Tenth Midwest Artificial Intelligence and Cognitive Science Conference (MAICS '99), Bloomington, IN. (to appear)

NOVICE VS. EXPERT TUTORS: A COMPARISON OF STYLE

Michael Glass glass@charlie.cns.iit.edu

Jung Hee Kim kimjung@charlie.cns.iit.eduu

Martha W. Evens csevens@minna.cns.iit.edu

Department of CS, Illinois Institute of Technology
10 W. 31st Street 236–SB Chicago, IL 60616

Joel A. Michael jmichael@rush.edu

Allen A. Rovick arovick@rush.edu

Department of Physiology, Rush Medical College
1750 W. Harrison Chicago, IL 60612

Abstract

The CIRCSIM-Tutor intelligent tutoring system project has been built on the basis of numerous studies of transcripts of expert human tutors (professors) teaching first year medical students. We also have transcripts of novice tutors (second year medical students) teaching the same material to medical students at the same level. In this paper we identify measurable differences in the teaching styles between the novices and experts. Examples of tutoring of identical topics were isolated from the novice- and expert-tutored transcripts and various dialogue acts were counted. The primary result is that expert tutors are more likely than novice tutors to query students for information as opposed to informing them directly.

Introduction

The CIRCSIM-Tutor project is building an intelligent tutoring system designed to tutor first-year medical students on the baroreceptor reflex, a mechanism for blood pressure regulation in the human body. The machine tutor is designed to imitate aspects of human tutorial dialogue. Our data for tutoring behavior and language comes from a corpus of fifty one- and two-hour tutoring sessions, conducted by typed communication between the tutor and student in separate rooms.

The sessions that we have analyzed for the construction of the machine tutor have all been taught by two of the authors, Joel Michael and Allen Rovick, who served as the medical students' own physiology professors. In addition to experience teaching the baroreceptor reflex in the classroom, they had considerable experience teaching this material one-on-one, in small groups [Michael 1993], and in computer-based teaching labs [Rovick and Michael 1992].

In addition to the sessions taught by our expert tutors, we have transcripts of teaching by "novice" tutors. The novice tutors for this experiment were four second-year medical students recruited for purposes of this experiment. They did not necessarily have prior tutoring experience, but they received training as described below. Altogether this experiment accumulated fifteen novice tutoring sessions, teaching one of the same problems as taught in many of the expert sessions.

This study does not endeavor to compare large samples of expert tutors to novice tutors. If we have only one expert tutor it is enough for our purpose: imitating some of the expert tutor's teaching style in a machine tutor. We have pre-test and post-test evidence that the expert tutors we use are effective, even after tutoring only one problem [Rovick and Michael 1992], so we know these are people it would be useful to imitate. It is our hope that by comparing their style to novice tutoring we can learn interesting characteristics of a style we know to be effective.

There have been a number of published analyses of the expert tutors' language behavior in the CIRCSIM-Tutor project expert transcripts. The tutors' use of negative acknowledgments was analyzed in [Spitkowsky and Evens 1993; Evens et al. 1993]. A more recent accounting covering all types of acknowledgments in detail appears in [Brandle 1998, 1999; Brandle and Evens 1997]. Hinting behavior was analyzed in [Hume et al. 1996a, b]. Language differences between typed and spoken tutoring were described in [Seu et al. 1991]. Student initiatives, where the student temporarily sets the discourse agenda, and also the tutors' responses to those initiatives, were described in [Sanders et al. 1992; Shah 1997].

Recently transcripts have been marked up to show the tutorial goal structure [Freedman et al. 1998b; Kim et al. 1998b]. These versions of the annotated expert transcripts have been used for purposes such as analyzing sentence

This work was supported by the Cognitive Science Program, Office of Naval Research under Grant No. N00014-94-0338 to Illinois Institute of Technology. The content does not reflect the position or policy of the government and no official endorsement should be inferred.

structure [Kim, Freedman and Evens 1998a] and automated analyses of tutorial schemata [Freedman et al. 1998a]. These are the marked up transcripts of expert tutors we used in the comparison in this paper.

Selection of Comparison Text

The essence of this experiment is observing tutoring language behavior in comparable novice- and expert-tutored texts. In order to understand what are comparable texts we need to briefly describe the tutoring task.

