International Journal of Artificial Intelligence in Education 19 (2009) 73-102
IOS Press

Learning, Interactional, and Motivational Outcomes in One-to-One Synchronous Computer-mediated versus Face-to-Face Tutoring

Stephanie Ann Siler, *Psychology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
siler@andrew.cmu.edu

Kurt VanLehn, *Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260, USA*
vanlehn@cs.pitt.edu

Abstract. Face-to-face (FTF) human-human tutoring has ranked among the most effective forms of instruction. However, because computer-mediated (CM) tutoring is becoming increasingly common, it is instructive to evaluate its effectiveness relative to face-to-face tutoring. Does the lack of spoken, face-to-face interaction affect learning gains and motivation? In this study, pairs of undergraduate students and tutors worked on physics problems either face-to-face or via a typed chat window. Although face-to-face tutoring took less time, students learned equal amounts in the two conditions. In both conditions, short tutor turns were associated with increased student learning. In both conditions, students who were more active had higher learning gains. Students in the CM condition who gained more produced more words per conversational turn. The same relationship was found in the FTF context only after back-channel feedback was taken out. A more direct measure of student activity, the relative proportion of student-initiated actions in problem-solving, was more strongly associated with student learning in the FTF context, but only for students with higher verbal SAT scores. Of the motivational variables we investigated, only students' ability goals (i.e. wanting to demonstrate one's ability to others) were influenced somewhat differently by the two contexts. These results suggest that although the difference in communication medium changes superficial characteristics of the tutoring such as its duration, most of the important pedagogical characteristics – learning gains, tutorial interaction, the activity measures associated with learning gains, and student motivation – were not affected.

Keywords. Tutoring, computer-mediated, face-to-face, motivation

INTRODUCTION

Face-to-face human-human tutoring is one of the most effective learning contexts (Bloom, 1984; Cohen et al., 1982). Though face-to-face (FTF) tutoring is still the most common tutoring context, computer-mediated (CM) tutoring is becoming increasingly common. This growth is in part due to the recent growth in online learning (Goodyear et al., 2001), including online college courses offered (Conole et al., 2002) and online colleges and universities, which offer online courses exclusively. Additionally, online tutoring services are becoming more widely used. In both online courses and online tutoring, tutors and students may communicate via chat programs. In this paper, we focus specifically on this component of online instruction, one-to-one text-based synchronous CM tutoring.

Both because of its growing availability and because of its potential applications to text-based computer tutoring, the learning and motivational outcomes of one-to-one synchronous CM tutoring are important to study. During CM tutoring, the tutorial dialogue is typed. Participants cannot see each other nor hear the prosodic information of speech, and thus they get less information about each other. Most importantly, they probably get less information about each other's affect. The lack of information might reduce the learning and motivational gains of tutoring. This hypothesis has a great deal of intuitive plausibility but it has not yet been adequately tested.

There have been many recent efforts at adding prosodic speech understanding and other affect sensing to computer tutors (Aist et al., 2002; D'Mello et al., 2005; Forbes-Riley et al., 2006; Johnson et al., 2004; Heylen et al., 2005; Moore et al., 2004; Moreno et al., 2001). These efforts are based on the assumption that such extra information would allow tutoring to be more effective. If it turns out that CM human tutoring is just as effective as FTF human tutoring, then adding prosodic speech and other affect sensing to computer tutoring that does not go beyond what human tutors do naturally may not improve the quality of CM tutoring.

Though there is a large volume of literature investigating CM communication, research on learning from CM compared to FTF tutoring is more limited. The most direct antecedent of the present study is experiment 1 of Litman et al. (2006), which compared one-to-one, synchronous human tutoring of conceptual physics in a CM text-only environment with a spoken, but not face-to-face, environment. The tutor in their study was a former physics professor and experienced tutor. Turn-taking in the CM condition was enforced; that is, one person had to wait until the other had sent their message before they could compose and send theirs. Although the spoken environment produced larger learning gains than the CM environment, the difference was not statistically reliable.[1]

Most other prior work comparing FTF to CM communication has not investigated learning from synchronous one-to-one CM tutoring, but rather:

- Learning from one-to-one tutorial support in *asynchronous* CM environments (e.g. email) embedded within a traditional (FTF) course (e.g. Schpilberg & Hubschman, 2003),
- *Collaborative* learning (rather than tutoring) in synchronous CM environments (e.g. Cummings et al., 1996; van der Meijden & Veenman, 2005; Pilkington & Walker, 2003; Tutty & Klein, 2008),
- Learning from both tutoring and *collaborative* groups in both synchronous and asynchronous environments (e.g. Dennis, 2003),
- *Collaborative* learning in *asynchronous* CM environments (e.g. Ocker & Yaverbaum, 1999; Marttunen & Laurinen, 2001; Cheung & Hew, 2004; Bordia, 1997),
- Learning in online courses (which typically involve both one-to-one *and* group learning) that include: *both* synchronous and asynchronous communication (Warren & Holloman, 2005), or *asynchronous* communication only (e.g. Neuhauser, 2002; Summers et al., 2005), and
- *Productivity* or *process* (not learning) differences from interactions in synchronous or asynchronous CM environments (e.g. Coleman et al., 1999; Condon & Cech, 1996; Jones et al., 2006; Heckman & Annabi, 2005; Lebie et al., 1995).

A large portion of studies comparing CM to FTF communication have investigated collaborative group learning rather than one-to-one tutoring interactions. Generally, comparisons of collaborative

---

[1] In an ANOVA with condition by test phase factorial design, Litman et al. (2006) found no reliable interaction. In an ANCOVA, there was a strong trend toward a reliable difference between conditions, $F(1,31) = 4.04$, $p = .053$, d = 0.74.

group learning in asynchronous environments have found no differences, provided that students were first given time to become familiar with the technology (Bordia, 1997). For example, in Ocker and Yaverbaum's (1999) experiment, groups of graduate students discussed and then produced a report on business case studies over two weeks while interacting either face-to-face or through asynchronous computer conferencing. After this period, their understandings of the case study were individually assessed. There was no difference between the face-to-face and computer-mediated groups on this measure of learning. Similarly, comparisons of synchronous CM groups to face-to-face groups generally report no differences in learning outcomes (e.g. Dennis, 2003; Pilkington & Walker, 2003), though there are some exceptions. For example, Tutty and Klein (2008) found that FTF students learned basic computer skills better than their CM counterparts.

Consistent with the majority of findings for groups, in one study of individual learning from tutoring in an asynchronous environment (Schpilberg & Hubschman, 2003), there was no difference in high school students' math grades after they were given either face-to-face help (during teachers' office hours) or help via email (non-restricted hours). Finally, several studies comparing online courses to traditional FTF courses have found no significant differences in learning outcomes (Neuhauser, 2002; Summers et al., 2005; Warren & Holloman, 2005).

CM groups may have some advantages in what they produce *during* group interactions, including collaborative decision-making, group work, and writing (Luppicini, 2007; Benbunan-Fich & Hiltz, 1999). For example, Cummings et al. (1996) found that CM groups produced essays with "higher integrative complexity" than FTF groups. However, this advantage may not hold for younger students. For example, van der Meijden and Veenman (2005) found that dyads of sixth-grade students in a FTF condition had higher group performance scores on math problems than those in a synchronous CM condition. It is important to note, however, that greater group performance does not necessarily translate into greater individual learning. For example, Tutty and Klein (2008) found that, though group project performance was better for students in synchronous CM dyads, FTF students demonstrated greater individual learning.

In the majority of reported studies, no learning outcome differences were found[2]. However, given the many differences between asynchronous and synchronous CM communication, between group learning and one-to-one tutoring, and between tasks, it is difficult to conclude from these results whether or not learning differences would exist for one-to-one synchronous CM and FTF tutoring. This suggests considering theoretical arguments for differences between CM and FTF tutoring. That is, how would we expect learning outcomes from synchronous CM tutoring to compare to FTF tutoring? The next three subsections review predictions stemming from differences in, respectively, the media (speech vs. text), interactivity, and motivational differences between CM and FTF tutoring.

Potential differences in learning due to differences in media

In FTF tutoring, tutors and students communicate by speaking and listening to each other, whereas in CM tutoring, they type and read each other's messages. This section considers learning differences due to listening vs. reading and speaking vs. typing.

---

[2] In all of these studies except for Tutty and Klein (2008), sample sizes did not exceed 30 per group. For a medium effect size, about 50 per group is necessary for an 80% power level at an alpha of .05. Thus, it is possible that significant results were not found due to sample size. However, the consistency of null results makes this less likely.

Does listening or reading lead to better comprehension? The answer appears to depend on many factors including age. For young children, some studies have shown an advantage of listening over reading. For example, Fletcher and Pumfrey (1988) found an advantage of listening over reading silently for 7 and 8-year-old children. For older children and adults, however, an advantage of reading over listening is a more common finding (e.g. Neville & Pugh, 1974; Rubin et al., 2000; Diakidoy et al., 2004). Thus, for older children and adults, including college-aged students who typically participate in studies comparing CM to FTF learning, including the present one, CM tutoring may give students a learning advantage on the comprehension end of communicating.

One potential contributing factor to an advantage of reading is that a permanent record exists that may be re-read (Hara et al., 2000). Re-reading may help students develop a better understanding and resolve previous misunderstandings due to initial reading errors. The ability to re-read messages may be particularly beneficial for students who are less skilled readers and thus less likely to understand a message initially. Re-reading may also help to consolidate information into long-term memory by repeated exposure to the same information. Rubin, Hafer, and Arata (2000) proposed the possibility that reading may have an advantage over listening because it requires more mental effort to process. If this additional mental effort translates into deeper processing of the reading material itself, then this may also contribute to an advantage of reading over listening.

