

# Learner Characteristics and Feedback in Tutorial Dialogue

**Kristy Elizabeth  
Boyer<sup>a</sup>**

**Robert  
Phillips<sup>ab</sup>**

**Michael D.  
Wallis<sup>ab</sup>**

**Mladen A.  
Vouk<sup>a</sup>**

**James C.  
Lester<sup>a</sup>**

<sup>a</sup>Department of Computer Science, North Carolina State University

<sup>b</sup>Applied Research Associates, Inc.

Raleigh, North Carolina, USA

{keboyer, rphilli, mdwallis, vouk, lester}@ncsu.edu

## Abstract

Tutorial dialogue has been the subject of increasing attention in recent years, and it has become evident that empirical studies of human-human tutorial dialogue can contribute important insights to the design of computational models of dialogue. This paper reports on a corpus study of human-human tutorial dialogue transpiring in the course of problem-solving in a learning environment for introductory computer science. Analyses suggest that the choice of corrective tutorial strategy makes a significant difference in the outcomes of both student learning gains and self-efficacy gains. The findings reveal that tutorial strategies intended to maximize student motivational outcomes (e.g., self-efficacy gain) may not be the same strategies that maximize cognitive outcomes (i.e., learning gain). In light of recent findings that learner characteristics influence the structure of tutorial dialogue, we explore the importance of understanding the interaction between learner characteristics and tutorial dialogue strategy choice when designing tutorial dialogue systems.

## 1 Introduction

Providing intelligent tutoring systems (ITSs) with the ability to engage learners in rich natural language dialogue has been a goal of the ITS community since the inception of the field. Tutorial dialogue has been studied in the context of a num-

ber of systems devised to support a broad range of conversational phenomena. Systems such as CIRCSIM (Evens and Michael 2006), BEETLE (Zinn et al. 2002), the Geometry Explanation Tutor (Aleven et al. 2003), Why2/Atlas (VanLehn et al. 2002), ITSpoke (Litman et al. 2006), SCOT (Pon-Barry et al. 2006), ProPL (Lane and VanLehn 2005) and AutoTutor (Graesser et al. 2003) support research that has begun to see the emergence of a core set of foundational requirements for mixed-initiative natural language interaction that occurs in the kind of tutorial dialogue investigated here. Moreover, recent years have witnessed the appearance of corpus studies empirically investigating speech acts in tutorial dialogue (Marineau et al. 2000), dialogues' correlation with learning (Forbes-Riley et al. 2005, Core et al. 2003, Rosé et al. 2003, Katz et al. 2003), student uncertainty in dialogue (Liscombe et al. 2005, Forbes-Riley and Litman 2005), and comparing text-based and spoken dialogue (Litman et al. 2006).

Recent years have also seen the emergence of a broader view of learning as a complex process involving both cognitive and affective states. To empirically explore these issues, a number of ITSs such as AutoTutor (Jackson et al. 2007), Betty's Brain (Tan and Biswas 2006), ITSpoke (Forbes-Riley et al. 2005), M-Ecolab (Rebolledo-Mendez et al. 2006), and MORE (del Soldato and Boulay 1995) are being used as platforms to investigate the impact of tutorial interactions on affective and motivational outcomes (e.g., self-efficacy) along with purely cognitive measures (i.e., learning gains). A central problem in this line of investigation is iden-

tifying tutorial strategies (e.g., Graesser et al. 1995) that can appropriately balance the tradeoffs between cognitive and affective student outcomes (Lepper et al. 1993). While a rich set of cognitive and affective tutorial strategies is emerging (e.g., Porayska-Pomsta et al. 2004), the precise nature of the interdependence between these types of strategies is not well understood. In addition, it may be the case that different populations of learners engage in qualitatively different forms of dialogue. Students with particular characteristics may have specific dialogue profiles, and knowledge of such profiles could inform the design of tutorial systems whose strategies leverage the characteristics of the target population. The extent to which different tutorial strategies, and specific instances of them in certain contexts, may be used to enhance tutorial effectiveness is an important question to designers of ITSs.

Given that human-human tutorial dialogue offers a promising model for effective communication (Chi et al. 2001), our methodology is to study naturally occurring tutorial dialogues in a task-oriented learning environment to investigate the relationship between the structure of tutorial dialogue, the characteristics of learners, and the impact of cognitive and motivational corrective tutorial strategies on learning and self-efficacy (Boyer et al. in press). A text-based dialogue interface was incorporated into a learning environment for introductory computer science. In the environment, students undertook a programming task and conversed with human tutors while designing, implementing, and testing Java programs.