The context of our transcripts is that the student solves a problem in blood pressure regulation. The problem is usually posed as an external event which perturbs blood pressure. The student then predicts qualitative changes for a number of causally related physiological variables. After predictions have been made, the tutor usually corrects the erroneous predictions one by one. There are three stages in the response to a disturbance of blood pressure; for each of the stages the cycle of student prediction and tutor remediation is repeated.

In comparing novice tutoring to expert tutoring we controlled for the following factors:

- Problem the student was solving
- Stage of the problem
- Physiological parameter being tutored

The fifteen novice tutoring transcripts all involve the students working the same problem, which postulates a cardiac pacemaker which starts pulsing at the wrong rate. We took from our expert texts the twenty-two previously marked up expert tutoring transcripts that include the same pacemaker problem. In each transcript we selected the first of the prediction/tutoring cycles (the first physiological stage), partly because this is the best understood and most comprehensively marked up part of the corpus, and partly because not all tutoring sessions included the later stages.

Within this first stage, we selected all tutoring instances dedicated to the physiological parameters stroke volume (SV) and central venous pressure (CVP). Predicting these variables tends to be difficult for the students, resulting in more and longer instances of tutoring. In some transcripts the parameter right atrial pressure (RAP) substituted for CVP, which is physiologically nearly identical. Each segment of selected text starts with the tutor introducing the variable into the conversation and ends when discussion of that variable has finished. One variable is represented at most once in each stage of the problem in each transcript; it is usually absent if the student predicted it correctly. Altogether we had twenty-three segments of tutoring dialogue from the first stage of tutoring, as summarized in Table 1.

		Expert	Novice
Var: SV	Segments	6	5
	Turns	26	26
Var: CVP	Segments	5	7
	Turns	35	24
Total	Segments	11	12
	Turns	61	50

Table 1. Quantities of Transcripts Selected

The medical students being tutored in both the expert and novice transcripts were first-year medical students, usually at approximately the same stage in the physiology course taught by the same professors. Although they came from several different cohorts of students (different school years), we believe that the populations tutored by the novices and experts are approximately the same in ability.

Novice Tutor Training

One variable we did not control for was the training history of the novice tutors. The tutors were put through separate courses of training in tutoring and in CIRCSIM-Tutor's problem domain, the baroreceptor reflex. Some novices received domain training and some received tutoring training. Domain training was designed to teach the topic without using tutoring. It consisted of several stages of reading texts and solving problems, with periodic testing and written non-interactive feedback. Tutoring training, conversely, avoided the baroreceptor reflex in favor of a different physiological feedback mechanism. It consisted of readings, being tutored by the experts, and practice tutoring with feedback. Since the novice tutors had learned the baroreceptor reflex in a previous school year, it was believed that the tutors who received only tutoring training still had a useful knowledge of the topic.

The novice transcripts could thus be divided into two groups depending on the training received by the novice tutor, but after controlling for other variables as described above there were not enough segments to control for training history.

Annotation and Counting of Text

An excerpt from a marked up transcript of expert tutoring is displayed in Figure 1. This excerpt contains the tutoring for one erroneously predicted variable, Stroke Volume (SV). Annotations are coded in SGML.

In this markup we show what we call "primitive dialogue acts." The fully annotated transcript contains descriptions of higher-level tutorial and dialogue acts (described later in this paper). Higher level goals can be realized by lower level goals until a primitive dialogue act is reached, which is a description of a piece of text.

The primitive dialogue acts in our expert transcripts are ***t-elicits*** (tutor elicits information from the student), ***t-informs*** (tutor asserts something), ***t-ack*** (tutor

<t-elicits info=value ...>
tu: So, if CVP is down what happens to (RAP and) SV?
<s-ans catg=correct>
st: RAP decreases, and SV must follow suit.
</s-ans>
</t-elicits>
<t-informs stage=DR, info=...>
tu: So, in DR HR is up, CO is up, but SV is down.
</t-informs>
<t-elicits ...>
tu: How is this possible? What must be true if all three of these predictions are correct?
<s-ans catg=correct>
st: That HR increases outstrip SV decreases in this case.
</s-ans>
<t-ack>
tu: Exactly!
</t-ack>
</t-elicits>

Figure 1. Expert SV Tutoring with Excerpted Annotations

acknowledges a student's answer), and ***s-ans*** (student answers). Note that student utterances are not categorized into ***informs*** or ***elicits*** because we are always studying the tutor's behavior. We like these primitive dialogue acts because they are reasonably unambiguous and cover almost the full range of utterances. Thus we can count how many ***elicits*** and ***informs*** are in a selected text segment, telling us how many times they were used in the process of completely correcting one variable.