Another difference between the two modes of communication is that generally, less time is acceptable between turns in FTF than in CM communication (Clark & Brennan, 1991). Thus, both tutor and student have less time to think about and produce a response in a FTF context. This decreased time may also lead to simultaneous encoding of the statement and production of a response in FTF communication (Condon & Cech, 2001), which would result in a heavier load on working memory. This may negatively impact both the students' comprehension of the message and the quality of the tutor's response in FTF communication, especially for students and tutors with poorer verbal processing skills. Longer turn lengths may be especially problematic for these students. On the other hand, more time per turn in the CM context may allow students the opportunity to re-read messages before responding, which may help students with lower reading comprehension in particular. This time difference suggests that FTF tutoring may be less effective than CM tutoring.

One advantage of FTF over CM tutoring is, because people generally speak faster than they can type (e.g. Olaniran, 1996), more information can be conveyed in the same period of time in FTF than CM communication (CMC). For this reason, in the FTF context, students may learn more in the same period of time, so FTF tutoring should be more time efficient. This was found in the Litman et al. (2006) experiment, where students in the spoken condition completed training in about half the time as students in the CM condition. Similarly, as reported by Bordia (1997) in a review of group learning, "a number of studies reported longer times taken by CMC groups to complete the allotted task" (p. 102).

Thus, from a media perspective alone, it appears that CM tutoring should be more effective than FTF tutoring, albeit considerably slower.

Possible differences in learning due to differences in interactivity

Even a cursory examination of transcripts shows that FTF dialogues are more interactive than CM dialogues, with more conversational turns for discussions of the same topics (e.g. Litman et al., 2006). In FTF dialogues, participants may speak simultaneously, which allows them to "back-channel" their understanding (Clark, 1996), to interrupt, or to complete an utterance started by the other participant.

In CM dialogues, including those of Litman et al. (2006) and the present study, the text typed by a participant is only sent to and seen by the other participant after the Enter key is pressed. Thus, CM participants can neither interrupt, back-channel, nor complete each other's utterances within a given turn.

Tutorial dialogues may vary between a highly didactic, lecturing style to a highly interactive, collaborative style. This difference can be measured with both surface features of the dialogues (e.g. word counts) and deeper codings. With respect to surface features, several correlations with learning have been found in studies of CM human-to-human tutoring:

- The number of student words per turn positively correlates with gains. That is, tutorial dialogues with longer student turns result in larger gains (Litman et al., 2006; Rosé et al., 2001; see also Core et al., 2003, for a related measure).
- The number of tutor words per turn positively correlates with gains. Longer tutor turns resulted in larger gains (Litman et al., 2006).

Similar correlations were not found with spoken tutoring (Litman et al., 2006). However, if we take the correlations at face value and note that both student and tutor turns are longer in CM than spoken tutoring (Litman et al., 2006), then the correlations predict an advantage for CM over FTF tutoring.

Deeper coding of tutorial dialogues suggests a resolution of this apparent contradiction, where both tutor and student words per turn are positively related to learning. Chi, Siler Jeong, Yamauchi, and Hausmann (2001) coded FTF tutorial dialogues using deeper categories for tutor turns (explanation, feedback, reading aloud, meta-cognitive, answering questions, asking content questions, scaffolding, and asking comprehension questions) and for student turns (reading aloud, self-explaining, asking questions, answering questions, responding to the tutor's scaffolding prompts, and reflecting). Using stepwise regression, Chi et al. found that only the number of student turns coded as "reflection", which were comprehension-monitoring statements, positively correlated with deep learning gains. For shallow learning gains, both the number of student turns coded as "responses to tutor scaffolding" and the number of tutor turns coded as "explanation" positively correlated with gains. Their interpretation is that the tutor giving frequent, long explanations produced only shallow learning; this would explain the second surface feature finding listed above, that the length of tutor turns sometimes correlates with learning gains. For deep learning, long, frequent student reflection turns produce learning gains, which is consistent with the first surface feature, that longer student turns often correlated positively with learning gains. This is clearly the more important correlation, as it appears in all analyses of tutorial dialogue and it relates to deep learning.

Thus, we hypothesize that it is primarily student activity or engagement that causes learning, regardless of tutoring context. In the present study, we looked at measures of student activity such as the number of student words, percent of words given by the student, and student turn length. Whichever context allows for higher levels of productive student engagement should produce higher learning gains.

Possible motivational effects of CM vs. FTF tutoring

In addition to favorable learning gains compared to other types of instruction, face-to-face tutoring has also been found to improve motivation-related factors such as students' interest in the tutored subject (e.g. Cohen et al., 1982). Motivation is a large, amorphous variable composed of many social and cognitive processes. Some relevant aspects are:

- Students' confidence in their ability to successfully perform a learning task, or their *efficacy beliefs*, which predicts their persistence in the task (Vollmeyer & Rheinberg, 2000) and their performances in math and science (Wigfield & Guthrie, 1997; Randhawa et al., 1993; Zimmerman et al., 1992).
- Students' motivations may be *intrinsic* and/or *extrinsic* to the task. Intrinsic motivations for performing a task include finding the task interesting or enjoyable and being curious about the task. Extrinsic motivations include objective rewards (e.g. grades; pay), the approval of teachers and peers, etc.
- *Achievement goals* also define purposes students have for engaging in achievement-related behavior; two types of achievement goals commonly studied are learning goals and performance goals (e.g. Dweck & Leggett, 1988). Students with learning goals seek challenges in an effort to *gain* competence in a domain, whereas students with performance goals want to *demonstrate* their competence in some way. Learning goals are often associated with deeper learning, whereas performance goals, including those that are normative or competitive (where students wish to do better than others) or outcome performance goals, which are "simply focused on obtaining positive outcomes" are generally associated with positive outcomes, albeit more shallow learning (Grant & Dweck, 2003). However, some types of performance goals (e.g. ability goals, in which students wish to be seen by others as competent) have been found to be detrimental to either type of learning when a set-back is encountered (Grant & Dweck, 2003).

It is plausible that tutors can influence motivation. For instance, tutors may influence students' beliefs about their efficacy (Lepper et al., 1993), they may embody extrinsic motivation, or they may encourage students to learn rather than perform.

The FTF vs. CM manipulation may differentially affect students' motivation. In FTF communication more channels of information are available to participants that are not available in CM communication, including visual information such as each other's facial expressions, gestures, general appearance and auditory information, such as a speaker's tone of voice, phrase inflections, and emphasis on particular words or phrases. Perhaps for these reasons, people view FTF communication as more personal than CM communication (Lebie et al., 1995; Coleman et al., 1999) and generally say they prefer FTF over CM communication (e.g. Adrianson, 2001). Furthermore, tutors have more potential ways to influence their students' affect in FTF than CM tutoring (Tu, 2000). For example, tutors can adopt a softer tone of voice or a sympathetic facial expression. Therefore, we expect inter-personal motivational goals in particular to be influenced more in the FTF context than in the CM context. However, again, little is known about whether and how tutoring contexts differ in their influence on students' confidence, intrinsic motivation, and achievement goals.

In the present study, we compared student learning and changes in motivational beliefs in one-to-one FTF and synchronous CM tutoring (Microsoft NetMeeting's chat window and whiteboard were used) during physics problem-solving and investigated features correlated with learning in both contexts[3]. Because novice or peer tutors (rather than highly experienced, professional teachers) often serve as tutors in FTF educational settings (Cohen et al., 1982), and adult novice tutors (e.g. graduate research assistants or advanced undergraduates) may serve as online tutors (e.g. Anderson, 2002), novice tutors were used in this study. We expected that, regardless of context, measures of student productivity would be correlated with learning outcomes. We did not expect to find any differences in learning, and therefore hypothesized that the two contexts would allow for equal levels of productive student engagement. However, we did expect learning in the FTF context to be more efficient than in the CM context. Finally, we expected that inter-personal motivational goals in particular would be influenced more in the FTF context than in the CM context. These findings will have important implications for FTF and synchronous CM human tutoring as well as computer tutoring systems that incorporate text-based natural language.

METHODS

Participants

Students: Undergraduates who had taken an introductory physics class in high school and/or college served as students (or tutees). The FTF condition consisted of 20 students (9 men and 11 women). The CM condition consisted of 20 students (14 men and 6 women). Students who were recruited through the psychology subject pool received course credit and those who were not received payment of $7/hour for their participation.

Tutors: Advanced undergraduate or graduate students in physics or engineering who demonstrated knowledge of the task domain and at least some tutoring experience, including informal tutoring experience (e.g. tutoring a friend), served as tutors. The FTF condition consisted of 20 tutors (16 men and 4 women), and the CM condition consisted of 20 tutors (also 16 men and 4 women). Tutors received $10 per hour for their participation.

Materials

Materials for students included the following:

1. Background questionnaire: Assessed factors that may influence student learning, including SAT math and verbal scores, physics classes taken and grades received in those classes.
2. Physics pre-test: Assessed students' knowledge of the physics concepts relevant in the tutoring questions. The physics pre-test included:
    a. Definitions of terms that assessed students' understanding of the individual physics concepts (e.g. distance, displacement, speed, velocity, acceleration),
    b. Short-answer questions that assessed their knowledge of the relationships between these concepts (i.e. kinematics equations, Newton's second law).
3. Pre-tutoring motivation questionnaire: The motivation questionnaire given to students was the Motivation for Reading Questionnaire (Wigfield & Guthrie, 1997), adapted to college students

[3] The FTF condition may be considered the "control" acting as a comparison for the CM condition. A (no-intervention) control condition was not necessary in this study because its pre/post design allows for the assessment of learning across the intervention.

and to the tutoring domain, physics (excluding questions that could not be sensibly adapted to the college-aged population and subject-domain of this study, for example, "My parents often tell me what a good job I am doing in reading"; "I make pictures in my mind when I read"; "I like mysteries"). This questionnaire included items that targeted various aspects of students' motivation, including their:

a. Self-confidence or efficacy in physics,
b. Desire for challenge in physics (consequence of holding a learning goal),
c. Curiosity about physics (intrinsic motivation),
d. Perceptions of the importance of being good in physics (learning goal),
e. Need for validation or recognition of ability from others (type of performance goal: ability goal),
f. Competitiveness in physics (a normative performance goal),
g. Beliefs about grades in physics (type of performance goal),
h. Compliance, or willingness to work hard in physics (or inversely, to avoid work).