The results of the study suggest that the choice of corrective tutorial strategy has a significant impact on the learning gains and self-efficacy of students. These findings reinforce those of other studies (e.g., Lepper et al. 1993, Person et al. 1995, Keller et al. 1983) that indicate that some cognitive and motivational goals may be at odds with one other because a tutorial strategy designed to maximize one set of goals (e.g., cognitive goals) can negatively impact the other. We contextualize our findings in light of recent results that learner characteristics such as self-efficacy influence the structure of task-oriented tutorial dialogue (Boyer et al. 2007), and may therefore produce important interaction effects when considered alongside tutorial strategy.

This paper is organized as follows. Section 2 describes the corpus study, including experimental design and tagging of dialogue and student problem-solving actions. Section 3 presents analysis and results. Discussion and design implications are considered in Section 4, and concluding remarks follow in Section 5.

## 2 Corpus Study

The corpus was gathered by logging text-based dialogues between tutors and novice computer science students. The learning task was to complete a Java programming problem that required students to apply fundamental concepts such as iteration, modularization, and sequential-access data structures. This study was conducted to compare the impact of certain corrective cognitive and motivational tutorial strategies on student learning and self-efficacy in human-human tutoring. Specifically, the study considered the motivational strategies of praise and reassurance (Lepper et al. 1993) and the category of informational tutorial utterances termed cognitive feedback (Porayska-Pomsta et al. 2004, Tan and Biswas 2006) that followed questionable student problem-solving action. Following the approach of Forbes-Riley (2005) and others (Marineau et al. 2000), utterances from a corpus of human-human tutorial dialogues were annotated with dialogue acts. Then, adopting the approach proposed by Ohlsson et al. (2007), statistical modeling techniques were employed to quantify the relative impact of these different tutorial strategies on the outcomes of interest (in this case, learning and self-efficacy gains).

### 2.1 Experimental Design

Subjects were students enrolled in an introductory computer science course and were primarily freshman or sophomore engineering majors in disciplines such as mechanical, electrical, and computer engineering.

The corpus was gathered from tutor-student interactions between 43 students and 14 tutors during a two-week study. Tutors and students were completely blind to each other's characteristics as they worked together remotely from separate labs. Tutors observed student problem-solving actions

(e.g., programming, scrolling, executing programs) in real time. Tutors had varying levels of tutoring experience, and were not instructed about specific tutorial strategies.

Subjects first completed a pre-survey including items about self-efficacy, attitude toward computer science, and attitude toward collaboration. Subjects then completed a ten item pre-test over specific topic content. The tutorial session was controlled at 55 minutes for all subjects, after which subjects completed a post-survey and post-test containing variants of the items on the pre-versions.

## 2.2 Problem-Solving Tagging

The raw corpus contains 4,864 dialogue moves: 1,528 student utterances and 3,336 tutor utterances. As a chronology of tutorial dialogue interleaved with student problem-solving (programming) actions that took place during the tutoring sessions, the corpus contains 29,996 programming keystrokes and 1,277 periods of scrolling – all performed by students. Other problem-solving actions, such as opening and closing files or running the program, were sparse and were therefore eliminated from the analyses. Of the 3,336 tutor utterances, 1,243 occur directly after “questionable” student problem-solving action. (The notion of “questionable” is defined below.) This subset of tutorial utterances serves as the basis for the tutorial strategy comparison.

Student problem-solving actions were logged throughout tutoring sessions. Two actions were under consideration for the analysis: typing in the programming interface and scrolling in the program editor window. To interpret the raw logged student problem-solving actions, these events were automatically tagged using a heuristic measure for correctness: if a problem-solving action was a programming keystroke (character) that survived until the end of the session, this event was tagged *promising*, to indicate it was probably correct. If a problem-solving act was a programming keystroke (character) that did not survive until the end of the session, the problem-solving act was tagged *questionable*. Both these heuristics are based on the observation that in this tutoring context, students solved the problem in a linear fashion and tutors did not allow students to proceed past a step that

had incorrect code in place. Finally, periods of consecutive scrolling were also marked *questionable* because in a problem whose entire solution fits on one printed page, scrolling was almost uniformly undertaken by a student who was confused and looking for answers in irrelevant skeleton code provided to support the programming task.

