give them because everyday tutors seem to be able to scaffold deeply in extended scaffolding episodes naturally, without any training. Obviously natural language understanding and an understanding of the content domain are crucial for appropriate scaffolding (although other knowledge and skills may also be required for more sophisticated type of scaffolding, such as the use entrapment strategy in a Socratic style of dialogue; Collins & Stevens, 1982). Therefore, it seems necessary for future generations of ITSs to incorporate natural language capabilities.

## Notes

1. Note, if one is tutoring reading, then reading skill itself becomes the content domain, and one may need to know something about reading, such as how to decompose words into their phonemes, before one can be an effective reading tutor.
2. Both students and tutors were also asked to draw the bloodpath after sentences 24 and 76. However, these data will be reported in Chi, Siler and Jeong (in preparation).
3. After the students completed the entire study, they were given a set of domain-relevant questions to take home and answer. The domain-relevant questions were generally returned to the experimenter within a week. These data will also not be presented in this paper because there were no differences between the groups.
4. Tutoring protocols from other tutoring studies in the literature may not contain such a large amount of explanations. If so, a major difference might be that this study used a conceptual domain, whereas the majority of the other studies were carried out in the context of problem solving.

## Acknowledgment

Funding for this research is provided by the Spencer Foundation to the first author, and in part by the National Science Foundation, Grant Number NSF (LIS): 9720359, to the Center for Interdisciplinary Research on Constructive Learning Environment (CIRCLE, <http://www.pitt.edu/~circle>). The authors appreciate the thoughtful comments from John Ander-

son, James Paul Gee, Art Graesser, Jim Greeno, and Kurt VanLehn. Copies for this paper may be requested from Micki Chi, or downloaded from the WEB: [www.pitt.edu/~Chi](http://www.pitt.edu/~Chi).

### **Appendix A. The text passage of 86 sentences**

1. Human life depends on the distribution of oxygen, hormones, and nutrients to the cells in all parts of the body and on the removal of carbon dioxide and other wastes.
2. These tasks are partially carried out by the circulatory system, which consists of the heart, and intricate network of blood vessels, and blood.
3. The blood, moving through the vessels serves as the transport medium for oxygen, nutrients, and other substances.
4. The heart is a muscular organ that pumps blood through the body.
5. The heart consists of cardiac muscles, nervous tissue, and connective tissues.
6. The septum divides the heart lengthwise into two sides.
7. The right side pumps blood to the lungs, and the left side pumps blood to the other parts of the body.
8. Each side of the heart is divided into an upper and a lower chamber.
9. Each lower chamber is called a ventricle.
10. In each side of the heart blood flows from the atrium to the ventricle.
11. One-way valves separate these chambers and prevent blood from moving in the wrong direction.
12. The atrioventricular valves (a-v) separate the atria from the ventricles.
13. The a-v valve on the right side is the tricuspid valve, and the a-v valve on the left is the bicuspid valve.
14. Blood also flows out of the ventricles.
15. Two semilunar (s-l) valves separate the ventricles from the large vessels through which blood flows out of the heart.
16. Each of the valves consists of flaps of tissue that open as blood is pumped out of the ventricles.
17. Blood returning to the heart, which has a high concentration, or density, of carbon dioxide and a low concentration of oxygen, enters the right atrium.
18. The atrium pumps it through the tricuspid valve into the right ventricle.
19. The muscles of the right ventricle contract and force the blood through the right semilunar valve and into vessels leading to the lungs.
20. Each upper chamber is called an atrium.
21. In the lungs, carbon dioxide leaves the circulating blood and oxygen enters it.
22. The oxygenated blood returns to the left atrium of the heart.
23. The oxygenated blood is then pumped through the bicuspid valve into the left ventricle.
24. Strong contractions of the muscles of the left ventricle force the blood through the semilunar valve, into a large blood vessel, and then throughout the body.
25. The contraction just described can also be described as having two phases.
26. In one phase, called systole, the ventricles contract and force blood into the arteries.
27. In another phase, called diastole, the ventricles relax, and the blood flows in from the atria.