We added the item "I find physics generally interesting", another aspect of intrinsic motivation (interest).

Because the motivation questionnaire was modified for a different age group (college students rather than elementary school children) and a different domain (physics rather than reading), an exploratory factor analysis was performed on students' responses to the pre-tutoring motivation questionnaire to determine the different underlying dimensions of student motivation assessed by the questionnaire for this subject population. Students indicated their agreement with items on the motivation questionnaires on a 7-point Likert scale (1 – strongly disagree; 4 – no opinion; to 7 – strongly agree).

4. Mid-tutoring motivation questionnaire: Identical to the pre-tutoring motivation questionnaire but with fewer statements (to minimize interruption to the tutoring session).
5. Mid-tutoring physics test: Included short-answer questions that assessed students' knowledge of the concepts and equations that were relevant in the solution to the question discussed in the final tutoring segment. Questions targeting concepts that were not addressed in the final tutoring segment were also included to prevent priming of only relevant concepts for the final question. To minimize the disruption to tutoring, definition questions were not included on the physics mid-test.
6. Post-tutoring motivation questionnaire: Identical to the pre-tutoring motivation questionnaire.
7. Physics post-test, which included the following:
   a. Terms assessing students' knowledge of the individual concepts relevant in the tutoring questions (same as physics pre-test),
   b. Short-answer questions requiring the application of various kinematics equations and Newton's second law,
   c. One near-transfer question that was isomorphic to the final question discussed in tutoring (different surface features but same deep structure).

Materials for tutors included the following:

1. Background questionnaire: Tutors completed a background questionnaire that assessed factors that may influence tutoring effectiveness, including their prior tutoring experience and tutor training.
2. Physics knowledge assessment: To ensure that tutors had adequate knowledge of the tutoring topics, tutors completed a physics knowledge assessment that consisted of questions assessing

the concepts discussed in the questions for the tutoring session. Those who did not pass this assessment did not serve as tutors.

3. Instructions for tutoring: Instructions about the tutoring session procedure.
4. Questions for tutoring, by segment of the tutoring session:
   - Segment 1: Two questions on trigonometric relations,
   - Segment 2: Two questions: (Q1) kinematics relation; (Q2) Newton's second law, definition of average acceleration,
   - Segment 3: One question: two-dimensional kinematics, Newton's second law,
   - Segment 4: One question: trigonometric relation, two-dimensional kinematics equations, Newton's second law, and definition of average acceleration. This was the most complex question covered in tutoring.

   To control for the specific content that tutors covered in the tutoring session, they were given the "ideal" solutions to each of the questions, which included a breakdown of all of the solution steps that were considered relevant in the solution to that question.
5. Tutor mid-tutoring motivation questionnaire: This questionnaire was identical to the mid-tutoring motivation questionnaire given to the students, but with brief instructions for tutors. The purpose of this questionnaire was to assess the accuracy of tutors' assessments of their students' overall motivation, confidence, and different dimensions of motivation. Tutors were instructed to predict their students' responses on the 7-point Likert scale for all items on the mid-tutoring motivation questionnaire.
6. Tutor mid-test (this will not be discussed here[4]).

Design

This study incorporated a 2 x 2 between-subjects design. The two independent variables were prior experience (Same or Different), and tutoring context (FTF and CM). In the Same conditions, each tutor was paired with a student and worked with that student in all four segments of the tutoring session, whereas in the Different condition, the students rotated among tutors so that tutors worked with a different student in each of the four tutoring segments. In this paper we compare *only* the FTF Same and CM Same conditions, which are more typical of real-world tutoring, where tutors and students work with each other for the duration of a tutoring session. We will refer to these simply as the FTF and CM conditions.

Procedure

About one week before the tutoring session, all students first completed the background questionnaire, then the physics pre-test. The experimental procedure is given in Table 1. Students were assigned to conditions so that pre-test scores were counter-balanced. All tutors completed the background questionnaire, then the physics knowledge assessment. Afterwards, tutors were given the instructions for tutoring, the questions for tutoring and the corresponding ideal answers and figures for each

---

[4] These materials are relevant for the discussion of the effect of tutoring context on the accuracy of tutors' assessments of their students, discussed in Siler and VanLehn (2003). These results will not be discussed here. Please refer to that paper for more information.

question. They were instructed to become familiar with the questions and answers before returning for the tutoring session.

Table 1
Experimental procedure

Timeline	Student activity	Tutor activity
Before the tutoring session	(1) Background questionnaire (2) Physics pre-test	(1) Background questionnaire (2) Physics knowledge assessment (3) Given instructions for tutoring (4) Given questions for tutoring session
Immediately before the tutoring session	(3) Pre-tutoring motivation questionnaire	(no corresponding tutor activity)
During the tutoring session	Segment 1 (15-minute time limit) Segment 2 (30-minute time limit) Segment 3 (15-minute time limit)	
	(4) Mid-tutoring motivation questionnaire (5) Physics mid-test	(5) Tutor mid-tutoring motivation questionnaire (6) Tutor mid-test (Likert ranking of students' general competence)
	Segment 4 (no time limit)	
After the tutoring session	(6) Post-tutoring motivation questionnaire (7) Physics post-test	(tutor is finished)

Approximately one week after their respective first sessions, students and tutors returned for the tutoring session. Students first completed the pre-tutoring motivation questionnaire. All tutors were given hardcopies of the questions that were to be discussed in the tutoring session and the solutions to each question, which they were permitted to refer to during the tutoring session. Tutors in the FTF condition were told to not allow their students to see the problem solutions. After students completed the pre-tutoring motivation questionnaire, each tutor-student pair was introduced face-to-face.

In the FTF condition, the tutor and student sat together at a desk in one room. All tutor-student pairs were given pens and blank paper and told that they could draw. In the CM condition, each tutor sat at a computer in one room and the student sat at another computer in a different room. The questions and corresponding figures were saved on the Desktops of all tutors' computers. Tutors and students in the CM condition were separately instructed on the use of Microsoft NetMeeting. They communicated with each other using this program's chat window and whiteboard. Turn-taking in the chat windows was not enforced; that is, tutors and students were able to send text messages to each other at any time. A message was viewed by the receiver only after the composer sent it by hitting the Enter key. All text that had been sent within a tutoring segment could be viewed by either the tutor or student. They were told that they could draw in the whiteboard during the tutoring session. Tutors and students informed the experimenter when they were comfortable enough using NetMeeting to begin the first segment.

In the FTF condition, the starting time of each of the four tutoring segments was when the tutor began reading the student the first question for that segment. In the CM condition, tutors pasted the figures for each question into the drawing window, and then pasted the questions into the chat window. The start time for the CM segments was when the tutors sent the first question for that segment to the student. The experimenter was in the hallway outside the rooms to answer any questions or address any problems that arose during the tutoring segments. Each of the first three tutoring segments had a time limit (15 minutes for the first segment, 30 minutes for the second, and 15 minutes for the third). There was no time limit for the fourth and final tutoring segment. For the first three tutoring segments, after the time limit had expired, tutoring was stopped if the tutor and student had not finished discussing the questions for that segment. In the CM condition, the chat and drawing windows were saved and closed, and new windows were opened. In the FTF condition, any drawings made were taken by the experimenter. All FTF tutoring segments were audio-taped and transcribed.

After the third tutoring segment, tutors and students were moved to separate rooms to complete the mid-tutoring motivation questionnaire and physics mid-test. All students first completed the mid-tutoring motivation questionnaire and then completed the physics mid-test. When both the tutor and the student finished the mid-tutoring motivation questionnaire and physics mid-test, they began the final tutoring segment.

When the tutor-student pair indicated that they were finished with the final problem, the student completed the post-tutoring motivation questionnaire and finally the physics post-test. Students who were not recruited from the psychology subject pool were paid and all were debriefed. All tutors were paid and debriefed.

RESULTS

There were no significant differences between FTF and CM conditions on student subject variables that may have influenced learning from tutoring, including physics pre-test scores, math SAT scores, verbal SAT scores, or total SAT scores. There were also no significant differences between tutoring conditions for tutor variables, including tutor pre-test, math SAT, and verbal SAT.

Student Learning by Context

Total post-test score: The total physics post-test score consisted of definitions, short-answer problems, and a problem isomorphic to the final question discussed in the tutoring session. There was no significant difference in post-test scores between the FTF and CM conditions (Table 2), $F(1, 38) = 0.47$, $p = .50$.

Table 2
Mean unadjusted and adjusted total post-test scores (standard deviation) by condition

	Total post-test score:		
Condition	Unadjusted	Adjusted for pre-test	Adjusted for mid-test
FTF	42.88(17.57)	40.02 (3.67)	41.46 (2.67)
CM	38.65 (21.30)	41.51 (3.67)	40.07 (2.67)

When student pre-test was factored out in an ANCOVA, there was still no difference between conditions in total post-test score, $F(1, 37) = 0.08$, $p = .78$[5]. There were no significant differences between conditions for any of the subparts of the post-test (i.e. term definitions, short-answer questions, or the isomorphic question). Thus, knowledge gains were similar for the two conditions. To see if there were any learning differences across the final tutoring segment (which had no time limit), mid-test scores were factored out of total post-test in an ANCOVA. There were again no differences, $F(1, 37) = 0.14$, $p = .72$. For just the definitional scores, although there was a significant pre-to-post-test increase, $F(1, 38) = 24.20$, $p < .001$, there was no time by condition interaction, $F(1, 38) = 0.49$, $p = .49$. Thus, students in the FTF and CM context gained similar amounts on this measure of conceptual learning. There was also a significant gain on the short-answer problems, $F(1, 38) = 43.55$, $p < .001$. But again, there was no difference between conditions in the amount of gain, $F(1, 38) = 0.78$, $p = .38$. Thus, on both of these measures of learning, although there were gains across the tutoring session, there were no context differences.