When marking up novice tutoring transcripts for this study we identified another primitive dialogue act we annotated as ***t-asks-confirmation***, which appears as tutor questions such as "do you understand?" and "right?"

We counted the number of turns of dialogue in addition to primitive dialogue acts. We included only tutor turns in the count. Turn-taking is enforced by our keyboard-to-keyboard conversation program CDS [Li et al. 1992]; the two parties cannot type simultaneously. One turn consists of one party typing, possibly including several sentences, until control is relinquished to the other party.

In our counts of tutor turns and behaviors we eliminated instances of simple conversational repair, such as embodied in the final two turns of the following:

tu: Now tell me some of the things that affect the SV?
st: IE and EDV.
tu: What is IE?
st: Oops, it should be IS.

Another feature we counted was physiological and anatomical concepts mentioned, referred to, or otherwise used by the tutor. The goal is to see how many concepts the tutor uses during the course of remediating one erroneous variable. In this example the concepts we counted are italicized. Each concept is counted only once:

tu: What else changes in response to *CO*?
st: Stroke volume.
tu: No, *stroke volume* contributes to CO. Think of it in terms of WHERE that increased CO ends up.
st: TPR.
tu: No, CO pumps more blood into the *arterial circulation* and takes away blood from the *venous circulation ….*
st: *CVP*
tu: Right.

In this excerpt the parameter stroke volume is broached by the student and the tutor imparts some domain knowledge about it, so stroke volume is counted. TPR is broached by the student and never picked up by the tutor, so it is not counted. The parameter CVP is broached by the student and the tutor is using it as part of the argument, so it is counted. The phrases "arterial circulation" and "venous circulation" are counted here as anatomical terms, standing for different collections of blood vessels.

Differences In Tutoring SV

Table 2 shows the all the counts from novice and expert transcript segments for the variable SV.

	Novice n	Novice n/seg	Expert n	Expert n/seg
Segments	5		6	
Turns	26	5.2	26	4.3
Elicits	13	2.6	17	2.8
Informs	52	10.4	18	3.0
Acks	13	2.6	15	2.5
Ask-conf	7	1.4	0	0.0
Total prim. acts	85	11.2	50	8.3
Concepts	56	10.4	30	5.0

Table 2. Tutoring Counts for SV

Among the primitive dialogue acts a striking difference is that the novices issued considerably more ***informs*** than the experts did, using an average of 10.4 ***informs*** during the course of tutoring SV while the experts used an average of only 3.0. Another striking difference is that the novices issued a number of ***ask-confirmation***s, 1.4 per segment, while the experts never engaged in that behavior. For the 4×2 contingency table (Table 3) showing counts of all four dialogue acts and two independent samples, χ^2 = 16.2, meaning the novices and experts differ in their issuance of primitive dialogue acts at the 0.01 significance level.

The absence of expert ***ask-confirmation*** acts confirms that the two expert tutors, who have been known to claim

	Novice	Expert
Elicit	13	17
Inform	53	18
Ack	13	15
Ask-conf	7	0

Table 3. SV Primitive Dialogue Acts

that these questions do not lead to productive tutoring, seem to practice what they preach.

There may be an empirical basis for this distaste. Graesser, Person and Magliano [1995], in an experiment with unskilled tutors, recognized the tutorial task of assessing a student's understanding of an answer. In 92% of the instances their tutors performed this task by asking questions of the "do you understand?" style, 7% of the time they asked follow-up questions, and occasionally they tried something more complicated. They found near-zero correlation between student achievement and "yes" answers to ***ask-confirmation***-style questions, meaning that a "yes" answer to "do you understand?" is useless for gauging student comprehension. Interestingly, "no" answers were correlated with higher achievement. Thus our experts' habit of only rarely asking these questions might be justified.