28. The result is the characteristic ‘lubb dup’ we call a heartbeat.
29. Each beat of the heart forces blood through blood vessels.
30. The large, muscular vessels that carry blood away from the heart are called arteries.
31. The thick walls of the arteries have three layers, an inner endothelium, a middle layer of smooth muscle, and an outer layer of connective tissue.
32. This structure gives arteries a combination of strength and elasticity that allows them to stretch as blood under high pressure enters them from the heart.
33. You can feel this stretching of arteries as the pulse in your wrist.
34. When the left ventricle contracts, it forces blood through the left semilunar valve into the body’s largest artery, the aorta.
35. From the aorta, blood travels through a network of smaller arteries, which in turn divide and form even smaller vessels called arterioles.
36. The arterioles branch into a fan of tiny vessels called capillaries.
37. Capillary walls are only one cell thick.
38. Movement of gases and nutrients takes place across the thin capillary walls mostly from areas of greater concentration to areas of lesser concentration.
39. This process is called diffusion.
40. Diffusion is the process by which molecules spread from an area of greater concentration (or density) to an area of lesser concentration.
41. Molecules, including gases and nutrients, are in constant, random motion, traveling in straight lines until they hit something, then they rebound, traveling in a different direction.
42. An example of diffusion occurs when you put a cube of sugar in water.
43. At first the molecules of sugar are more concentrated in and around the sugar cube and less concentrated in the water farther from the cube.
44. The sugar molecules will spread away from the cube into the water.
45. Eventually, the sugar molecules spread throughout the water so that they are equally distributed.
46. When the concentration of the molecules of a substance is the same throughout a space, a state called equilibrium exists.
47. Once equilibrium is established, the random movement of molecules continues, and equilibrium is maintained.
48. Some kinds of molecules can also pass through a membrane.
49. They do this by moving between the molecules that make up the membrane.
50. If a substance can pass through a membrane, the membrane is permeable to it.
51. As with all diffusion, molecules spread from an area of greater concentration on one side of the membrane to one of lesser concentration on the other.
52. This is what occurs with the gases and nutrients which cross the thin membrane wall of the capillaries.
53. If the concentrations of oxygen and nutrients are higher in the blood than in the surrounding cells, these substances will move from the blood into the cells.
54. In this way cells receive oxygen and nutrients and in a similar manner wastes leave the cells and enter the blood.
55. In the body cells, oxygen is consumed.

56. De-oxygenated blood flows through capillaries that merge and form larger vessels called venules.
57. Several venules in turn unite to form a vein, a large blood vessel that carries blood to the heart.
58. The walls of veins, which are also composed of three layers, are thinner and less muscular than those of arteries.
59. By the time blood reaches the veins, it is under much less pressure than it is in the arteries.
60. Thus, the veins are not as strong or as flexible as the arteries.
61. Many veins pass through skeletal muscles.
62. During movements, these muscles contract, squeezing blood through the veins.
63. Valves prevent the blood from moving backward or downward.
64. These valves, which are usually located above the point at which two veins meet, allow blood to flow in only one direction through the veins.
65. However, when muscle tone is poor, a person's valves may not fully come together, which will often allow some blood to seep backward through the valves.
66. The English scientist William Harvey (1578–1657) first showed that the heart and blood vessels formed one continuous, closed system of circulation.
67. To communicate effectively, scientists often say this system is divided into various subsystems.
68. The two primary subsystems of circulation are pulmonary circulation, in which the blood travels between the heart and lungs, and systemic circulation, in which the blood travels between the heart and all other body tissues.
69. Pulmonary circulation is the movement of blood from the heart to the lungs and back to the heart.
70. Blood from the right ventricle flows through the right semilunar valve into the pulmonary artery and then to the lungs.
71. The pulmonary artery is the only artery that carries de-oxygenated blood.
72. At the lungs the pulmonary artery divides into two smaller arteries, one leading to each lung.
73. These arteries branch into arterioles and then into capillaries.
74. In the capillaries, the carbon dioxide diffuses out of the blood during exhalation, and the oxygen diffuses into the blood during inhalation.
75. Oxygenated-blood then flows into venules, which merge into the pulmonary veins that lead to the left atrium of the heart.
76. The pulmonary veins are the only veins that carry oxygenated-blood.
77. Systemic circulation is the movement of blood to all parts of the body, except the lungs, and then back to the heart.
78. Blood from the left ventricle enters systemic circulation through the aorta.
79. Systemic circulation is further divided into several other subsystems.
80. These subsystems are coronary circulation, renal circulation, and hepatic portal circulation.
81. Coronary circulation is the subsystem of systemic circulation that supplies blood to the heart itself.