Differences in tutoring efficiency

Next, we looked at tutoring efficiency. Because the final segment was the only one without an imposed time limit, we chose to assess efficiency across this segment[6]. Time efficiency was defined as the ratio of total post-test score adjusted for mid-test score to time on task.

Table 3
Efficiency measures for the final tutoring segment: means (and standard deviations)

Condition	Segment 4 time (minutes)	Efficiency [a]
FTF	29.35 (14.80)	18.42 (12.00)
CM	49.04 (15.55)	9.17 (5.49)

[a] Efficiency values were multiplied by 1,000.

As in earlier studies (e.g. Litman et al., 2006), times were longer in the CM condition (about 1.7 times longer), a highly significant difference (Table 3), $F(1, 38) = 16.81$, $p < .001$. Correspondingly, time efficiency was greater in the FTF condition, $F(1, 38) = 9.83$, $p = .003$[7].

How would this compare to time efficiency differences if participants in the CM context had used speech-to-text technology instead of typing messages? In this study, only the times that messages were sent were recorded, and not the time typing of the message began. Thus, we cannot get an exact measure of typing times. However, we can get an approximation of the time spent typing. Given the total number of words exchanged and using an average typing rate of 40 wpm, we first approximated segment 4 times if tutors and students had spoken to each other instead of typing. An important note

[5] There was no aptitude-treatment interaction: the student pre-test by condition interaction was not significant, $F(1, 36) = 0.87$, $p = .36$.

[6] According to Bordia (1997, p. 112): “Providing a limited amount of time puts pressure on subjects to somehow complete the task. This may distort the experimental manipulation and may introduce unintended effects.” For these reasons, we chose to target the final tutoring segment.

[7] The variance in time efficiency was significantly lower in the CM than in the FTF context. However, this is likely just a mathematical consequence of dividing adjusted post-test scores by a larger number (segment 4 time).

is that this assumes that tutors and students did not type simultaneously, though they almost certainly did at times. Therefore, almost certainly, more time was removed from the CM condition than their actual typing time, and this can be considered an absolute best-case scenario. After converting typing to speaking rate[8], there was no longer a time difference, $F(1, 38) = 0.17$, $p = .68$. Efficiencies were calculated as the same total post-test scores with mid-test scores factored out, divided by the new estimates of time. There was no longer a difference in efficiency (Table 4), $F(1, 38) = 0.29$, $p = .59$. However, the assumption that CM students would have learned the same amount in this reduced time is unlikely. This analysis does not show that there would be no time or efficiency differences had the participants in the CM context spoken rather than typed; rather, it merely does not rule out the possibility that times and efficiencies in a CM context could approach those in a FTF context.

Table 4
Efficiency measures for the final tutoring segment: typed converted to spoken: means (and standard deviations)

Condition	Segment 4 time (minutes)	Efficiency [a]
FTF	29.35 (14.80)	18.42 (12.00)
CM (converted)	31.18 (13.40)	16.38 (11.91)

[a] Efficiency values were multiplied by 1,000.

Motivational Outcomes

The different dimensions of student motivation that were addressed by the motivation questionnaire were found using exploratory principal components factor analysis with the Varimax rotation method. This was performed on all 80 students' (i.e. students in both contexts and experience conditions) responses to the 26 statements on the pre-tutoring motivation questionnaire. All 80 students were used to maximize the power of the factor analysis. Factorability was acceptable (KMO = .78). Seven component factors with eigenvalues greater than one emerged from this analysis. These seven factors accounted for over 73% of the total variance. The first component factor, which comprised 33.75% of the total variance, tapped strongly into a factor related to confidence in one's physics ability. This factor loaded most highly (above the criterion .60) on the following statements:

- I know I will do well in the tutoring session and on the physics post-test today
- I am generally good in physics
- I am generally better at solving physics problems than most people I know
- I like hard, challenging physics problems
- If I find it interesting, I can understand difficult physics concepts
- I am usually able to learn difficult physics concepts

The second component factor, which comprised 11.34% of the total variance, loaded most highly on four statements related to competition with other students in physics, a normative performance goal:

---

[8] An average speaking rate of 150 wpm (Wald & Bain, 2008) and an average typing rate of 40 wpm (Karat et al., 1999) were used in the conversion: Time (if spoken) = Time (CM actual) – [total words*(1/40 – 1/150)].

- I try (or tried) to do better on physics exams than my friends
- I would like to get the highest scores on physics tests[9]
- I would like to finish physics problems before other students in my class
- I would like being the only one who knew an answer to a physics question

The third component factor, which comprised 6.63% of the total variance, loaded most highly on four statements related to physics curiosity, a measure of intrinsic motivation:

- If the physics topic is interesting, I might investigate it more later
- I read to learn about physics topics that interest me
- I like to read about physics in my free time
- I only read about physics when I have to (e.g. for a class) (negative loading)

The fourth component factor, which comprised 6.50% of the total variance, loaded most highly (above .60) on two statements related to desire for challenge in physics, a consequence of holding a learning goal:

- I like when physics questions make me think
- Even if the physics topic is difficult, if I find it interesting I will try hard to understand it

The fifth component factor, 5.82% of the total variance, loaded most highly on two statements that assessed students' need for validation of their physics ability from others, another type of performance goal. As stated previously, this goal has been called an ability performance goal in the motivation literature (e.g. Grant & Dweck, 2003). We will refer to this factor as "ability goal" henceforth. Because it taps into a wish for others to consider them competent, this can also be considered a type of inter-personal goal:

- I would like my physics instructor to think I am good in physics
- I like (or would like) to get compliments for my physics ability

The sixth component factor, 5.14% of the total variance, loaded most highly on three statements that can be considered related to students' desire to get good grades in their physics class, an outcome performance goal:

- In comparison to other subjects, it is (or was) very important for me to do well in physics
- I study (or studied) mainly to improve my grade in the class
- I do (or did) as little work in my physics class as possible (negative loading)

No statements loaded onto the seventh and final component factor above .60. There were five statements (including "I find physics generally interesting") that did not load on any factor above the criterion .60.

Changes in motivational beliefs by context

When considering students' overall average motivation scores (the sum of Likert scores across items divided by the number of items), there was no main effect of time, nor was there an overall context by

---

[9] The underlined questions were not included on the mid-tutoring motivation questionnaire.

time (pre/post-tutoring) interaction (Table 5). Thus, overall average motivation scores were stable across time for students in both contexts.

Table 5
Motivational measure means by time and context

Measure	Context	Time: Pre-tutoring	Time: Post-tutoring	Time	Interaction
Overall average score (26 items)	FTF CM	4.31 (0.85) 4.08 (0.85)	4.33 (0.75) 4.03 (0.90)	(n.s.)	(n.s.)
F1: Confidence	FTF CM	4.17 (1.28) 4.13 (1.11)	4.22 (1.07) 4.07 (1.27)	(n.s.)	(n.s.)
F2: Competition	FTF CM	4.73 (1.21) 4.50 (1.26)	4.75 (0.96) 4.55 (1.20)	(n.s.)	(n.s.)
F3: Curiosity	FTF CM	3.11 (1.49) 3.04 (1.08)	2.83 (1.57) 2.90 (1.22)	$p = .07$	(n.s.)
F4: Challenge	FTF CM	5.15 (1.28) 4.75 (1.22)	4.83 (1.42) 4.45 (1.20)	$p = .07$	(n.s.)
F5: Ability goal	FTF CM	4.60 (1.43) 3.83 (1.17)	4.65 (1.17) 4.40 (1.26)	$p = .03$	$p = .07$
F6: Grades	FTF CM	4.68 (1.11) 4.53 (1.46)	4.88 (1.02) 4.38 (1.55)	(n.s.)	(n.s.)
Interest in physics	FTF CM	4.85 (1.42) 4.70 (1.56)	4.60 (1.54) 4.10 (1.52)	$p = .02$	(n.s.)

However, when looking at the individual factors[10], there were some significant changes. Students' curiosity scores (Factor 3) showed a strong trend of declining from pre- to post-tutoring, $F(1, 38) = 3.37$, $p = .07$, and this decline was similar across tutoring contexts (Table 5). Similarly, students' challenge scores also showed a strong trend of declining, $F(1, 38) = 3.59$, $p = .07$, again with similar decreases across contexts. The only motivational factor that showed a strong trend of a time by context interaction was ability goal (Factor 5), $F(1, 38) = 3.54$, $p = .07$. Again, these two items were "I would like my physics instructor to think I am good in physics" and "I like (or would like) to get compliments for my physics ability". FTF students' scores were stable across time whereas CM students' scores increased from pre- to post-tutoring. Finally, students' agreement with the statement "I find physics generally interesting" decreased significantly from the beginning of the tutoring session, $F(1, 38) = 6.45$, $p = .02$. Though decreases were larger for CM students than for FTF students, this difference was not significant ($p = .30$).

In summary, though there were pre-to-post tutoring changes for most of the motivational measures, there was only one time by context interaction that showed a trend approaching significance, for the ability goal factor (validation), where CM students' scores increased with time while FTF students' scores were stable. As stated earlier, ability performance goals are generally found to be negatively associated with performance. If this relationship between ability goal and performance is causal, then this result favors learning in the FTF context.

The only factor for which there was a context difference in the accuracy of *tutors' predictions* of their students' mid-tutoring responses on the motivation questionnaire was the challenge factor. That

[10] Individual factor scores were averages across student responses to all items comprising that factor.

is, this was the only factor for which there was a significant context by tutor prediction interaction, $F(2, 37) = 3.77$, $p = .03$. In the FTF context, there was a significant positive relationship between tutors' predictions of students' challenge scores (r = +.65, p = .002). However, in the CM context, this correlation was not significant (r = +.09, p = .70). Because only one of the six factors differed, these results suggest that there is only a slight advantage in the FTF context in obtaining affective information, which parallels the results for student motivational gains just discussed.