The relative frequencies of ***elicits*** vs. ***informs*** would seem to represent a fairly direct measure of how much the teaching style represents active learning as opposed to passive learning [Collins and Stevens 1982]. The use of ***ask-confirmation*** could conceivably be a linguistic tic, as opposed a consequence of tutoring strategy. But the ratio of questions to assertions might reflect a more fundamental difference. Therefore in order to test whether we have found a structural difference we compared only the two primitive dialogue acts of ***elicits*** and ***informs***. The 2×2 contingency table for SV tutoring (Table 3 with the last two rows removed) yields $\chi^2 = 8.8$, which indeed shows that novices and experts differ in their behavior at the 0.01 level of significance.

The ratio of ***elicits*** to the sum of ***informs*** + ***elicits*** is 0.2 for the novices and 0.5 for the experts. The novices are informing a lot more, on average, for each question they ask. We consider this to be our most important result. It confirms our choice in trying to embody a question-asking active learning style within CIRCSIM-Tutor.

Graesser, Person and Magliano [1995] report a similar result in their study of unskilled tutors, listing active student learning among the components which are notably "underdeveloped, defective, or virtually non-existent" in unskilled tutoring.

Another big numerical difference which hints at variance between the novices and the experts is the number of concepts per segment. It takes the novices an average of 11.2 concepts to tutor SV and experts an average of 5.2. Inspection of the examples reveals what seems to be a pattern: the expert tutors tend to teach SV by moving forward in causal reasoning from its determinants. Figure 1 above has just such an example. The novices are different. All four novice tutors are represented among the five segments of novice SV tutoring; they all share a similar pattern. They start by asking a relatively open-ended question, then spend the rest of the segment informing the student about a good many details. Here is an excerpted example. The tutor's last sentence is a good example of ***ask-confirmation***:

tu: Last parameter: SV. Why does that not change?

st: OK, I am a little unsure on this one now.… I think it actually does change, since there is more blood now moving into the arterial system … and the Frank-Starling thing says that SV will go up.

tu: You are correct that there is actually a change in the SV. What happens is that there are three main influences on the SV. Can you remember them?

st: No.

tu: There is preload, the intotropic state (IS), and afterload. Preload can be thought of here as CVP. Afterload can be thought of here as MAP, and we have already determined that…. [*proceeds to talk about MAP, ventricles, fiber lengths, muscle cells, Frank-Starling, a buildup of volume, etc., until finally*...] a decrease in CVP will combine to actually DECREASE stroke volume. Does this make sense, am I going too fast?

st: …

By contrast, the expert's opening question quoted in Figure 1 cuts to the chase: "So if CVP is down what happens to (RAP and) SV?" The large bag of other physiological and anatomical concepts is brought to bear only if needed. Another notable difference in the use of concepts is that the experts use far fewer anatomical concepts than the novices did in tutoring SV, preferring to invoke physiological parameters and their relationships.

Differences in Tutoring CVP

Table 4 shows the all the counts from novice and expert transcript segments for the variable CVP.

	Novice		Expert	
	n	n/seg	n	n/seg
Segments	7		5	
Turns	24	3.4	35	7.0
Elicits	12	1.7	26	5.2
Informs	40	5.7	23	4.6
Acks	12	1.7	17	3.4
Ask-conf	6	0.9	2	0.4
Total prim. acts	70	10.0	68	13.6
Concepts	39	5.6	31	6.2

Table 4. Tutoring Counts for CVP

	Novice	Expert
Elicit	12	26
Inform	40	23
Ack	12	17
Ask-conf	6	2

Table 5. CVP Primitive Dialogue Acts

Again we can build a contingency table of primitive dialogue acts (Table 5). Computing $\chi^2 = 12.6$ shows that the difference between experts and novices is significant at the 0.01 level.

The relative use of ***elicits*** versus ***informs*** follows the same pattern in tutoring CVP as it does in tutoring SV. In tutoring CVP, the ratio of ***elicits*** to the total of ***elicits*** + ***informs*** is 0.23 for novices and 0.53 for experts. Computing $\chi^2 = 9.6$ for the 2×2 contingency table containing only ***elicits*** and ***informs*** shows that the difference in behavior is significant at the 0.01 level. Again we can conclude that the experts engage in a more active style of tutoring.