82. Renal circulation is the subsystem of systemic circulation that moves blood through the kidneys and back to the heart.
83. Hepatic portal circulation is the subsystem of systemic circulation that moves the blood through the digestive tract and liver.
84. Blood is the liquid connective tissue that forms the transport medium of the circulatory system.
85. The major function of the blood is to transport nutrients and oxygen to the cells and to carry away carbon dioxide and nitrogenous wastes from the cells.
86. Blood also transfers body heat to the body surface and plays a major role in defending the body against disease.

## **Appendix B. Pre- and post-test terms and four categories of post-test questions**

1. Aorta
2. Artery
3. Atrium
4. Blood
5. Capillary
6. Carbon Dioxide
7. Circulatory System
8. Concentration
9. Diffusion
10. Equilibrium
11. Heart
12. Lungs
13. Nutrients
14. Oxygen
15. Permeable
16. Pulmonary Circulation
17. Septum
18. Systemic Circulation
19. Valve
20. Vein
21. Ventricle

### **Posttest Questions:**

#### **Category 1 - Text Explicit Questions:**

1. Why are there valves in the heart?
2. Which artery carries deoxygenated blood?
3. From the heart, where does blood travel in systemic circulation?
4. What does the heart do?
5. Why is blood vital to human life? That is, what is its primary purpose?
6. Which vein carries oxygenated blood?
7. Where is the blood oxygenated?



8. Where does the blood go in coronary circulation and why?
9. Where does the right ventricle pump blood to?
10. Where does the blood go to when it comes out of the left ventricle of the heart?
11. Which part of the heart does the blood go to after it's oxygenated?
12. Where does the blood go to after it delivers oxygen to body cells?
13. Does blood travel from the left ventricle to the right atrium?
14. What happens to sugar molecules when you put a sugar cube in a glass of water?
15. What eventually happen to the sugar molecules in the glass of water?
16. Describe the motion of molecules, including gases and nutrients (in the blood stream).

Category 2 - Text Implicit Questions:

1. Why do we have valves in veins, but not in arteries and capillaries?
2. What is it that you are feeling when you take your pulse?
3. Why are capillary walls only one cell thick?
4. Why is there an artery that carries deoxygenated blood?
5. Why do vessels get increasingly smaller as they get close to the body cells, and get increasingly larger as they get nearer to the heart?
6. Does the blood change in any way as it passes through the heart?
7. Is the septum solid or porous? That is, can blood pass through it from one side of the heart to the other?
8. Does blood in veins flow in both directions, up toward the heart, and down away from the heart?
9. Is the blood always carried in blood vessels or does it float freely in the body as well?
10. In which kind of blood vessels (arteries, veins, or capillaries) is the blood pressure the lowest? Why?
11. Besides the role of valves in the veins, what keeps blood from the lower parts of the body moving up toward the heart through the veins, against gravity?
12. Why are the artery walls thicker than vein walls?
13. What keeps the blood flowing in the given direction when it leaves the heart?
14. How do carbon dioxide and wastes get into the blood stream from the cells?
15. How can we tell whether a blood vessel is a vein or an artery, on the basis of its purpose?

Category 3 - Mental Model Implicit Questions:

1. Why don't we have valves in pulmonary veins?
2. Why does the circulatory system need an atrium? Why can't blood flow straight through each ventricle in a pipe-like fashion?
3. What is the consequence at the cellular level of having a hole in the septum?
4. Why is your right ventricle less muscular than your left ventricle?
5. Why do we sometimes refer to the heart as a "double pump"?
6. The artery that carries blood from the right side of the heart to the lungs (the pulmonary artery) carries about the same amount of blood as the artery that carries blood from the left side of the heart to the rest of the body (aorta). Why do they carry the same amount of blood?
7. Why can't the blood go straight to the lungs after it comes back from the body? In other words, why does the blood need to go to the heart first before it goes to the lungs?
8. What is the difference between blood pressure and pulse rate?