## 2.3 Dialogue Act Tagging

Because utterances communicate through two channels, a cognitive channel and a motivational/affective channel, each utterance was annotated with both a required cognitive dialogue tag (Table 1) and an optional motivational/affective dialogue tag (Table 2). While no single standardized dialogue act tag set has been identified for tutorial dialogue, the tags applied here were drawn from several schemes in the tutorial dialogue and broader dialogue literature. A coding scheme for tutorial dialogue in the domain of qualitative physics influenced the creation of the tag set (Forbes-Riley et al. 2005), as did the four-category scheme (Marineau et al. 2000). A more expansive general dialogue act tag set also contributed commonly occurring acts (Stolcke et al. 2000). The motivational tags were drawn from work by Lepper (1993) on motivational strategies of human tutors.

Table 1 displays the cognitive subset of this dialogue act tag set, while Table 2 displays the motivational/affective tags. It should be noted that a cognitive tag was required for each utterance, while a motivational/affective tag was applied only to the subset of utterances that communicated in that channel. If an utterance constituted a strictly motivational/affective act, its cognitive channel was tagged with EX (EXtra-domain) indicating there was no relevant cognitive content. On the other hand, some utterances had both a cognitive component and a motivational/affective component. For example, a tutorial utterance of, “That looks great!” would have been tagged as positive feedback (PF) in the cognitive channel, and as praise (P) in the motivational/affective channel. In contrast, the tutorial move “That’s right,” would be tagged as positive feedback (PF) in the cognitive channel and would not be annotated with a motivational/affective tag. Table 3 shows an excerpt from the corpus with dialogue act tags applied.

Table 1: Cognitive Channel Dialogue Acts

Act	Description	Tutor and Student Example Utterances	Average Count Per Session (Standard Deviation)	
			Student	Tutor
Question (Q)	Questions about goals to pursue, domain concepts, etc.	“Where should we start?” “How do I declare an array?”	6.5 (4.2)	0.6 (0.8)
Evaluative Question (EQ)	Questions that explicitly inquire about student knowledge state or correctness of problem-solving action.	“Do you know how to declare an array?” “Is that right?”	9.7 (7.1)	7.0 (5.0)
Statement (S)	Declarative assertion.	“You need a closing bracket there.” “I am looking for where this method is declared.”	4.9 (4.1)	46.2 (21.0)
Acknowledgement (ACK)	Positive acknowledgement of a previous statement.	“Okay.” or “Yeah.” “Alright.”	3.8 (5.0)	2.5 (1.9)
Extra Domain (EX)	A statement not related to the computer science discussion.	“Hello” or “You’re Welcome” “Can I use my book?”	1.0 (2.1)	1.1 (2.4)
Positive Feedback (PF)	Unmitigated positive feedback regarding problem solving action or student knowledge state.	“Yes, I know how to declare an array.” “That is right.”	2.7 (2.5)	12.0 (7.6)
Lukewarm Feedback (LF)	Partly positive, partly negative feedback regarding student problem solving action or student knowledge state.	“Sort of.” “You’re close.” or “Well, almost.”	0.7 (1.2)	2.3 (2.5)
Negative Feedback (NF)	Negative feedback regarding student problem solving action or student knowledge state.	“No.” “Actually, that won’t work.”	2.1 (1.8)	1.3 (2.1)

The entire corpus was tagged by a single human annotator, with a second tagger marking 1,418 of the original 4,864 utterances. The resulting kappa statistics were 0.76 in the cognitive channel and 0.64 in the motivation channel.

### 3 Analysis and Results

Overall, these tutoring sessions were effective: they yielded learning gains (difference between posttest and pretest) with mean 5.9% and median 7.9%, which were statistically significant ( $p=0.038$ ), and they produced self-efficacy gains

