

Expert Tutors' Feedback Is Immediate, Direct, and Discriminating

Sidney D'Mello¹, Blair Lehman¹, and Natalie Person²

¹Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152

²Department of Psychology, Rhodes College, Memphis, TN 38152

[sdmello|balehman]@memphis.edu person@rhodes.edu

Abstract

Feedback is critical in both human and computer tutoring because it has directive, facilitative, and motivational functions. An understanding of the feedback strategies of expert human tutors is essential for ITSs that aspire to model such tutors. Although previous research suggests that expert tutors provide indirect and delayed feedback, methodological concerns limit the generalizability of these findings. In order to alleviate some of these methodological concerns, we conducted a fine grained analysis of the feedback strategies of 10 expert tutors across 50 sessions. We analyzed the likelihood that tutors provide positive, negative, and neutral feedback immediately following students' correct, partially correct, error ridden, vague, or no answers. Our results support the conclusion that expert tutors feedback is direct, immediate, discriminating, and largely domain independent. We discuss the implication of our results for the development of an ITS that aspires to model expert tutors.

Introduction

Over the past 25 years Intelligent Tutoring Systems (ITSs) have emerged as powerful tools to promote active knowledge construction particularly at deeper levels of comprehension (Psotka, Massey, & Mutter, 1988; Sleeman & Brown, 1982). The ITSs that have been successfully implemented and tested have produced learning gains with an average effect size of one sigma, which is roughly equivalent to one letter grade (Corbett, 2001; VanLehn et al., 2007). When compared to classroom instruction and other naturalistic controls, the 1.0 effect sizes obtained by ITSs is superior to the .39 effect for computer-based training, .50 for multimedia, and .40 effect obtained by novice human tutors (Cohen, Kulik, & Kulik, 1982; Corbett, 2001; Dodds & Fletcher, 2004; Wisher & Fletcher, 2004). It is however less than the 2 sigma effect obtained by expert tutors for mathematics in naturalistic contexts (Bloom, 1984). The naturalistic setting is important because ITSs and accomplished tutors have produced equivalent learning gains when face-to-face

communication is replaced with computer-mediated communication (VanLehn et al., 2007).

It might be the case that the 1.0 sigma effect in learning gains represents an upper bound for ITSs that model novice human tutors. These tutors generally have a little more content knowledge than the tutee and have received little to no training on effective pedagogical methods. Novice human tutors do not adhere to ideal tutoring models or employ sophisticated methods or strategies that have been identified in the ITS literature (Graesser, Person, & Magliano, 1995; McArthur, Stasz, & Zmuidzinas, 1990; Person, Graesser, Magliano, & Kreuz, 1994). It might be the case, however, that expert tutors use ideal models and sophisticated strategies. Hence, building ITSs that model the strategies of expert tutors might be the key to cracking the barrier between the 1.0 sigma effect obtained by current ITSs and the 2.0 sigma effect attributed to expert tutors.

Building an ITS that models the strategies of expert tutors at a fine-grained level requires an analysis of the pedagogical models they adhere to, their question asking strategies, their models of student knowledge, their motivational tactics, and how they handle students' errors and misconceptions. One important aspect of the expert tutoring puzzle, and the focus of the current paper, is the nature of their feedback. Our goal is to analyze the feedback strategies of expert human tutors with an eye for integrating any insights gleaned into an ITS modeled after expert tutors.

Nature of Expert Tutors' Feedback

Feedback is critical in both human and computer tutoring because it is *directive* (i.e., tells students what needs to be fixed), *facilitative* (i.e., helps students conceptualize information), and has *motivational* functions (Black & William, 1998; Lepper & Woolverton, 2002; Shute, 2008). Feedback strategies of tutors have received considerable attention from educational researchers, with a handful of meta-analyses devoted exclusively to the effectiveness of feedback as a pedagogical and motivational tool (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Shute, 2008).

Feedback is generally conceived to vary along *directness* and *timing* dimensions. Although most of the studies have focused on novice human tutors, a precious few studies have assessed the nature of expert tutors' feedback (Fox, 1991, 1993; Kulik & Kulik, 1988; Littman, Pinto, & Soloway, 1990; McKendree, 1990). These are briefly described below.