9. What would happen if the valves between the atria and the ventricles are stuck open and wouldn't close?
10. How does the oxygen get into the blood and how does carbon dioxide exit from the blood in the lungs?
11. When you put a cube of sugar in a glass of water, the sugar molecules will diffuse throughout the water until they are equally distributed or a state of equilibrium is reached. Why ISN'T a state of equilibrium reached in the human body between the molecules of oxygen in the blood and the molecules of oxygen in the body cells?
12. Humans get many of the nutrients they need from the food they eat. If food is broken down in the digestive system, how do nutrients get into the blood stream to be delivered to cells?
13. Suppose you have a glass of sugar water (20 ml of water mixed with 1 spoonful of sugar). Suppose you also have a balloon filled with 5 ml of water and 2 spoonfuls of sugar. Is the concentration of sugar in the glass greater than, less than, or equal to the concentration of sugar in the balloon?
  - a) The balloon is permeable to both water and sugar. You taste the water in the glass, then you put the balloon into the glass and allow the system to come to equilibrium. You taste the water again. Does the water in the glass taste sweeter, less sweet, or as sweet as it tasted before? Why?
  - b) Does any sugar or water pass through the membrane after equilibrium is reached?
  - c) Why do the sugar molecules spread in the water?
  - d) Now say you have a glass of sugar water which is 20 ml water and two spoonfuls of sugar and a balloon which is 5 ml water and one spoonful of sugar. Does the balloon or glass have the higher concentration of sugar water? Again you put the balloon into the water. What will happen?
14. Draw arrows describing the directions of motion of the oxygen and nutrient molecules in the bloodstream in the capillaries as they're diffusing into the body cells. What do you think happens to the movement of molecules after they've diffused into the cells?
15. Say you have a permeable membrane separating water with a certain concentration of sugar on the left side of the membrane from water with a slightly lower concentration of sugar on the right side. During diffusion, will any sugar move across the membrane from the right side to the left side (from lower to higher concentration)? Why or why not?
  - a) If you were to add equal amounts of sugar to each side of the membrane (which have equal amounts of water) after equilibrium has been reached, what would happen with respect to movement of sugar across the membrane? What would happen if you added sugar to the left side but not the right?
  - b) Does the number of particles passing through the membrane (from the left to the right side) at any point in time depend on the concentration difference between the two sides of the membrane? That is, are the same number of sugar molecules passing through the membrane from the left to the right side at the beginning of diffusion as compared to later on (when the concentrations of sugar are close to equal on either side of the membrane)?

16. Describe in your own words why the net movement of molecules is from higher to lower concentration (such as in the case of oxygen movement between the blood and body cells). Give as much detail as possible.
17. If the heart stopped pumping for one minute, would all of the waste and carbon dioxide in the cells go into the blood stream?

Category 4 - Health Questions:

1. Doctors warn against having too much of a fatty substances called cholesterol in your diet, because cholesterol builds up on the lining of the coronary arteries and clogs them. Why is this dangerous?
2. Exercise strengthens muscles. Why is this good for circulation?
3. Why do older people who have poor circulation also have cold feet and hands?
4. When we get our blood pressure taken at the doctor's office, they read us two numbers, such as 120/80. The upper number measures the systolic pressure, and the lower number measures the diastolic pressure.
  - (a) What do these numbers measure and why is the upper number larger than the lower number?
  - (b) Why is it dangerous for the upper number to be too high?
5. Many antismoking groups claim that nicotine reacts with cells in the brain which cause smokers to become addicted to cigarettes. If smoke is inhaled into the lungs, how does nicotine get to the cells in the brain?
6. Why is it that often, when the heart stops functioning properly, it is the failure of the left ventricle (not the right ventricle)?
7. It's well known that high blood pressure is dangerous. Is it dangerous to have low blood pressure too? Why?
8. Alcohol initially expands the peripheral blood vessels. As a result, the heart beats faster right after we drink alcohol. How would the expansion of blood vessels lead to a faster heart beat?
9. People sometimes feels dizzy when they suddenly stand up after sitting on the ground for awhile. Why do you think this happens?
10. The short-term effect of exercise is faster heart beat. However, the long-term effect of exercise is slower heart beat. Why?
11. When people faint, one of the first thing you need to do is to lower their head (e.g., by putting pillows under their feet). Why does this help people to regain consciousness?