How did students learn?

There were no learning differences between the FTF and CM conditions, though the FTF condition was more time-efficient. But were the features associated with learning similar across the two contexts, as predicted? We will first look at context differences for student activity, tutor activity, and tutor-student interactivity. Then, to address this question, we will look at the relationship between each of these types of activity and student learning in each context.

Student and tutor activity by context

Consistent with prior findings (e.g. Litman et al., 2006), there were large differences between the spoken and typed communication modalities. In the final tutoring segment (Table 6), compared to the CM context, the FTF context had more student words, more tutor words, more turns and less time per turn. These differences are expected due to the difference between speaking vs. typing.

Table 6
Student and tutor activity in the final tutoring segment: means (and standard deviations)

Measure	FTF	CM	$F(1,33)$, p
Student words	616.20 (407.64)	177 (116.49)	21.11, $p < .001$
Tutor words	1965.80 (1275.91)	796.70 (299.49)	15.78, $p < .001$
Turns	224.67 (119.35)	80.80 (46.18)	24.40, $p < .001$
Time (sec) per turn	10.19 (10.61)	47.27 (29.01)	22.15, $p < .001$
Student % words	24.62 (11.61)	18.01 (9.53)	3.42, $p = .07$
Mean student words per turn	5.82 (3.89)	4.48 (1.71)	1.90, $p = .18$
Mean tutor words per turn	23.11 (25.02)	23.68 (11.11)	0.01, $p = .93$

In terms of relative word production, students in the FTF context showed a strong trend of being more active than students in the CM context. That is, students in the FTF condition produced a higher percentage of total words in the final tutoring session than students in the CM condition (Table 6). Thus, students in the FTF context were more active (at least overtly) than students in the CM context. However, there was no difference between conditions in mean student or tutor turn length, which may give the impression that the FTF and CM contexts were similarly interactive. However, FTF discussions generated almost three times as many turns, showing that the FTF context was actually more interactive.

One would expect tutoring to be more adaptive in the FTF condition than the CM condition due to the increased information contained in speech, gesture, etc., of FTF tutoring. Consistent with this hypothesis, the variance in number of student and tutor words was significantly greater in the FTF condition ($F = 6.46$, $p = .02$; $F = 12.81$, $p = .001$, respectively), as was the variance in number of turns

(F = 14.17, p = .001)[11]. Because a greater variance in tutor words, per se, is not evidence of more adaptive tutoring in the FTF context, additional evidence is needed to support this hypothesis. If tutor words were adaptive in the FTF context, one may expect the number of tutor words to be related to student mid-test score only in the FTF context; specifically, one may expect a negative relationship if tutors tend to say more to students who are less competent. In fact, there was a significant context by student mid-test score interaction, $F(1, 31) = 5.25$, $p = .03$. In the FTF context, there was a significant negative relationship between student mid-test score and tutor words, r = -.54, $p = .04$. However, in the CM context there was no significant relationship, r = -.29, $p = .22$. Thus, these results are consistent with the hypothesis that the greater variances in the FTF context may be attributable to more adaptive tutoring[12]. The variance in time per turn was greater in the CM condition (F = 9.23, p = .005), which could be a consequence of a greater variance in typing than in speaking speed.

Activity and student learning

Let us turn now to the relationship between different types of activity in the final tutoring segment and student learning in that segment (total post-test score, with mid-test score factored out using ANCOVA), starting with the CM condition (Table 7, last column). There were significant positive pair-wise relationships with student words and mean student turn length – students who typed more words learned more. There was a significant negative relationship between tutor turn length and student learning – when tutors typed fewer words, students learned more. Given these two relationships, one would expect percent student words to be strongly positively correlated with learning, and it was. However, the two relationships had opposite slopes, so when student and tutor turns were aggregated, words per turn only showed a strong trend of being correlated with learning.

Table 7
Summary of pair-wise correlations with student learning over the final tutoring segment

Measure	FTF	CM
Student words	n.s.	r = +.52, $p = .02$
Mean student words per turn	n.s.	r = +.58, $p = .008$
Tutor words	n.s	n.s.
Mean tutor words per turn	r = -.54, $p = .04$	r = -.52, $p = .02$
Student % words	n.s.	r = +.71, $p < .001$
Mean words per turn	r = -.55, $p = .03$	r = -.43, $p = .06$

To see which of these CM variables (some of which were inter-correlated) were most closely associated with student learning, we included student words, percent student words, mean student turn length, and mean tutor turn length in a stepwise regression analysis. Mean student turn length remained positively related to student learning, r(partial) = +.66, $p = .002$, and mean tutor turn length remained negatively related, r(partial) = -.62, $p = .005$. This suggests that student and tutor turn length are independent contributors to student learning.

[11] The difference in variance could also be due to a floor effect in the CM condition.
[12] Similar trends (though with lower p-values) were found for student words and turns. However, this is not surprising since student words, tutor words, and turns were all significantly correlated with each other.

To illustrate how mean tutor and student turn length related to learning in the CM context, excerpts are given from the final tutoring segment for the CM students with the lowest and highest gains, respectively, across this final tutoring segment (Figures 1 and 2). Note that in the low-learning CM excerpt, the tutor dominated the discussion and completed every step of the problem solution with the exception of calculating the numerical answer. In contrast, in the high-learning excerpt, the student tended to talk more and produced the equation on his own, while the tutor hung back and let the student do as much as he could.

T: this problem uses most of the stuff we already did
first we want to find the initial vertical velocity
S: ok
T: we will use the equation we used to find the Vi of the man jumping
S: ok
T: 2*a*(Xfinal-Xinitial)=Vf^2-Vi^2
we know the max height is 5m so we will use that
S: is 5m the Xfinal?
T: a= gravity, Xf=5m, Xi=0,Vf=0: because the object is at its max height and has essentially stopped, and Vi is what we are looking for
S: ok
T: remember that this is only the vertical component of the velocity, and solve the equation
S: the answer is 10

Fig.1. Excerpt from the CM session with the lowest learning gain.

S: OK, vertical v is going to relate to max height (5M) and force of gravity (10 m/sec/sec)
T: Yep
S: am I allowed to pull the formulas out of my wallet?
…
S: ending velocity will be zero, correct?
T: We can't do this without the formulas. I say go ahead.
S: [writes equation] that's out of memory -- how wrong is it?
T: That' right.
S: No kidding? Cool
T: Yes.
S: Now what?

Fig.2. Excerpt from the CM session with the highest learning gain.

Now let us turn to the FTF condition (Table 7, middle column). The mean tutor turn length was negatively related to student learning, as was the aggregated words per turn – dialogues with fewer words per turn produced more learning. These results suggest that more interactive tutoring produced higher learning gains in the FTF context. However, because student words per turn was *not* significantly related to their learning, the aggregate words per turn relationship to student learning is due to only the tutors' words per turn.

Figures 3 and 4 are excerpts from the FTF tutoring sessions that resulted in the lowest and the highest learning gains, respectively. In the low-learning excerpt, the tutor not only produced the relevant equation, but also used the equation to calculate the target variable's value. After the tutor had derived the value without interacting with the student, the student showed confusion by asking the tutor vague clarification questions. This example demonstrates how lack of interaction in the form of a lengthy tutor turn may impair student learning in the FTF context. In contrast, for the highest-learning excerpt, the tutor did not give the student the equation, even after the student indicated in her response that she needed some equations to move forward in the problem. Rather, the tutor "scaffolded" the student, giving her only minimal support in producing the equation.

T: The first thing we want to find is this Vi, this initial velocity. That would be the easiest thing to do. Let's see… maximum height is 5 meters. So we have that Vi it's broken into Vi cos 45, and Vi sin 45. And then we have… so you know that the maximum height is 5 meters. So we have delta y equals 5 meters. And then we know a in the y-direction equals 10 meters per second squared. Let's see… And what you need is… let's start with the equation 2 a delta x, which would be final minus v initial squared… So we know what, V? If we just go from… so that would be in the y-direction. So your v, that's Viy, you know the final y velocity, right here, is zero, when it hits the ground. So you have twenty equals Viy squared… times five… so you have Viy equal to ten. So, that's believable.
[pause]

S … Viy squared?

T: Yeah, that's Viy squared.

S: Square root?

T: Yeah, the square root of that's ten.

S … and then the Vi?

T: This V final is zero. So you got Viy squared, what is that minus? It would be minus. We can go… square root of a negative number. We have to take the absolute value. Should be… it's just… it is ten. The negative number, it's in there, but it's not important. You know that velocity…

S: No I mean… the square root.

T: Yeah, that's the square root of 100 is ten. So, it'd give you viy.

S: viy…

T: Ok, this is… This isn't Vi times y. This is a sub, like a subscript of y. That's just what I mean with those two. That would be Vi in the y-direction.

S: Ok

T: That's the way it gets in the end. That's where physics gets in the way, you get too many subscripts to keep track of. And you see how, you understand by trigonometry you get that to equal ten?

S: Ok

Fig. 3. Excerpt from the FTF session with the lowest learning gain.