Looking at the segments of tutoring we isolated from the transcripts, we see that the novice tutoring contains long sequences of explanation. Here is an example, again ending with an instance of ***ask-confirmation***:

tu: You were right in predicting CO to increase and that, in turn, would cause MAP to increase. CO also has an effect on another cv variable. Do you know which one this is?
st: SV
tu: No. SV is one of the determinants of CO, along with HR, but CO does not DIRECTLY affect SV. You are on the right track though. CO is inversely related to CVP. This makes sense because an increase in CO means that blood is going to be removed from the central veins that drain into the heart at an accelerated rate. If an increase in CO causes the central venous volume to decrease then the CVP will decrease as well. Does this make sense to you?

By contrast an expert tutor segment contains many more turns, where the tutors more frequently elicit answers from the student. The average number of turns per CVP tutoring segment for the expert tutors, 7.0, is twice that of the novice tutors. In the CVP tutoring segments, we also note that the experts used 1.9 primitive dialogue acts per turn, as opposed to 2.9 for the novices. This would seem to be indicative of an expert teaching style where the student gets more turns per idea communicated from the tutor.

It is noticeable that the experts appeal to the same concepts as the novices do, (e.g. the volume of blood in the central venous compartment) but ask more questions of the student instead of giving the answer. In contrast to the SV tutoring, here the novices and experts seem to be using the same physiological and anatomical arguments, but the differences are due to the style of tutoring. Contrast the novice tutoring above to this example of expert tutoring. (This is one of the sessions where RAP is being tutored instead of CVP.)

tu: Do you know what determines the value of RAP?
st: No.
tu: RAP is essentially the same as central venous pressure. Does this help?
st: Is right atrial pressure CO x TPR-pa?
tu: No. (CO x TPR is MAP)
st: Need more help.
tu: Sorry about that! The pressure in the right atrium is essentially the same as the pressure in the central venous compartment. This (CVP) is determined by the compliance of the system and the volume of blood in the central venous compartment. What determines that volume?
st: Is the volume determined by venous tone?
tu: No. … [*another, more direct hint is given...*]

Note in the above the tutor's use of an ***ask-confirmation*** dialogue act, in the form of the question "does this help?" The ***ask-confirmation*** occurs when there is still an open question for the student to answer. In the novice excerpts, it often occurs just after the tutor has given an answer.

Comparison of Hierarchical Tutoring Goals

Figure 2 shows a more complete annotation of the excerpt of expert tutoring shown in Figure 1. Here we show the hierarchy of tutorial goals. These goals are not based in the human tutor's beliefs about tutoring; rather, they are structure which we impose on the dialogue in an attempt to model the tutoring behavior.

The highest goal shown, encompassing the entire segment, is ***t-tutors-variable*** for variable SV. Satisfying ***t-tutors-variable*** will entail satisfying several other goals hierarchically underneath it, in this case two tutorial goals ***t-moves-forward*** which establishes the correct value for SV and ***t-explores-anomaly*** which lets the student explain an apparent SV-related anomaly in the causally-related set of predictions. Each of these two can be realized as a sequence of more specific tutoring goal*s*, and so on.

The lowest level of the tutoring hierarchy is occupied by the primitive goals of ***t-informs*** or ***t-elicits***. The primitives for student answers (***s-ans***) and tutor acknowledgments of the student answers (***t-ack***) are not part of the tutoring goal hierarchy, as they cannot possibly be planned ahead of time by the tutor.