**Appendix C. Samples of protocols for tutors #1 and #10, at sentences 7 and 85.  
Tutoring condition (Study 1)**

Tutor #1 Student MH

Sentence 7) The right side pumps blood to the lungs and the left side pumps blood to other parts of the body.

T: That's right, the right side receives the blood, pumps it into the lungs, the lungs bring it back into the left side and the left side pumps it to the left side through the aorta.

S: It is like one big loopy thing.

T: Yeah, it is like a big loop, you have your heart there and here is your septum right here. . . and there is another division there. You have your right atria here and your right ventricle and your blood. . . when it come back up to the body, it is deoxygenated. It goes into the right atrium.

S: It come in here and then it goes out near the lungs and then it goes back up through here.

T: Right. It goes in through here, it goes to the pulmonary artery and goes to the lungs and comes back to the pulmonary vein into the left atria, goes through the left ventricle, the left ventricle goes through the aorta, which arches up like this and it goes back to rest of the body and up through the body and it completes its cycle again. That is what they said later on, that is closed system.

Tutor #10 Student LC

Sentence 85) The major function of the blood is to transport nutrients and oxygen to the cells and to carry away carbon dioxide and nitrogenous wastes from the cells.

T: You want to carry your food and all your oxygen and everything, okay? You want to carry that to all of your cells. You want to take that to and carry away and from the cells you want take away that carbon dioxide and nitrogenous wastes. That is kind of like the same thing. It is kind of like bad liquids. Like is like bad gases, we want to get rid of bad liquid wastes.

S: Nitrogenous.

T: Yeah, you want to get that out of your cells because that. . . if you have it in you too long it kind of poisons you. You want to get rid of all that stuff. You want to get good oxygen in. Everything like that.

#### Appendix D. Example of Shallow Scaffolding (marked by the bracket)

Sentence 84) These arteries branch into arterioles and then into capillaries.

S: Like arteries do.

T: Uh-huh.

S: And then.

T: So what would you assume comes next?

S: After capillaries?

T: Uh-huh.

S: Can I look? [pause] Venules. So then, you have venules because they go to the capillaries, to the venules, is that how you say it? To the veins

T: And we're inside the lungs now?

S: What?

T: Is where this where the picture is inside the lungs?

S: Well this is here. This is where diffusion happens.

T: **And what would you expect to be passed back and forth?**

S: Oxygen and carbon dioxide. Do you want me to put that?

T: You don't have to.

S: O.k.

**Example of Deep Scaffolding (marked by the bracket)**

- 60.10) If a substance can pass through a membrane, the membrane is permeable to it.  
 S: So, it explains itself. If something is permeable to something, then that thing can pass through the other thing.  
 T: So how would the-  
 S: And if it's impermeable, it can't.
- [ T: **And how would that relate back to the- capillary walls.**  
 S: Well the capillary walls,  
 T: Can you explain?  
 S: Well this is how I learned it.  
 T: Uh-huh  
 S: In the cell, it's made up of these things and then it has these protein thing like this. They're really, really big. And then there's a little substance like oxygen and it can just go through here, but a big substance like sugar which is tons of letters has to go through the protein first.  
 ...  
 T: **And how does, how does that relate to the cell membrane or the capillary?**  
 S: Well if it's too big, if something's too big, to go into the capillary membrane through the capillary membrane, it can't because the capillary membrane is only one cell thick, but then maybe if it has protein, it can. O.k.  
 T: O.k.  
 S: Alright.