Directness of Feedback

Directness of feedback pertains to the degree to which tutor's explicitly provide negative feedback to student errors and positive feedback to student accomplishments. The claim has been made that expert tutors rarely provide direct feedback to students (Lepper, Aspinwall, Mumme, & Chabay, 1990; Lepper & Woolverton, 2002; Merrill, Reiser, Ranney, & Trafton, 1992). In a corpus of expert tutors (N=2), Lepper and colleagues found the tutors to be indirect in their feedback regardless of whether the feedback was positive or negative. They suggest that good tutors avoid overtly stating negative feedback or even implying that the student has made an error. Direct feedback is replaced with a series of increasingly direct questions in an effort to elicit the correct response from the student. When probed, the expert tutors attributed their indirect style to enhancing motivation and self-efficacy. Surprisingly, Lepper and colleagues have also found positive feedback to be delivered in an indirect style, presumably in an attempt to lessen the evaluative nature of academics.

Similarly, Merrill et al. (1992) compared the strategies of human tutors to computer tutors, finding critical differences in the delivery of feedback. Human tutors were more subtle and flexible in their delivery. They asked probing questions of varying levels of directness to reveal student errors instead of providing direct and diagnostic feedback (Merrill et al., 1992).

Timing of Feedback

The second important feedback dimension pertains to the *timing* of feedback delivery (Shute, 2008). Immediate feedback occurs right after the student has responded to the tutor's question, while delayed feedback would occur sometime later. It should be noted that there are pedagogical advantages to both immediate and delayed feedback and comparisons between the feedback mechanisms have produced mixed results (Shute, 2008).

It has been claimed that the size of student errors and how tutors classify the errors (productive or unproductive) impacts the timing of the feedback (Lepper & Woolverton, 2002; Merrill et al., 1992). Hence, errors that would block students from ever reaching a solution are immediately handled, whereas less threatening errors are monitored carefully but are not immediately addressed. Nevertheless, findings from these studies seem to indicate that expert tutors do not provide direct feedback and imply that they delay their feedback or provide none at all.

Research Goals

In summary, it is generally acknowledged that expert tutors provide indirect and delayed feedback. However, before we accept these conclusions too cavalierly it is important to highlight some methodological problems that threaten our understanding of expert tutoring. First, several of the studies on expert tutoring fail to indicate how many expert tutors were included in the analyses (Aronson, 2002; Fox, 1991, 1993; Merrill et al., 1992). Second, all of the reported studies included six or fewer expert tutors, with the majority including only one or two experts (Glass, Kim, Evens, Michael, & Rovick, 1999; Hay & Katsikitis, 2001; Lajoie, Faremo, & Wiseman, 2001). Third, the same sample of expert tutors is used in multiple studies. For example, the same five tutors are included in the Graesser et al., Jordan and Siler, and VanLehn et al. studies (Graesser, Person, Harter, & Group, 2000; Jordan & Siler, 2002; VanLehn et al., 2004). Putnam's tutors are included in the Merrill et al. studies (Merrill et al., 1992; Putnam, 1987). A fourth problem with these studies is that it is unclear as to what constitutes an expert tutor. In some of the studies, the expert tutors are Ph.D.s with extensive teaching and/or tutoring experience (Evens, Spitkovsky, Boyle, Michael, & Rovick, 1993; Glass et al., 1999; Graesser et al., 2000; Jordan & Siler, 2002), whereas in others the experts are graduate students that work in tutoring centers (Fox, 1991, 1993).

These are some of the problems that warranted an investigation of the feedback strategies of a large expert tutoring corpus that alleviates the aforementioned methodological concerns. In particular, we investigated whether the feedback of 10 expert tutors over 50 naturalistic tutorial sessions was direct and immediate or indirect and delayed. We also investigated if feedback strategies were modulated by domain (math versus science).