Thus far, our hypothesis that student activity should predict greater learning gains in both contexts was satisfied only in the CM context. Student activity, as measured by words per conversational turn, was only associated with student learning in the CM context. But this may not be a good estimate of student activity in problem solving, especially in the FTF context, where tutors' back-channel responses (responses used to indicate that one is listening such as "mm-hmm" or "uh-huh") may break up longer student turns, hiding a possible relationship with learning (refer to the middle portion of the dialog in Figure 4 for examples of tutor back-channel responses). For example, in the FTF context, a lengthy student explanation may have been broken into several turns as a result of tutor back-channel feedback, whereas in the CM context this explanation would be coded as one turn. Thus, "removing" back-channel feedback could potentially have the effect of allowing similar relationships between features of dialog and learning outcomes to be uncovered. For this reason, we calculated FTF student words per turn ignoring tutor back-channel responses[13]. Again, there was no pair-wise relationship between the two, r = -.03, *p* = .92, or partial relationship when tutor turn length was also included in the regression, r(partial) = -.15, *p* = .59. However, when students' back-channel responses were removed in addition to tutors', the relationships between measures of student and tutor activity and student learning now look very similar to those in the CM context (Table 8)[14]. Now, measures of student activity including words spoken, words per turn, and percentage of words spoken were positively related to their learning. When a backward step-wise regression was run, only student words per turn remained significantly related to learning.

T: and, uh, we know that this height is gonna be what? This is the 5 meters. OK? So, um... do you... do you think that you have enough information to find the vertical velocity in the vert-- this Vy?
S: Uh, I don't know. I'm not really sure [what I'd do.]
T: Do you need some equations?
S: I guess, yeah. Cause I can't do anything without them.
T: Do know what kind of equation you'd like-- what quantities it should relate?
S: velocity
T: mm-hmm.
S: distance, I guess
T: mm-hmm.
S: uh, acceleration
T: mm-hmm, OK. Do you remember the equation that we-- from the other problems? The one with the acceleration and distance?
S: the 2 a...
T: mm-hmm.
S: times...
T: distance. delta x.
S: OK, yeah. Is that it?
T: yep, yep. good, good.

Fig.4. Excerpt from the FTF session with the highest learning gain.

Table 8
Summary of pair-wise correlations with student learning over the final tutoring segment (tutor and student back-channel feedback (BCFB) removed in FTF context)

Measure	FTF (no BCFB)	CM
Student words	r = +.52, *p* = .048	r = +.52, *p* = .02
Mean student words per turn	r = +.60, *p* = .02	r = +.58, *p* = .008
Tutor words	n.s	n.s.
Mean tutor words per turn	r = -.51, *p* = .05	r = -.52, *p* = .02
Student % words	r = +.49, *p* = .06	r = +.71, *p* < .001
Mean words per turn	r = -.48, *p* = .07	r = -.43, *p* = .06

[13] That is, the number of words in consecutive student turns immediately before and after tutor back-channel responses (e.g. "yeah", "mm-hmm") were combined and considered to be one turn.
[14] One student was an outlier in terms of student words, words per turn, and percent words spoken (2.9, 3.2, and 2.3 SDs above the mean, respectively). This student was excluded from these analyses.

To supplement the prior analyses and more directly test our hypothesis that student productivity is associated with learning regardless of context, we looked at a deeper, more direct measure, the percentage of problem-solving steps in the first topic of the final problem that were student-initiated. This included who initiated equation choice, who initiated finding the values for each of the terms in the equation, and who initiated solving the equation. An example of a student initiating equation choice is shown below.

S: Mm-hmm, mm-hmm, well you don't know anything about time but you do know about height? So you want to use this [equation]? Do you want to use that [equation]?

Fig. 5. Example of student initiating equation choice (FTF context).

The proportion of steps that were student-initiated was positively correlated with student words per turn in the FTF context ($r = +.64$, $p = .008$). This suggests that student words per turn reflects productive student activity. When student-initiated steps were included in a backward regression analysis with student and tutor words per turn (excluding back-channel feedback), and math and verbal SAT scores, only tutor words per turn and student-initiated problem steps remained significantly related to student learning, $r = -.71$, $p = .01$; $r = +.63$, $p = .03$, respectively. Thus, like student words per turn, this deeper measure of student activity may be a good predictor of learning, and possibly a better predictor than student words per turn. To investigate interactions between student activity and other measures, an ANCOVA was run. Because tutor turn length was negatively related to student learning in both contexts, this measure was included in the model. Additionally, because students' ability to understand tutor turns (particularly longer ones) may have contributed to the negative relationship between tutor words per turn and student learning (especially in the FTF context), students' SAT verbal scores were included in the model. If longer tutor turns were negatively related to student learning because students were having more difficulty processing them in the relatively short allowable time between turns, we would expect that students with lower verbal ability would have more difficulty with longer tutor turns than students with higher verbal ability. In other words, we would expect a stronger negative relationship between tutor words per turn and student learning for students with lower verbal SAT scores than for students with higher scores. And in fact, there was a tutor words per turn by student verbal SAT score interaction that approached significance, $F(1, 8) = 4.88$, $p = .058$.

Additionally, there was an interaction between student verbal SAT scores and the percentage of initial topic solution steps that were student initiated that approached significance, $F(1, 8) = 4.86$, $p = .059$[15]. To investigate these interactions, students were classified as "high" or "low" verbal ability (i.e. those with verbal SAT scores above and below the mean of 572, respectively). For high-verbal students, a multiple regression model with both student initiation and tutor words per turn as independent variables showed a significant positive relationship between the percentage of student-initiated problem steps and learning across the final tutoring segment[16], r(partial) = +.97, $p = .006$. However, there was no significant relationship between student learning and tutor turn length,

[15] When the same analysis was run using math scores rather than verbal scores, there were no interactions between math scores and percentage of student-initiated problem steps or math scores and tutor words per turn that were significant or approached significance.

[16] As measured by the total post-test score with mid-test scores factored out.

r(partial) = +.22, p = .72. Unlike the higher-verbal students, for lower verbal students, there was a non-significant negative relationship between student initiation and learning, r(partial) = -.25, p = .59. Thus, lower-verbal students did not appear to benefit from their problem-step initiations during tutoring. However, there was a significant negative relationship between tutor words per turn and student learning, r = -.85, p = .02. Thus, consistent with our prediction, long tutor turns were negatively related to students learning only for lower-verbal students. Additionally, the hypothesis that greater student initiating activity is associated with greater student learning in the FTF context was supported only for students with higher verbal scores.

It does not appear that this negative relationship was due to tutors simply taking over actions that would have been more beneficial for the lower-verbal student to do themselves: there was no significant relationship between student-initiated steps and tutor words per turn, r = -.34, p = .40. A more plausible interpretation is that lower-verbal students had difficulty initially understanding or remembering the content of longer tutor statements. This interpretation is supported by the dialog shown in Figure 3, which is an excerpt from a tutoring session where the tutor's turns are relatively long and the student was classified as lower-verbal; this student had the lowest learning gain across the final tutoring segment. As evidenced by the subsequent dialog, this student did not understand the tutor's initial lengthy explanation of how to find the initial vertical velocity.

In the CM context, as in the FTF context, the proportion of steps that were student-initiated was again positively correlated with student words per turn in the FTF context (r = +.60, p = .01), which is consistent with the assumption that student words per turn reflects productive student activity. When student-initiated steps were included in a backward regression analysis with student and tutor words per turn, and math and verbal SAT scores, only tutor words per turn and student words per turn remained significantly related to student learning, r = -.65, p = .006; r = +.70, p = .002. Thus, in the CM context, this deeper measure of student activity does not appear to be a better predictor of overall learning than overall student words per turn. To investigate interactions between student-initiated activity, tutor words per turn, and verbal SAT scores, an ANCOVA was run. There were no 2- or 3-way interactions that were significant or approached significance. Only student initiating activity and tutor turn length were related to student learning. With these two variables entered in a regression model, student initiating activity was positively associated with learning, r(partial) = +.53, p = .03[17]. Additionally, tutor turn length was negatively associated with student learning, r(partial) = -.55, p = .02. But in the CM context, tutor turn length did not differentially impact higher and lower-verbal students' learning like it did in the FTF context. One possible explanation for this context difference is that lower-verbal students in the CM context re-read and/or used more time to process tutor messages, reducing the advantage the better readers had. The explanation that longer tutor turns were indicative of tutors taking over actions that would have been more beneficial for students to have done themselves is not supported because both variables independently contributed to learning gains. Moreover, the correlation between student-initiated activity and tutor words per turn was not significant, r = -.11, p = .66.

Because they are related to student learning in both contexts, student-initiated problem-solving steps appear to be a good measure of "productive student activity". Because there were no significant differences in learning between contexts, we would expect there to be no differences between contexts

[17] There was a higher correlation between student turn length and student learning than between the measure of student activity in this analysis (i.e. percent of student-initiated action in the first topic of the final question) and learning, r(partial) = +.66, p = .002. Thus, student turn length is a better predictor of learning.

in the proportion of actions that were student-initiated. And in fact, there was no difference between conditions on this measure. In the FTF context, students initiated on average 27.19% of these steps, whereas in the CM context, students initiated on average 34.07%, a non-significant difference, $F(1, 32) = 0.39$, $p = .54$.

DISCUSSION

The primary question addressed in the present study was how do student learning and motivation outcomes compare in face-to-face and synchronous CM tutoring contexts. There was no difference between the CM and FTF contexts in students' overall learning, as measured by their total post-test scores, nor were there any differences on any of the subsections of the post-test. This is consistent with the Litman et al. (2006) study of human tutoring, which did not find a significant difference in learning gains between a CM condition and a condition where students spoke with (but could not see) their tutors. The lack of a difference is also consistent with some studies comparing synchronous or asynchronous on-line and face-to-face learning (e.g. Ocker & Yaverbaum, 1999; Schpilberg & Hubschman, 2003; Dennis, 2003; Pilkington & Walker, 2003). How might these results translate into group tutoring results? Gallupe et al. (1992) found that as group size increased, per-person productivity on a brainstorming activity decreased in FTF groups, but did not in asynchronous CM groups. This result may be due to factors such as production blocking and social loafing in FTF groups. This raises the possibility that if group size had increased, FTF group performance would decrease relatively more than CM group performance, allowing the possibility that CM groups would out-perform FTF groups on this task. However, task differences make this analogy questionable, and additional studies are necessary to answer this question directly.