**Appendix E. A list of content-free prompts**

1. Any thoughts about that sentence?
2. Any thoughts on that?
3. Anything that comes to your mind?
4. Anything that you want to say about that?
5. Could you explain what you are thinking?
6. Could you tell me what is your thinking of this sentence?
7. Could you explain the concept of the idea discussed on this sentence?
8. Do you have any idea or thoughts about this sentence?
9. Please explain what the sentence says.
10. Please tell me about this sentence here.
11. What are you thinking about?
12. What do you think?
13. What does this sentence mean?
14. What does this sentence tell us?
15. What could you learn from this sentence?

16. What is your thinking on that?
17. What's going on here?
18. What does all this mean to you?
19. Why don't you explain what this sentence means?
20. You look like you are thinking about something. What is it?
21. Any more you want to say about this?
22. Anything else?
23. Anything else you can tell me about it?
24. Anything else you want to say about this sentence?
25. Anything else to say about it?
26. Could you explain what you mean by [a whole sentence]?
27. Could you think of anything else?
28. Could you restate what you just said?
29. Is there something else you want to tell me?
30. What do you mean?
31. What do you mean by [a whole sentence]?
32. What do you mean when you say that?
33. What else can you say?
34. What were you thinking about when you say [a whole sentence]?
35. Why don't you explain this sentence again?
36. Why don't you try again?
37. Could you explain/put this in your own words?
38. Could you explain what's happening in real, simple terms?
39. Could you explain why [a whole sentence]?
40. Could you explain how?
41. Could you elaborate on what you just said?
42. Could you be a little bit more specific?
43. Could you clarify what you just said?
44. Do you have any ideas/thoughts on why that might be the case?
45. I don't quite understand what you said?
46. I'm not sure what are you talking about.
47. What makes you say that/think so?
48. Why do you suppose that it's the case?
49. Why do you think so?
50. Why do you say that?
51. Why?
52. How?
53. What could you learn from this sentence?
54. What do you understand from the sentence?
55. What's the main point of this sentence?
56. Could you connect what you just read with what you have read before?
57. How does it work?
58. How does it do that?
59. How's that?



60. How come?
61. Then, what happens next?
62. What about the other one?
63. What's happening there?
64. What do they look like?
65. What is it related to?
66. How's its structure?
67. What does it do for us?
68. Where can you find it?
69. Where is it?

**Appendix F. Samples of protocols for tutors #1 and #10, at sentences 7 and 85.  
Prompting condition (study 2)**

Tutor #1 Student JL

Sentence 7) The right side pumps blood to the lungs and the left side pumps blood to the other parts of the body.

T: What do those two sentences mean to you?

S: Um, the right side of your heart pumps blood to your lungs and the left side pumps blood to the other parts.

T: What about the first sentence that you read?

S: And it's divided by the septum.

T: Okay. **What do you, what do you think a septum is?**

S: Um, I don't know.

T: Okay, okay. No pressure, no pressure. Okay, all right, read the next three sentences.

Tutor #10 Student BT

Sentence 85) The major function of the blood is to transport nutrients and oxygen to the cells and to carry away carbon dioxide and nitrogenous wastes from the cells.

S: The major function of the blood is to transport nutrients and. . . (pause)

T: Okay.

S: All right, so the major function of blood.

T: Uh huh. **What does blood do?**

S: Carry away. . . carry away. . . carbon dioxide that went from the cells.

T: Okay. So what does that mean? What is that whole statement saying? Like what is it trying to tell you?

S: It's telling me that blood transports nutrients and oxygen to the cells.

T: Uh hmm.

S: And carries away, oxygen, the wastes.

T: Yeah.