When we started this experiment we wanted to compare novice and expert tutoring transcripts on the basis of tutorial goals. It proved difficult to annotate the novice tutors' transcripts using the same style of goal structure we used for the expert tutors. The problem, subjectively, is that

<t-tutors-variable var=SV>
<t-moves-forward>
<t-tutors-consequence from=CVP from-val=down to=SV>
<t-elicits info=value ...>
tu: So, if CVP is down what happens to (RAP and) SV?
<s-ans catg=correct>
st: RAP decreases, and SV must follow suit.
</s-ans>
</t-elicits>
</t-tutors-consequence>
</t-moves-forward>
<t-explores-anomaly info=HR-vs-SV>
<t-presents-anomaly>
<t-informs info=...>
tu: So, in DR HR is up, CO is up, but SV is down.
</t-informs>
</t-presents-anomaly>
<t-tutors-anomaly>
<t-elicits ...>
tu: How is this possible? What must be true if all three of these predictions are correct?
<s-ans catg=correct>
st: That HR increases outstrip SV decreases in this case.
</s-ans>
<t-ack>
tu: Exactly!
</t-ack>
</t-elicits>
</t-tutors-anomaly>
</t-explores-anomaly>
</t-tutors-variable>

Figure 2. Excerpt of Expert Tutoring SV with Annotation of Hierarchical Tutoring Goals

the novices often seemed less organized in their tutoring behavior. An example of the problem is shown here, where the novice tutor is attempting to teach the variable SV but because of an errant student answer the dialogue veers off into a variable EDV, which is never related by the tutor to the topic at hand:

tu: Okay, you predicted SV after TPR. What factors determine the stroke volume?
st: SV is increased with increased IS and with increased EDV--preload
tu: Which of those caused SV to remain unchanged? Or to be more direct, you predicted IS wouldn't change. Has EDV changed?
st: Don't know. Could have increased venous return because of increase in CO but….
tu: Okay, there are 2 points we need to clear up. Let's finish up EDV and then talk about the relationship between CO and venous return. …[*then tutors EDV*]

After finishing EDV the tutor proceeds to a discussion of variable CO without first explaining the eventual relevance it will turn out to have to SV, the topic under discussion.

The novice tutors were often using stratagems similar to the ones the experts use. However their style of dialogue, with various topical excursions, was difficult to annotate with the hierarchy of tutorial goals from the expert transcripts. As we continue to analyze the differences in tutoring styles we think that further effort here will be fruitful.

Conclusions

We compared transcripts of both novice and expert tutors teaching the same baroreceptor reflex problem, counting various discourse features and the variety of physiological concepts invoked. In order to make a fair comparison of the counts, we compared segments of text which tutored the same topics to conclusion, regardless of length. We performed this experiment on the teaching of two different physiological variables.

The most clear result is that the ratio of eliciting acts (asking questions of the student) to informing acts (telling things to the student) is much higher for experts than for novices. This indicates, we believe, a more active learning experience. Another difference is that the novices were much more likely than experts to ask questions similar to "do you understand?"

In the teaching of one variable, the novices used noticeably fewer dialogue turns than the experts did. This seemed to be an effect of their style of informing more and eliciting less.

We would like to compare the novices and experts on the basis of annotations of tutoring strategies and tactics with which we have marked up our expert transcripts. But novice tutor dialogues were less organized, so that effort needs to be tried again in a different way.

References

Brandle, Stefan. 1999. "A Framework for Issuing Acknowledgments in Tutorial Discourse." Forthcoming.

Brandle, Stefan. 1998. Using Joint Actions to Explain Acknowledgments in Tutorial Discourse: Application to Intelligent Tutoring Systems. Ph.D. dissertation, Dept. of Computer Science, Illinois Institute of Technology.

Brandle, Stefan and Martha W. Evens. 1997. Acknowledgments in Tutorial Dialogue. In Eugene Santos, ed.., *Eighth Midwest AI and Cognitive Science Conference, Dayton, OH*, pp. 13–17. Menlo Park, CA: AAAI Press, AAAI Technical Report CF–97–01,

Collins, Allan and Albert L. Stevens. 1982. Goals and Strategies of Inquiry Teachers. In Glaser, Robert, ed.., *Advances in Instructional Psychology*, v. 2, pp. 65–119. Hillsdale, NJ: Lawrence Erlbaum.

Evens, Martha W., John Spitkowsky, Patrick Boyle, Joel A. Michael and Allen A. Rovick. 1993. Synthesizing Tutorial Dialogues. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, Boulder, CO*. Hillsdale, NJ: Lawrence Erlbaum.