Although there was no difference in gains, the time required to achieve those gains was significantly less in the FTF context. That is, tutoring in the FTF context was more time efficient. This finding is consistent with the Litman et al. (2006) study and with general findings for group interactions (Bordia, 1997). In CM tutoring, the student must wait while the tutor types a response, so any time the student is not involved in productive activity would be wasted.

Although the tutors dominated the discussion in both the FTF and CM contexts, the percentage of words produced by the students showed a strong trend of being higher in the FTF context than in the CM context. In addition, there was a non-significant trend of FTF students producing more words per turn. The Litman et al. (2006) study found the opposite trends. That is, the Litman et al. CM students generated a significantly higher percentage of words than the spoken-tutoring students, and had a significantly higher number of words per turn as well. The greater student activity in the Litman et al. study compared to the present one may be due to a difference in the way typed communication was handled. In the Litman et al. CM context, turn taking was enforced. In particular, the tutor could not send a message until the student had finished and sent his or her message. In our study, an impatient tutor could send a message while the student was still typing[18]. Doing so may have reduced the overall number of student words and words per turn in the CM context of our study.

Face-to-face tutoring has been found to increase students' interest in the tutoring domain (e.g. Cohen et al., 1982). However, in the present study, over the course of the tutoring session, students'

[18] It was not possible to tell when and how often this occurred from the existing data.

reported interest in tutoring actually decreased in both tutoring contexts[19]. But there was no difference between contexts in the magnitude of this drop. Students' change in confidence did not differ across tutoring contexts, nor did confidence change across time. Thus, there was no evidence that tutoring context had different effects on students' domain interest or self-confidence.

The only motivational factor for which there was a strong trend towards a time by context interaction was ability goal (validation); this was the only factor that targeted inter-personal motivational beliefs. Surprisingly, FTF students' scores were stable across time whereas CM students' scores increased from pre- to post-tutoring. This result is surprising because CM communication is usually considered to be less personal than FTF communication because fewer social cues are available such as facial expression and voice intonation. Thus, we did not expect that CM students would be sensitive to these items. One possible reason why CM students' ability goal increased is that students' need for validation from the tutor was not satisfied due to the limited social cues in the CM context that can convey tutors' approval of student ability (e.g. vocal intonations). Because stronger ability goals have been linked to weaker performance (Grant & Dweck, 2003), this result favors the FTF context. However, because this was not significant at the .05 level and was the only context difference for our measures of student confidence, intrinsic motivations, and different achievement goals, our findings are cause for optimism for the effect of CM tutoring on motivational measures compared to FTF tutoring. These results may not be good news for those developing affect sensing to computer tutoring that does not go beyond what human tutors do naturally. However, more studies that address this issue are important to support these results and to assess other motivational measures, especially those that are known to be related to student learning. Additionally, the only motivational measure that FTF tutors demonstrated accurate assessment of (which CM tutors did not) was the challenge factor.

The remaining findings concern correlations between learning gains and various measures of student and tutor activity. In order to make sense of these findings, we will assume that two independent factors are primarily responsible for determining the size of the student's learning gain.

1. A student factor that is not measured by the pre-test – how productive they are during the tutoring session. We previously hypothesized that, regardless of tutoring context, more productive student engagement should cause higher learning gains.
2. A factor that we will call "tutorial skill". More skillful tutors cause higher student gains.

Let us assume that random sampling caused equivalent distribution of student engagements and tutoring skill in the FTF and CM conditions, and that neither student engagement nor tutoring skill changed during the course of tutoring. Since by hypotheses, these two factors are the major ones determining learning gains, these assumptions explain why the two conditions displayed the same learning gains. Now the stage is set for explaining the within-condition findings as well.

In the CM condition, student turn length was positively correlated with learning, which was also found in the Litman et al. (2006) and Rosé et al. (2001) studies. For typed communication, student turns with more words were associated with more learning. Litman et al. (2006) suggested that this effect may be due to tutors receiving more information about the students' beliefs, which would allow the tutor to more effectively adapt the tutoring to the student's needs. However, if this were true, we

[19] One possible reason that students' reported interest decreased is that, unlike tutoring in real-world settings, in this study, tutors and students could not choose which problems they discussed, which may be of more interest to them.

would expect longer tutor turn lengths, which presumably contain more feedback for the student, to be positively related to student learning, when, in fact, we found the opposite. Thus, we believe that student engagement is the more likely explanation for the positive relationship. A positive relationship between student words per turn and learning is a robust finding across studies. If this relationship can be established as causal, then tutors in a CM context may be advised to try to encourage more student elaboration and activity in their turns.

In the FTF condition, student turn length was not correlated with learning. This is consistent with Litman et al. (2006), who found no correlation in the FTF condition between student turn length and learning gains. Our hypothesis that student engagement is causing both increased learning and increased turn length should apply equally well to FTF as it did to CM. Why do the two contexts differ? A major difference is that FTF tutors and students may provide "back-channeling," (Clark, 1996) to each other within a conversational turn. Figure 4 provides illustrations. When transcribed, these interjections break a spoken utterance up into several smaller turns. Thus, back-channeling may hide associations between student activity and learning in the FTF context. Since back-channeling within a turn cannot occur in the CM context, it displays the expected association: longer student turns are correlated with larger learning gains.

When we accounted for (verbal) tutor and student back-channel responses by factoring them out of activity measures, we found significant positive correlations between student activity (i.e. words, words per turn, and percent words) and learning. In fact, the pair-wise relationships between tutor and student activity and student learning in the FTF context were strikingly similar to those in the CM context. So it appears that student activity measures such as turn length, which is a good predictor of learning in the CM context, may also be a good predictor of student learning in the FTF context after back-channeling is removed.

In addition, we found that another, deeper, measure of engagement, student-initiated problem-solving steps, was predictive of student learning in the FTF context. However, this was true only for higher-verbal students. Why did lower verbal FTF students fail to benefit from their initiations? One possibility is that the additional memory load they experienced during tutoring precluded them from retaining what they had learned, regardless of how engaged or active they were.

For both CM and FTF conditions, the number of tutor words per turn was negatively correlated with student learning. This makes sense if we assume that long tutor turns were confusing. This should be especially true in this task domain, quantitative physics, where the tutor turns often contained mathematical expressions and variable terms in those equations that can be difficult to convey. These may be particularly difficult to convey verbally, as in the FTF context. However, the similar negative correlations in both contexts do not suggest that the mathematical nature of the task was more problematic in the FTF context. This may be because in this study, in the FTF context, tutors and students could write anything they wanted, including equations, which could supplement these verbal explanations and resolve potential confusions that arise if presented only verbally. Thus, the quantitative nature of the task does not appear to explain why the number of tutor words per turn was negatively correlated with student learning.

Another possible explanation for the negative relationship between tutor words per turn and student learning is that longer turns were related to tutors performing actions that students would have benefited from doing themselves. If this were true, we would expect there to be negative correlations between student-initiated problem-solving steps and tutor words per turn. However, there were no significant correlations. So there is no evidence that this caused the negative relationship between turn length and student learning.

Perhaps longer tutor turns were negatively associated with learning because longer turns are simply more difficult to parse and understand. To assess this possibility, we looked at the relationship between tutor turn length and learning as a function of student verbal SAT scores. In the CM condition, the negative relationship between tutor words per turn and student learning was similar for higher and lower-verbal students. Thus, in the CM context, the possibly confusing longer tutor statements affected higher and lower-verbal students similarly. This could be because students with lower verbal ability were able to re-read anything they did not initially understand, thus negating any comprehension differences due to reading ability. But why was the overall relationship between tutor words per turn and student learning negative in the CM context? In the Litman et al. (2006) study, tutor turn length was *positively* correlated with learning in the CM condition. The Litman tutor was not only highly skilled, but he had more practice tutoring students on the problems, so it is plausible that his long turns may have been better crafted and easier to understand than the long turns of our tutors. More specifically, his explanations may have been generally more coherent than those on-the-fly explanations generated by the tutors in this study, requiring fewer inferences to understand. And, because back-channel feedback is not possible within a turn in CM tutoring, less coherent explanations may have a larger negative impact than in FTF tutoring. Additionally, the nature of the tutoring task differed in the present and Litman et al. (2006) studies. In Litman et al., students first gave typed essay responses to qualitative physics questions before the tutor began providing feedback on their responses, whereas in the present study, students were never given an opportunity to solve the problem on their own first. If tutors' statements were more often in response to students' work in the Litman et al. study than in the present one, then this may have made them more understandable to the Litman students. These final two explanations may account for both the positive relationship between tutor words per turn and learning in the Litman et al. study and the negative results in the CM context of the present study.

In the FTF context, tutor turn length was negatively correlated with student learning only for students with lower verbal SAT scores. There was no significant relationship between tutor words per turn and learning for students with higher verbal SAT scores. We would expect this pattern if tutors were not trying to make their longer turns more coherent, thus requiring more inferences to understand, which may be more difficult for less verbal students. In the Litman et al. study, there was no significant relationship between tutor words per turn and student learning in their spoken communication condition. If it is true that the Litman tutor gave better, more coherent explanations than the tutors in the present study, then this could explain why in their study, wordier tutor turns were not related to student learning in general, but in the present study, wordier tutor turns were negatively related to learning for students with lower, but not higher verbal ability. However, further research is necessary to determine reasons for different relationships between tutor words per turn and student learning and establish causal patterns. Only then can tutors in a CM setting be advised of how much explanation or elaboration to give their students.