Table 2: Motivational/Affective Channel Dialogue Acts

Act	Description	Tutor and Student Example Utterances	Average Count Per Session (Standard Deviation)	
			Student	Tutor
Confusion (C)	Explicit expression of confusion. Indicates disorientation beyond that indicated by negative feedback (which indicates the student lacks a particular piece of knowledge).	"I have no idea what to do." "I'm lost."	0.8 (1.2)	-
Frustration (F)	Explicit expression of frustration.	"Grrr!" "This is so frustrating."	0.1 (0.4)	0.0 (0.3)
Excitement (E)	Explicit expression of excitement.	"Sweet!" "Cool!"	0.4 (0.7)	0.3 (0.6)
Praise (P)	Statement intended to emphasize a student's success. This goes beyond positive feedback, which serves as factual confirmation of correctness.	"Great job on that part!" "That's perfect."	-	4.2 (5.7)
Reassurance (R)	Statement intended to minimize a student's failure.	"That part was hard." "Don't worry about it."	0.4 (0.6)	1.3 (1.6)
Other Emotion (O)	Utterance that conveys affective or motivational content but for which there is no pre-defined tag.	"Ha ha." "I'm sorry."	0.7 (1.0)	1.5 (2.1)

(difference between pre-survey and post-survey scores) with mean 12.1% and median 12.5%, which were also statistically significant ( $p < 0.0001$ ). Analyses revealed that statistically significant relationships hold between tutorial strategy and learning, as well as between tutorial strategy and self-efficacy gains.

### 3.1 Analysis

First, the values of learning gain and self-efficacy gain were grouped into binary categories ("Low", "High") based on the median value. We then applied multiple logistic regression with the gain category as the predicted value. Tutorial strategy, incoming self-efficacy rating, and pre-test score

were predictors in the model. The binarization approach followed by multiple logistic regression was chosen over multiple linear regression on a continuous response variable because the learning instruments (10 items each) and self-efficacy questionnaires (5 items each) yielded few distinct values of learning gain, meaning the response variable (learning gain and self-efficacy gain, respectively) would not have been truly continuous in nature. Logistic regression is used for binary response variables; it computes the odds of a particular outcome over another (e.g., "Having high learning gain versus low learning gain") given one value of the predictor variable over another (e.g., "The corrective tutorial strategy chosen was positive cognitive feedback instead of praise").

Table 3: Dialogue Excerpts

Utterance	Cognitive Tag	Motivational/Affective Tag
<b>Tutor:</b> Is there a way you could only write the code to extract the digit once but have it go through each of the five digits?	EQ	
<b>Student:</b> Create a loop?	EQ	
<b>Tutor:</b> Yes!	PF	P
<b>Student:</b> Would a for loop be best?	EQ	
<b>Tutor:</b> Yes	PF	
<b>Tutor:</b> What are the three things we need for a loop?	EQ	
<b>Student:</b> Conditions, what we want it to do, then what increments you want it to increase by.	S	

### 3.2 Results

After accounting for the effects of pre-test score and incoming self-efficacy rating (both of which were significant in the model with  $p < 0.001$ ), observations containing tutorial encouragement were 56% less likely to result in high learning gain than observations without explicit tutorial encouragement ( $p = 0.001$ ). On the other hand, an analogous model of self-efficacy gain revealed that tutorial encouragement was 57% more likely to result in high self-efficacy gain compared to tutorial responses that had no explicit praise or reassurance ( $p = 0.054$ ). These models suggested that *the presence of tutorial encouragement in response to questionable student problem-solving action may enhance self-efficacy gain but detract from learning gain.*

Another significant finding was that observations in which the tutor used cognitive feedback plus praise were associated with 40% lower likelihood of high learning gain than observations in which the tutor used purely cognitive feedback. No impact was observed on self-efficacy gain. These results suggest that in response to questionable student problem-solving action, *to achieve learning gains, purely cognitive feedback is preferred over cognitive feedback plus praise, while*

*self-efficacy gain does not appear to be impacted either way.*

Among students with low incoming self-efficacy, observations in which the tutor employed a standalone motivational act were 300% as likely to be in the high self-efficacy gain group as observations in which the tutor employed a purely cognitive statement or a cognitive statement combined with encouragement ( $p = 0.039$ ). In contrast, among students with high initial self-efficacy, a purely motivational tactic resulted in 90% lower odds of being in the high self-efficacy gain group. These results suggest that *standalone praise or reassurance may be useful for increasing self-efficacy gain among low initial self-efficacy students, but may decrease self-efficacy gain in high initial self-efficacy students.*

Considering strictly cognitive feedback, positive feedback resulted in 190% increased odds of high student self-efficacy gain compared to the other cognitive strategies ( $p = 0.0057$ ). Positive cognitive feedback did not differ significantly from other types of cognitive strategies in a Chi-square comparison with respect to learning gains ( $p = 0.390$ ). The models thus suggest when dealing with questionable student problem-solving action, *positive cognitive feedback is preferable to other types of cognitive feedback for eliciting self-efficacy gains, but this type of feedback is not*