Freedman, Reva, Yujian Zhou, Michael Glass, Jung Hee Kim, and Martha W. Evens. 1998a. Using Rule Induction to Assist In Rule Construction for a Natural-Language Based Intelligent Tutoring System. In *Twentieth Annual Conference of the Cognitive Science Society, Madison*, pp. 362–367. Hillsdale, NJ: Lawrence Erlbaum.

Freedman, Reva, Yujian Zhou, Jung Hee Kim, Michael Glass, and Martha W. Evens. 1998b. SGML-Based Markup as a Step toward Improving Knowledge Acquisition for Text Generation. In AAAI 1998 Spring Symposium: Applying Machine Learning to Discourse Processing. Menlo Park, CA: American Association for the Advancement of Artificial Intelligence.

Graesser, Arthur C., Natalie K. Person, and Joseph P. Magliano. 1995. Collaborative Dialogue Patterns in Naturalistic One-to-One Tutoring. Applied Cognitive Psychology, v. 9, pp. 495–522.

Hume, Gregory, Joel A. Michael, Allen A. Rovick, and Martha W. Evens. 1996a. Hinting as a Tactic in One-On-One Tutoring. *Journal of the Learning Sciences*, v. 5, no. 1, pp. 23–47.

Hume, Gregory, Joel Michael, Allen Rovick, and Martha W. Evens. 1996b The Use of Hints by Human and Computer Tutors: the Consequences of the Tutoring Protocol. In Proceedings of the 2nd International Conference on the Learning Sciences, Evanston, IL.

Kim, Jung Hee, Reva Freedman, and Martha W. Evens. 1998a. Relationship Between Tutorial Goals and Sentence Structure in a Corpus of Tutoring Transcripts. In *Ninth Midwest AI and Cognitive Science Conference, Dayton OH*, pp. 124–131. Menlo Park, CA: AAAI Press, Technical report No. CF–98–01.

Kim, Jung Hee, Reva Freedman, and Martha W. Evens. 1998b. Responding to Unexpected Student Utterances in CIRCSIM-Tutor v. 3: Analysis of Transcripts. In *Proceedings of the Eleventh Florida Artificial Intelligence Research Symposium, Sanibel Island, FL*, pp. 153–157. Menlo Park, CA: AAAI Press.

Li, Jun, Jai Seu, Martha W. Evens, Joel A. Michael, and Allen A. Rovick. 1992. Computer Dialogue System (CDS): A System for Capturing Computer-Mediated Dialogue. *Behavior Research Methods, Instruments, & Computers*, v. 24, no. 4, pp. 535–540.

Michael, Joel A. 1993. Teaching Problem Solving in Small Groups. *Annals of the New York Academy of Sciences*, v. 701, pp. 37–48. Special Issue on Promoting Active Learning in the Classroom.

Rovick, Allen A. and Joel A. Michael. 1992. The Predictions Table: A Tool for Assessing Students' Knowledge. *American Journal of Physiology*, v. 263, no. 6, part 3 (also known as *Advances in Physiology Education*, v. 8) pp. S33–S36.

Sanders, Gregory A., Martha W. Evens, Allen A. Rovick, and Joel A. Michael. 1992. An analysis of how Students Take the Initiative in Keyboard-to-Keyboard Tutorial Dialogues in a Fixed Domain. In *Proceedings of the Fourteenth Annual Meeting of the Cognitive Science Society, Bloomington, IN*, pp. 1086–1091. Hillsdale, NJ: Lawrence Erlbaum.

Seu, Jai, Ru-Charn Chang, Jun Li, Martha W. Evens, Joel A. Michael and Allen A. Rovick. 1991. Language Differences in Face-to-Face and Keyboard-to-Keyboard Tutoring Sessions. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, Chicago*, pp. 576–580. Hillsdale, NJ: Lawrence Erlbaum.

Shah, Farhana. 1997. Recognizing and Responding to Student Plans in an Intelligent Tutoring System: CIRCSIM-Tutor. Ph.D. dissertation, Dept. of Computer Science, Illinois Institute of Technology.

Spitkowsky, John and Martha W. Evens. 1993. Negative Acknowledgments in Natural Language Tutoring Systems. In Proceedings of the Fifth Midwest Artificial Intelligence and Cognitive Science Society Conference, Carbondale, IL, pp. 41–45.