Expert Tutoring Corpus

The corpus consisted of 50 tutoring sessions between students and expert tutors on algebra, geometry, physics, chemistry, and biology. The students were all having difficulty in a science or math course and were either recommended for tutoring by school personnel or voluntarily sought professional tutoring help.

The expert tutors were recommended by academic support personnel from public and private schools in a large urban school district. All of the tutors had long-standing relationships with the academic support offices that recommended them to parents and students. The criteria for being an expert tutor were (a) have a minimum of five years of one-to-one tutoring experience, (b) have a secondary teaching license, (c) have a degree in the subject that they tutor, (d) have an outstanding reputation as a private tutor, and (e) have an effective track record (i.e., students who work with these tutors show marked

improvement in the subject areas for which they receive tutoring).

Fifty one-hour tutoring sessions were videotaped and transcribed. To capture the complexity of what transpires during a tutoring interaction, two coding schemes were developed to classify every tutor and student dialogue move (Person, Lehman, & Ozbun, 2007). A total of 47,256 dialogue moves were coded in the 50 hours of tutoring.

The Tutor Coding Scheme consisted of 24 categories inspired by previous tutoring research on pedagogical and motivational strategies and dialogue moves (Cromley & Azevedo, 2005; Graesser et al., 1995; Lepper & Woolverton, 2002). The moves consisted of various forms of information delivery (direct instruction, explanation, example, etc.), questions and cues to get the student to do the talking (hints, prompts, pumps, forced choices, etc.), feedback (positive, negative, neutral), motivational moves (general motivation statement, solidarity statement), humor, and off-topic conversation.

A 16 category coding scheme was also developed to classify all student dialogue moves. Some of the student move categories captured the qualitative nature of a student dialogue move (e.g., correct answer, partially-correct answer, error-ridden answer), whereas others were used to classify student questions and actions (e.g., reading aloud or solving a problem).

Although detailed descriptions of the coding schemes are beyond the scope of this paper, of relevance to this paper is the coding of feedback and answer moves. Student answers were coded as (a) *no* answers (e.g. “Umm.” “Mmm.”), (b) *error ridden* answers (e.g. “Prokaryotes are human and eukaryotes are bacteria”), (c) *vague* answers (e.g. “Because it helps to, umm, you know”), (d) *partial* answers (e.g. “It has to do with the cells”), and (e) *correct* answers (e.g. “In meiosis it starts out the same with one diploid”). These five answer categories comprised 28.6% of all student moves. 14% of students’ answers were correct, 5.7% partial, 4.6% vague, and 2.8% error-ridden. No answers occurred 1.5% of the time.

There were three feedback categories that comprised 15.6% of all tutor moves. The categories were *positive* (e.g., “correct”, “right”, “exactly”), *negative* (e.g., “no.” “uh uh.”), and *neutral* (e.g., “I see”), comprising 12.5%, 1.6%, and 1.5% of tutor moves, respectively.

Four trained judges coded the 50 transcripts on the dialogue move schemes. Cohen’s kappas were computed to determine the reliability of their judgments. The kappa scores were .92 for the tutor moves and .88 for the student moves.

Data Analysis

The analyses began by creating a time series of student and tutor moves for each session. On average, there were 945 moves per time series ($SD = 343$). Time series ranged from 467 to 1870 moves with a median of 925 moves.

We used the likelihood metric (D’Mello, Taylor, & Graesser, 2007) to compute the likelihood of a transition

between any two moves (see Eq. 1). The metric allows us to compute the likelihood of a transition between any two moves after correcting for the base rate of M_{t+1} . It includes a normalization factor (i.e., the denominator) so that any two likelihoods can be compared even if the prior probabilities of the moves differ; i.e., $L(M \rightarrow X)$ and $L(M \rightarrow Y)$ can be compared even if, $\Pr(X) \neq \Pr(Y)$. This comparison is compromised with mere conditional probabilities (i.e., $\Pr(M_{t+1}|M_t)$).

$$L(M_t \rightarrow M_{t+1}) = \frac{\frac{\Pr(M_{t+1} \cap M_t)}{\Pr(M_t)} - \Pr(M_{t+1})}{1 - \Pr(M_{t+1})} \quad (\text{Eq. 1})$$

According to Equation 1, if $L(M_t \rightarrow M_{t+1}) > 0$, we can conclude that move M_{t+1} follows M_t above and beyond the prior probability of experiencing M_{t+1} (i.e., above chance levels). If, on the other hand, $L(M_t \rightarrow M_{t+1}) = 0$, then M_{t+1} follows M_t at the chance level. Furthermore, if $L(M_t \rightarrow M_{t+1}) < 0$, then the likelihood of move M_{t+1} following M_t is lower than the base rate of experiencing M_{t+1} (i.e., below chance).