*found to be better or worse than other cognitive feedback for effecting learning gains.*

## 4 Discussion

The study found that the presence of direct tutorial praise or encouragement in response to questionable student problem-solving action increased the odds that the student reported high self-efficacy gain while lowering the odds of high learning gain. The study also found that, with regard to learning gains, purely cognitive feedback was preferable to cognitive feedback with an explicitly motivational component. These empirical findings are consistent with theories of Lepper et al. (1993) who found that some cognitive and affective goals in tutoring are “at odds.” As would be predicted, the results also echo recent quantitative results from other tutoring domains such as qualitative physics (Jackson et al. 2007) and river ecosystems (Tan and Biswas 2006) that, in general, overt motivational feedback contributes to motivation but cognitive feedback matters more for learning.

Of the corrective tutorial strategies that were exhibited in the corpus, positive cognitive feedback emerged as an attractive approach for responding to plausibly incorrect student problem-solving actions. Responding positively (e.g., “Right”) to questionable student actions is an example of indirect correction, which is recognized as a polite strategy (e.g., Porayska-Pomsta et al. 2004). A qualitative investigation of this phenomenon revealed that in the corpus, tutors generally followed positive feedback in this context with more substantive cognitive feedback to address the nature of the student’s error. As such, the positive feedback approach seems to have an implicit, yet perceptible, motivational component while retaining its usefulness as cognitive feedback.

This study found that explicit motivational acts, when applied as corrective tutorial approaches, had different impacts on different student subgroups. Students with low initial self-efficacy appeared to benefit more from praise and reassurance than students with high initial self-efficacy. In a prior corpus study to investigate the impact of learner characteristics on tutorial dialogue (Boyer et al. 2007), we also found that learners from different populations exhibited significantly different dialogue profiles. For instance, high self-efficacy

students made more declarative statements, or assertions, than low self-efficacy students. In addition, tutors paired with high self-efficacy students gave more conversational acknowledgments than tutors paired with low self-efficacy students, despite the fact that tutors were not made aware of any learner characteristics before the tutoring session. Additional dialogue profile differences emerged between high and low-performing students, as well as between males and females. Together these two studies suggest that learner characteristics influence the structure of tutorial dialogue, and that the choice of tutorial strategy may impact student subgroups in different ways.

## 5 Conclusion

The work reported here represents a first step toward understanding the effects of learner characteristics on task-oriented tutorial dialogue and the use of feedback. Results suggest that positive cognitive feedback may prove to be an appropriate strategy for responding to questionable student problem-solving actions in task-oriented tutorial situations because of its potential for addressing the sometimes competing cognitive and affective needs of students. For low self-efficacy students, it was found that direct standalone encouragement can be used to bolster self-efficacy, but care must be used in correctly diagnosing student self-efficacy because the same standalone encouragement does not appear helpful for high self-efficacy students. These preliminary findings highlight the importance of understanding the interaction between learner characteristics and tutorial strategy as it relates to the design of tutorial dialogue systems.

Several directions for future work appear promising. First, it will be important to explore the influence of learner characteristics on tutorial dialogue in the presence of surface level information about students’ utterances. This line of investigation is of particular interest given recent results indicating that lexical cohesion in tutorial dialogue with low-performing students is found to be highly correlated with learning (Ward and Litman 2006). Second, while the work reported here has considered a limited set of motivational dialogue acts, namely praise and reassurance, future work should target an expanded set of affective dialogue acts to

facilitate continued exploration of motivational and affective phenomena in this context. Finally, the current results reflect human-human tutoring strategies that proved to be effective; however, it remains to be seen whether these same strategies can be successfully employed in tutorial dialogue systems. Continuing to identify and empirically compare the effectiveness of alternative tutorial strategies will build a solid foundation for choosing and implementing strategies that consider learner characteristics and successfully balance the cognitive and affective concerns surrounding the complex processes of teaching and learning through tutoring.

### Acknowledgments

The authors wish to thank Scott McQuiggan and the members of the Intellimedia Center for Intelligent Systems for their ongoing intellectual contributions, and the Realsearch Group at NC State University for extensive project development support. This work was supported in part by the National Science Foundation through Grant REC-0632450, an NSF Graduate Research Fellowship, and the STARS Alliance Grant CNS-0540523. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Support was also provided by North Carolina State University through the Department of Computer Science and the Office of the Dean of the College of Engineering.