## References

- Adams, B. C., Bell, L. C., & Perfetti, C. A. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes*, 20, 307–323.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences*, 4(2), 167–207.
- Azmitia, M., & Perlmutter, M. (1989). Social influences on children's cognition: state of the art and future directions. In H. W. Reese (Ed.), *Advances in child development and behavior*. New York: Academic Press.
- Barwise, J., & Perry, J. (1983). *Situation and attitudes*. Cambridge, MA: MIT Press.
- 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.
- Brinton, D. A., Jarvis, J. Q., & Harris, D. L. (1984). A small-group instruction experiment in medical education. *Journal of Medical Education*, 59, 13–18.
- Brown, A. L., & Palinscar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. Resnick (Ed.), *Cognition and instruction: issues and agendas*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18, 32–42.
- Chan, C. K., Burtis, P. J., Scardamalia, M., & Bereiter, C. (1992). Constructive activity in learning from text. *American Educational Research Journal*, 29, 97–118.
- Chi, M. T. H., Siler, S., & Jeong, H. (In preparation). Can tutors diagnose student's misunderstandings?
- Chi, M. T. H. (2000). Self-explaining: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology*, (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: a practical guide. *Journal of the Learning Sciences*, 6(3), 271–315.
- Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, 10, S33–S49.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, 15, 145–182.
- Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439–477.
- Clark, H. H. (1996). *Using Language*. New York: Cambridge University Press.
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.
- Clark, C. M., Snow, R., & Shavelson, R. J. (1976). Three experiments in learning to teach. *Journal of Teacher Education*, 27, 174–180.
- Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23, 13–20.
- Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, 19, 237–248.
- Cole, M. (Ed.). (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. *Journal of the Learning Sciences*, 7(3 & 4), 387–427.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collins, A., & Stevens, A. (1982). Goals and methods of inquiry teachers. In Glaser, R., (Ed.), *Advances in instructional psychology*, Vol. 2. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cooper, C. R., Ayers-Lopez, S., & Marquis, A. (1982). Children's discourse during peer learning in experimental and naturalistic situations. *Discourse Processes*, 5, 177–191.

- Cox, R., McKendree, J., Tobin, R., Lee, J., & Mayes, T. (In press). Vicarious learning from dialogue and discourse: a controlled comparison. *Instructional Science*.
- Evens, M., Spitkovsky, P. B., Boye, Michael, J., & Rovick, A. (1993). *Synthesizing tutorial dialogues*. Paper presented at the fifteenth annual meeting of the Cognitive Science Society.
- Fantuzzo, J. W., King, J. A., & Heller, L. A. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: a component analysis. *Journal of Educational Psychology*, 84(3), 331–339.
- Fitz-Gibbon, C. T. (1977). *An analysis of the literature of cross-age tutoring*. Washington, DC: National Institute of Education, (ERIC Document Reproduction Service No. ED 148 807).
- Forman, E. A., & Cazden, C. B. (1985). Exploring Vygotskian perspectives in education: the cognitive value of peer interaction. In J. V. Wertsch (Ed.), *Culture, communication and cognition: Vygotskian perspectives*. New York: Cambridge University Press.
- Fox, B. (1991). Cognitive and interactional aspects of correction in tutoring. In P. Goodyear (Ed.), *Teaching knowledge and intelligent tutoring*. Norwood, NJ: Ablex Publishing Corp.
- Graesser, A. C., Person, N., & Magliano, J. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology*, 9, 359–387.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104–137.
- Greeno, J. G., & Middle School Mathematics Through Applications Project Group. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53, 5–26.
- Griffin, P., & Humphrey, F. (1978). Task and talk. In R. Shuy & P. Griffin (Eds.), *The study of children's functional language and education in the early years*. Final report to the Carnegie Corporation of New York. Arlington, VA: Center for Applied Linguistics.
- Hausmann, R. B., & Chi, M.T.H. (In preparation). Does timing of computer generated prompts facilitate learning from self-explanations?
- Hume, G. D., Michael, J. A., Rovick, A. A., & Evens, M. W. (1993). The use of hints as a tutorial tactic. In M. C. Polson (Ed.), *Proceedings of the fifteenth annual conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Juel, C. (1991). Cross-age tutoring between student athletes and at-risk children. *The Reading Teacher*, 45, 178–186.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1–48.
- Lave, J. (1988). *The culture of acquisition and the practice of understanding* (IRL88–0007): Institute for Research on Learning.
- Leinhardt, G. (In press). Instructional explanations: a commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook for research on teaching* (4th Ed.). Washington, DC: American Educational Research Association.
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), 75–95.
- Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. L. (1991). Motivational techniques of expert human tutors: lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction*, 7(3), 197–244.
- McKendree, J., Stenning, K., Mayers, T., Lee, J., & Cox, R. (1998). Why observing a dialogue may benefit learning: the vicarious learner. *Journal of Computer Assisted Learning*, 14, 110–119.
- Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: guided learning by doing. *Cognition and Instruction*, 13(3), 315–372.
- Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: a comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences*, 2(3), 277–306.
- Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: working for cognitive change in school*. Cambridge, MA: Cambridge University Press.

- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117–175.
- Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension*. New York: Holt, Rinehart & Winston.
- Perret-Clermont, A. N., Perret, J. F., & Bell, N. (1991). The social construction of meaning and cognitive activity in elementary school children. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition*. Washington, D.C.: American Psychological Association.
- Poole, D. (1990). Contextualizing IRE in an eighth-grade quiz review. *Linguistics and Education*, 2, 185–211.
- Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal*, 24(1), 13–48.
- Reisner, E. R., Petry, C. A., & Armitage, M. (1990). *A review of programs involving college students as tutors or mentors in grades K-12*. Washington D.C.: Policy Studies Associates, Inc.
- Renkl, A. (1997). Learning from worked-examples: a study on individual differences. *Cognitive Science*, 21(1), 1–29.
- Roschelle, J. (1992). Learning by collaboration: convergent conceptual change. *Journal of the Learning Sciences*, 2, (3 & 4), 235–276
- Rogoff, B. (1986). Adult assistance of children's learning. In T. E. Raphael (Ed.). *The contexts of school-based literacy*. New York: Random House
- Rogoff, B. (1990). *Apprenticeship in thinking: cognitive development in social context*. New York: Oxford Press
- Rose, C. P., Moore, J. D., VanLehn, K., & Allbritton, D. (2000). A comparative evaluation of Socratic versus didactic tutoring. *Learning Research and Development Center Technical Report, #LRDC-BEE-, 1*.
- Scardamalia, M., Bereiter, C., & Steibachm, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science*, 8, 173–190.
- Schliemann, A. D., & Nunes, T. A. (1990). A situated schema of proportionality. *British Journal of Developmental Psychology*, 8, 259–268.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and observers. *Cognitive Psychology*, 21, 11–232.
- Silver, M., & Wilkerson, L. (1991). Effects of tutors with subject expertise on the problem-based tutorial process. *Academic Medicine*, 66(5), 298–300.
- Skon, L., Johnson, D., & Johnson, R. (1981). Cooperative peer interaction effort versus individual effort: effect on the acquisition of cognitive reasoning strategies. *Journal of Educational Psychology*, 73, 83–92.
- Slavin, R. E. (1987). Making Chapter 1 make a difference. *Phi Delta Kappan*, 69(2), 110–119.
- Stenstrom, A. B. (1984). *Questions and responses in English conversation*. Lund: Gleerup.
- Stone, A. (1993). What is missing in the metaphor of scaffolding? In E. Forman, N. Minick, & A. Stone (Eds.), *Contexts for learning*. New York: Oxford University Press.
- Suchman, L. A. (1987). *Plans and situated actions*. New York: Cambridge.
- Swanson, J. H. (1992, April). What does it take to adapt instruction to the individual?: a case study of one-to-one tutoring. Paper presented at the American Educational Research Association meeting, San Francisco.
- Towle, A. (1989). *Modern Biology*. New York: Holt, Rinehart and Winston.
- Tudge, J., & Rogoff, B. (1989). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In M. Bornstein & J. Bruner (Eds.), *Interaction in human development*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- VanLehn, K., Siler, S., Murray, C., & Bagget, W. B. (In press). Human tutoring: why do only some events cause learning? *Cognition and Instruction*.
- von Glaserfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality*. New York: Norton.
- von Glaserfeld, E. (1989). Cognition, construction of knowledge, and teaching. *Synthese*, 80, 121–140.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Webb, G. (1983). The tutorial method. Learning strategies and student participation in tutorials: some problems and suggested solutions. *Programmed Learning and Educational Technology*, 29, 117–121.

- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183–194.
- Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist*, 24, 345–376.
- Woods, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89–100.