In summary, our major finding is that FTF tutoring was *not* more effective than CM tutoring for learning or for motivational gains. This finding may be surprising to many researchers, especially those that are working hard to build spoken-language interfaces to tutoring systems. On the bright side, FTF tutoring was more efficient than CM tutoring, so a spoken language tutoring system may have considerable practical importance because it takes less time to achieve the same gains as a typed-language tutoring system. When typing time was converted to speaking time in the CM context, the difference in time efficiency between the CM and FTF contexts was no longer significant. However, because it is possible that a reduction in overall time may result in reduced student learning, this analysis did *not* show that spoken language CM tutoring would be as efficient as FTF tutoring; it merely did not rule out this possibility. This question can only be answered by directly comparing these two conditions. Lastly, Chi et al. (2001) compared three hypotheses for why students learn so much from human tutors. The tutor-centered hypothesis is that tutoring effectiveness arises from the tutor's pedagogical skills. The student-centered hypothesis is that it arises from the student's active generation during the learning task, and the interactive coordination hypothesis is that it arises from the joint effort of both the tutor and student. The student-centered hypothesis seems to be most clearly supported in both contexts by our finding of a positive correlation between learning gains and student activity measures (words per turn and student-initiated problem steps); it is less clear which of these hypotheses is supported by our finding of an overall negative relationship between tutor turn length and student learning. However, for future studies that investigate these hypotheses, our results suggest that they could be done in the somewhat more tractable CM context.

ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research, Cognitive and Neural Sciences Division MURI grant N00014-00-1-0600 and NSF Grant 9720359 to Circle, a center for research on intelligent tutoring. We also wish to thank Chris Schunn, Jonathan Schooler, and Susan Fussell, for their helpful guidance in the earlier stages of this research. And thanks to David Klahr, Ken Koedinger, Kevin Willows, and Jodi Davenport for their helpful comments on drafts of this paper.

REFERENCES

Adrianson, L. (2001). Gender and computer-mediated communication: Group processes in problem solving. *Computers in Human Behavior*, 17, 71-94.

Aist, G., Kort, B., Reilly, R., Mostow, J., & Picard, R. (2002). Experimentally augmenting an intelligent tutoring system with human-supplied capabilities: Adding human-provided emotional scaffolding to an automated reading tutor that listens. In *Proceedings of the ITS 2002 Workshop on Empirical Methods for Tutorial Dialogue Systems* (pp. 16–28), online. San Sebastian, Spain.

Anderson, D. (2002). Interfacing email tutoring: Shaping an emergent literate practice. *Computers and Composition,* 19, 71-87.

Benbunan-Fich, R., & Hiltz, S. (1999). Impacts of asynchronous learning networks on individual and group problem solving: A field experiment. *Group Decision and Negotiation*, 8, 409-426.

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(6), 4 -16.

Bordia, P. (1997). Face-to-face versus computer-mediated communication: A synthesis of the experimental literature. *Journal of Business Communication,* 34, 99-120.

Cheung, W. S., & Hew, K. F. (2004). Evaluating the extent of ill-structured problem solving processes among pre-service teachers in an asynchronous online discussion and reflection log learning environment. *Journal of Educational Computing Research*, 30(3), 197-227.

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

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.) *Perspectives on Socially Shared Cognition* (127-149). Washington, DC: APA.

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

Coleman, L. H., Paternite, C. E., & Sherman, R. C. (1999). A reexamination of deindividuation in synchronous computer-mediated communication. *Computers in Human Behavior,* 15, 51-65.

Condon, S. L., & Cech, C. G. (1996). Functional comparison of face-to-face and computer-mediated decision-making interactions. In Herring, S. (Ed.) *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 65-80). Philadelphia: John Benjamins.

Condon, S. L., & Cech, C. G. (2001). Profiling turns in interaction: Discourse structure and function. *Proceedings of the 34th Hawaii International Conference on System Sciences.* Maui, Hawaii.

Conole, G., Hall, M., & Smith, S. (2002). An evaluation of an online course for medical practitioners. *Educational Technology & Society*, 5(3), 66-75.

Core, M. G., Moore, J. D., & Zinn, C. (2003). The Role of Initiative in Tutorial Dialogue. Presented at the 10th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary, April, 2003.

Cummings, A., Schlosser, A., & Arrow, H. (1996). Developing complex group products: Idea combination in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work*, 4, 229-251.

Dennis, J. K. (2003). Problem-based learning in online vs. face-to-face environments. *Education for Health,* 16, 198-209.

Diakidoy, I., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2004). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology an International Quarterly*, 26(1), 55-80.

D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R. W., & Graesser, A. C. (2005). Integrating Affect Sensors in an Intelligent Tutoring System. *Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces*, 7-13.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256-273.

Fletcher, J., & Pumfrey, P. D. (1988). Differences in text comprehension amongst 7-8-year-old children. *School Psychology International,* 9(2), 133-145.

Forbes-Riley, K., Litman, D., Silliman, S., & Tetreault, J. (2006). Comparing Synthesized versus Pre-recorded Tutor Speech in an Intelligent Tutoring Spoken Dialogue System. *Proceeding 19th International FLAIRS (Florida Artificial Intelligence Research Society) Conference*, Melbourne Beach, FL.

Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, C. M., & Nunamaker, J. F., Jr. (1992). Electronic brainstorming and group size. *Academy of Management Journal,* 35, 350-369.

Goodyear, P., Salmon, G., Spector, J. M., Steeples, C., & Tickner, S. (2001). Competences for online teaching: A special report. *Educational Technology Research and Development,* 49(1), 65-72.

Grant, H., & Dweck, C.S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology*, 85, 541-553.

Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science,* 28*,* 115-152.

Heckman, R., & Annabi, H. (2005). A content analytic comparison of learning processes in online and face-to-face case study discussions. *Journal of Computer-Mediated Communication,* 10(2), article 7.

Heylen, D., Nijholt, A., & op den Akker, R. (2005). Affect in tutorial dialogues. *Applied Artificial Intelligence,* 19, 287-311.

Johnson, W. L., Rizzo, P., Bosma, W., Kole, S., Ghijsen, M., & van Welbergen, H. (2004). Generating socially appropriate tutorial dialog. In E. André et al. (Eds.) *Proceedings of the ISCA Workshop on Affective Dialogue Systems (ADS)* (pp. 254–264). Berlin:Springer-Verlag.

Jones, R. H., Garralda, A., Li, D. C. S., & Lock, G. (2006). Interactional Dynamics in On-Line and Face-to-Face Peer-Tutoring Sessions for Second Language Writers. *Journal of Second Language Writing.* 15(1), 1-23.

Karat, C.M., Halverson, C., Horn, D., & Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems, *CHI 99 Conference Proceedings* (pp. 568–575).

Lebie, L., Rhoades, J. A., & McGrath, J. E. (1995). Interaction process in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work*, 4(2-3), 127-152.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J.-L. (1993). 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* (pp. 75-105). Hillsdale, NJ: Erlbaum.

Litman, D. J., Rosé, C. P., Forbes-Riley, K., VanLehn, K., Bhembe, D., & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education,* 16(2), 145-170

Luppicini, R. (2007). Review of computer mediated communication research for education. *Instructional Science,* 35, 141-185.

Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. *Instructional Science*, 29*,* 127-153.

Moore, J. D., Porayska-Pomsta, K., Varges, S., & Zinn, C. (2004). Generating tutorial feedback with affect. *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference* (pp. 923-928). Miami Beach, Florida.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated agents? *Cognition and Instruction,* 19(2), 177-213.

Neuhauser, C. (2002). Learning style and effectiveness of online and face-to-face instruction. *American Journal of Distance Education.* 16(2), 99-113.

Neville, M. H., & Pugh, A. K. (1974). Context in reading and listening: A comparison of children's errors in cloze tests. *British Journal of Educational Psychology,* 44(3), 224-232.

Ocker, R. J., & Yaverbaum, G. J. (1999). Asynchronous computer-mediated communication versus face-to-face collaboration: Results of student learning, quality, and satisfaction. *Group Decision and Negotiation,* 8, 427-440.

Olaniran, B. A. (1996). A model of group satisfaction in computer-mediated communication and face-to-face meetings. *Behavior and Information Technology,* 15(1), 24-36

Pilkington, R. M., & Walker, S. A. (2003). Facilitating debate in networked learning: Reflecting on online synchronous discussion in higher education. *Instructional Science*, 31, 41-63.

Randhawa, B. S., Beamer, J. E., & Lundberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of Educational Psychology,* 85 (1), 41-8.

Rosé, C. P., Moore, J. D., VanLehn, K., & Allbritton, D. (2001). A comparative evaluation of Socratic versus didactic tutoring. In J. D. Moore & K. Stenning (Eds.) *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 897-902). Mahwah, NJ: Erlbaum.

Rubin, D. L., Hafer, T., & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education,* 49(2), 121-133.

Schpilberg, B., & Hubschman, B. (2003). Face-to-face and computer mediated tutoring: A comparative exploration on high school students' math achievement. Presented at the American Educational Research Association conference, Chicago, IL.

Siler, S. A., & VanLehn, K. (2003). Accuracy of tutors' assessments of their students by tutoring context. In R. Alterman & D. Hirsch, (Eds.) *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Summers, J. J., Waigandt, A., & Whittaker, T. A. (2005). A comparison of student achievement and satisfaction in an online versus a traditional face-to-face statistics class. *Innovative Higher Education,* 29(3), 233-250.

Tu, H. (2000). On-line learning migration: from social learning theory to social presence theory in a CMC environment*. Journal of Network and Computer Application*, 23, 27-37.

Tutty, J. I., & Klein, J. D. (2008). Computer-mediated instruction: A comparison of online and face-to-face collaboration. *Educational Technology Research and Development,* 56(2), 101-124.

van der Meijden, H., & Veenman, S. (2005). Face-to-face versus computer-mediated communication in a primary school setting. *Computers in Human Behavior*, 21(5), 831-859.

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction,* 10(4), 293-309.

Wald, M., & Bain, K. (2008). Universal access to communication and learning: The role of automatic speech recognition. *Universal Access in the Information Society*, 6(4), 435-447.

Warren, L. L., & Holloman, H. L. Jr. (2005). On-line Instruction: Are the outcomes the same? *Journal of Instructional Psychology,* 32(2), 148-151.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology,* 89, 420-433.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal,* 29(3), 663-676.