Transition likelihoods were computed for all possible combination of moves resulting in a 40×40 matrix for each session. Two-tailed one sample t-tests were used to test whether the mean likelihood for any given transition was significantly greater than (excitatory), less than (inhibitory), or equal to zero (no relationship between immediate and next move).

Results

Our data analysis strategy allowed us to assess the likelihoods of four classes of transitions: $L(\text{Tutor} \rightarrow \text{Student})$, $L(\text{Student} \rightarrow \text{Tutor})$, $L(\text{Tutor} \rightarrow \text{Tutor})$, and $L(\text{Student} \rightarrow \text{Student})$. However, the current paper focuses on one particular set of transitions: $L(\text{Student}_{\text{Answer}} \rightarrow \text{Tutor}_{\text{Feedback}})$. Table 1 presents mean transition likelihoods for the 15 $\text{Tutor}_{\text{Feedback}} \rightarrow \text{Student}_{\text{Answer}}$ transitions.

Table 1. Mean transition likelihoods

Student Answer	Tutor Feedback		
	Negative	Neutral	Positive
No	-.007*	.006	-.041
Error	.291**	.038*	-.075**
Vague	.013*	.016	.020
Partial	.016	.067**	.135**
Correct	-.008**	.015*	.359**

* $p < .05$, ** $p < .001$

Direct vs. Indirect Feedback

We examined the patterns of tutor feedback to evaluate whether expert tutors provided direct feedback. Feedback patterns for each of the five answer categories are described below.

Error-Ridden Answers. Tutors provide negative ($d = 1.2$) and neutral ($d = .35$), but not positive ($d = -1.68$) feedback to error-ridden answers. In addition to the one-sample t -tests that independently compare each transition to chance (zero), we also performed a 3-way repeated measures ANOVA to test for differences in the feedback strategies. The ANOVA indicated that the main effect for feedback was significant and quite robust, $F(2, 96) = 65.3$, $Mse = .026$, $p < .001$, partial $\eta^2 = .576$. Bonferroni-posthoc tests revealed the following feedback ordering at the $p < .05$ significance level: Negative > Neutral > Positive.

Partially-Correct Answers. It appears that tutors provide neutral ($d = .52$) and positive ($d = .59$) feedback to partially-correct answers (see Table 1). Negative feedback followed partially-correct answers at chance rates. An ANOVA comparing the likelihood of the three feedback moves given partially-correct answers was significant, $F(2, 98) = 7.15$, $p = .001$, partial $\eta^2 = .127$. Bonferroni post-hoc tests indicated that the likelihoods of neutral and positive feedback following a partially-correct answer were on par and significantly greater than negative feedback.

Correct Answers. Tutors provide positive ($d = 2.04$) and neutral ($d = .35$), but not negative ($d = -1.14$) feedback to correct answers. An ANOVA comparing the likelihood of the three feedback moves after correct answers was significant, $F(2, 98) = 182.7$, $p < .001$, partial $\eta^2 = .789$. Bonferroni-posthoc tests revealed the following ordering of feedback patterns: Positive > Neutral > Negative.

No Answers and Vague Answers. The t -tests indicated that tutors do not provide feedback when students do not provide an answer. A similar pattern is observed for vague answers. Although negative feedback appears to follow vague answers, the effect was quite small ($d = .3$). Furthermore, an ANOVA comparing the three feedback categories when the student provided a vague answer was not significant, $p = .868$. Hence, similar to no answers, it appears tutors do not provide feedback to vague answers.