### References

Kristy Elizabeth Boyer, Robert Phillips, Michael Wallis, Mladen Vouk, and James Lester. In press. Balancing cognitive and motivational scaffolding in tutorial dialogue. To appear in *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*.  
 Kristy Elizabeth Boyer, Mladen Vouk, and James Lester. 2007. The influence of learner characteristics on task-oriented tutorial dialogue. *Proceedings of AIED*, pp. 127-134. IOS Press.  
 Vincent Aleven, Kenneth R. Koedinger, and Octav Popescu. 2003. A tutorial dialog system to support self-explanation: Evaluation and open questions. *Proceedings of the 11th International Conference on Artificial Intelligence in Education*, pp. 39-46. Amsterdam. IOS Press.

Vincent Aleven, Bruce McLaren, Ido Roll, and Kenneth Koedinger. 2004. Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. J. C. Lester, R. M. Vicari, and F. Paraguaçu (Eds.), *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, pp. 227-239. Berlin: Springer Verlag.  
 Albert Bandura. 2006. Guide for constructing self-efficacy scales. T. Urdan and F. Pajares (Eds.): *Self-Efficacy Beliefs of Adolescents*, pp. 307-337. Information Age Publishing, Greenwich, Connecticut.  
 Michelene T. H. Chi, Nicholas De Leeuw, Mei-Hung Chiu, and Christian LaVancher. 1994. Eliciting self-explanations improves understanding. *Cognitive Science*, 18:439-477.  
 Michelene T. H. Chi, Stephanie A. Siler, Heisawn Jeong, Takashi Yamauchi, and Robert G. Hausmann. 2001. Learning from human tutoring. *Cognitive Science*, 25(4):471-533.  
 Mark G. Core, Johanna D. Moore, and Claus Zinn. 2003. The role of initiative in tutorial dialogue. *Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics*, pp. 67-74.  
 Teresa del Soldato and Benedict du Boulay. 1995. Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education*, 6(4):337-378. Association for the Advancement of Computing in Education, USA.  
 Martha Evens and Joel Michael. 2006. One-on-One Tutoring by Humans and Computers. Mahwah, New Jersey: Lawrence Erlbaum Associates.  
 Kate Forbes-Riley and Diane Litman. 2005. Using bigrams to identify relationships between student certainty states and tutor responses in a spoken dialogue corpus. *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*. Lisbon, Portugal.  
 Kate Forbes-Riley, Diane Litman, Alison Huettner, and Arthur Ward. 2005. Dialogue-learning correlations in spoken dialogue tutoring. Looi, C-k., Mccalla, G., Bredeweg, B., Breuker, J. (Eds.): *Proceedings of AIED*, pp. 225-232. IOS Press.  
 Arthur C. Graesser, George T. Jackson, Eric Mathews, Heather H. Mitchell, Andrew Olney, Mathew Ventura, Patrick Chipman, Donald R. Franceschetti, Xiangen Hu, Max M. Louwerse, Natalie K. Person, and the Tutoring Research Group. 2003. Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, pp. 474-479.  
 Arthur C. Graesser, Natalie K. Person, and Joseph P. Magliano. 1995. Collaborative dialogue patterns in naturalistic one-to-One tutoring. *Applied Cognitive Psychology*, 9(6):495-522. John Wiley & Sons, Ltd.

- G. Tanner Jackson and Art Graesser. 2007. Content matters: An investigation of feedback categories within an ITS. Luckin, R., Koedinger, K. R., Greer, J. (Eds.): *Proceedings of AIED 2007*, 158:127-134. IOS Press.
- Sandra Katz, David Allbritton, and John Connelly. 2003. Going beyond the problem given: How human tutors use post-solution discussions to support transfer. *International Journal of Artificial Intelligence in Education*, 13:79-116.
- John M. Keller. 1983. Motivational design of instruction. Reigeluth, C.M. (Ed.): *Instructional-Design Theories and Models: An Overview of Their Current Status*, pp. 383-429. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.
- H. Chad Lane and Kurt VanLehn. 2005. Teaching the tacit knowledge of programming to novices with natural language tutoring. *Computer Science Education*, 15:183-201.
- Mark R. Lepper, Maria Woolverton, Donna L. Mumme, and Jean-Luc Gurtner. 1993. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. Lajoie, S.P., Derry, S. J. (Eds.): *Computers as Cognitive Tools*, pp. 75-105. Lawrence Erlbaum Associates, Inc., Hillsdale NJ.
- Jackson Liscombe, Julia Hirschberg, and Jennifer J. Venditti. 2005. Detecting certainness in spoken tutorial dialogues. *Proceedings of Interspeech*, 2005.
- Diane J. Litman, Carolyn P. Rosé, Kate Forbes-Riley, Kurt VanLehn, Dumisizwe Bhembe, and Scott Silliman. 2006. Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence in Education*, 16:145-170.
- Johanna Marineau, Peter Wiemer-Hastings, Derek Harter, Brent Olde, Patrick Chipman, Ashish Karnavat, Victoria Pomeroy, Sonya Rajan, Art Graesser, and the Tutoring Research Group. 2000. Classification of speech acts in tutorial dialog. *Proceedings of the Workshop on Modeling Human Teaching Tactics and Strategies of ITS 2000*, pp. 65-71. Montreal, Canada.
- Stellan Ohlsson, Barbara Di Eugenio, Bettina Chow, Davide Fossati, Xin Lu, and Trina C. Kershaw. 2007. Beyond the code-and-count analysis of tutoring dialogues. Luckin, R., Koedinger, K. R., Greer, J. (Eds.): *Proceedings of AIED 2007*, 158:349-356. IOS Press.
- Natalie K. Person, Roger J. Kreuz, Rolf A. Zwaan, and Arthur C. Graesser. 1995. Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction*, 13(2):161-188. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.
- Heather Pon-Barry, Karl Schultz, Elizabeth Owen Bratt, Brady Clark, and Stanley Peters. 2006. Responding to student uncertainty in spoken tutorial dialogue systems. *International Journal of Artificial Intelligence in Education*, 16:171-194.
- Kaška Porayska-Pomsta and Helen Pain. 2004. Providing cognitive and affective scaffolding through teaching strategies: Applying linguistic politeness to the educational context. J.C. Lester, Vicari, R. M., Paragauçu, F. (Eds.): *Proceedings of ITS 2004*, LNCS 3220:77-86. Springer-Verlag Berlin / Heidelberg.
- Genaro Rebolledo-Mendez, Benedict du Boulay, and Rosemary Luckin. 2006. Motivating the learner: an empirical evaluation. Ikeda, M., Ashlay, K. D., Chan, T.-W. (Eds.): *Proceedings of ITS 2006*, LNCS 4053:545-554. Springer Verlag Berlin / Heidelberg.
- Carolyn P. Rosé, Dumisizwe Bhembe, Stephanie Siler, Ramesh Srivastava, and Kurt VanLehn. 2003. The role of why questions in effective human tutoring. Hoppe, U., Verdejo, F., Kay, J. (Eds.): *Proceedings of AIED 2003*, pp. 55-62. IOS Press.
- Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. 2000. *Computational Linguistics*, 26:339-373.
- Jason Tan and Gautam Biswas. 2006. The role of feedback in preparation for future learning: A case study in learning by teaching environments. Ikeda, M., Ashley, K., Chan, T.-W. (Eds.): *Proceedings of ITS 2006*, LNCS 4053:370-381. Springer-Verlag Berlin / Heidelberg.
- Kurt VanLehn, Pamela W. Jordan, Carolyn P. Rosé, Dumisizwe Bhembe, Michael Bottner, Andy Gaydos, Maxim Makatchev, Umarani Pappuswamy, Michael Ringenberg, Antonio Roque, Stephanie Siler, and Ramesh Srivastava. 2002. The architecture of Why2-Atlas: A coach for qualitative physics essay writing. *Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, LNCS 2363:158-167.
- Arthur Ward and Diane Litman. 2006. Cohesion and learning in a tutorial spoken dialog system. *Proceedings of the 19<sup>th</sup> International FLAIRS (Florida Artificial Intelligence Research Society) Conference*. Melbourne Beach, FL.
- Claus Zinn, Johanna D. Moore, and Mark G. Core. 2002. A 3-tier planning architecture for managing tutorial dialogue. *Intelligent Tutoring Systems, Sixth International Conference*. LNCS 2363:574-584. Springer-Verlag, London, UK.