Immediate vs. Delayed Feedback

We investigated whether tutors provided immediate or delayed feedback by assessing whether a student's answer at time t was more likely to be followed by one of the feedback categories or any other tutor move at $t + 1$ (i.e. the turn immediately following the student's answer). The analyses proceeded by grouping the three feedback moves into one general feedback category and collapsing the remaining 24 tutor moves into a non-feedback category. Paired-sample t -tests then compared the likelihood of a

feedback move versus a non-feedback move immediately following a student answer.

The results indicated that correct answers were significantly more likely to be followed by a feedback move than a non-feedback move ($L_{COR \rightarrow FDB} = .371$, $L_{COR \rightarrow NO\ FDB} = .054$, $p < .001$, $d = 1.41$). Error-ridden answers were also more likely to be followed by feedback compared to another tutor move ($L_{ERR \rightarrow FDB} = .280$, $L_{ERR \rightarrow NO\ FDB} = .105$, $p = .085$, $d = .47$).

In contrast, no answers and vague answers were significantly more likely to be followed by a non-feedback move than a feedback move (no answer: $L_{NO \rightarrow FDB} = -.042$, $L_{NO \rightarrow NO\ FDB} = .646$, $p < .001$, $d = 1.92$; vague answer: $L_{VAG \rightarrow FDB} = .051$, $L_{VAG \rightarrow NO\ FDB} = .370$, $p < .001$, $d = 1.22$).

Finally, feedback and non-feedback moves were equally likely to follow partial answers ($L_{PAR \rightarrow FDB} = .226$, $L_{PAR \rightarrow NO\ FDB} = .242$, $p = .876$, $d = .04$).

Domain Differences in Feedback Profiles

The expert tutoring corpus included sessions on a number of math and science domains such as physics, chemistry, biology, algebra, and geometry. It might be the case that the feedback strategies of the expert tutors are modulated by the tutoring domain. Hence, we analyzed whether feedback profiles differed across the 31 math and 19 science sessions.

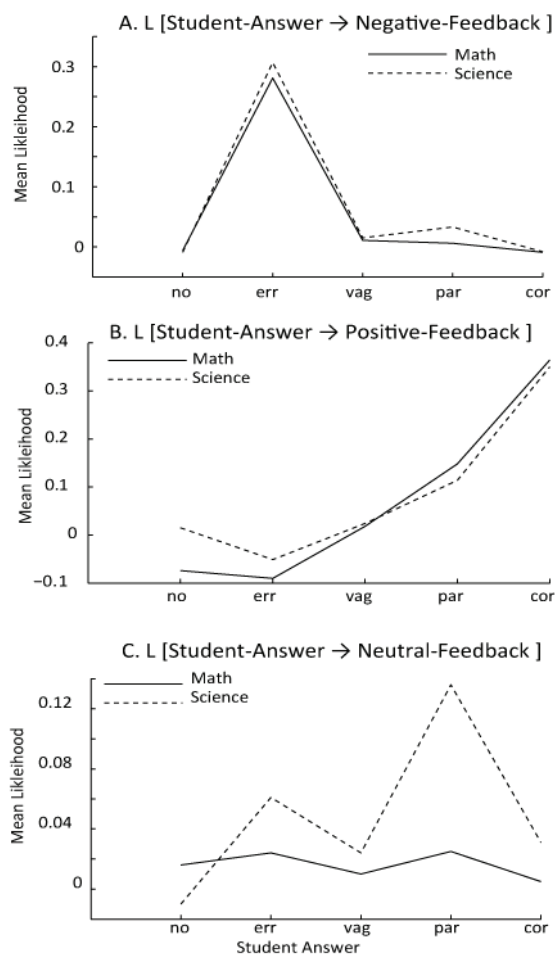
Two 5×2 (*answer* $\times$ *domain*) ANOVAs with *answer* (no, error-ridden, vague, partially-correct, correct) as a within subjects factor and *domain* (math or science) as a between subjects factor did not yield a significant *answer* $\times$ *domain* interaction for negative feedback ($p = .927$) or positive feedback ($p = .323$). Hence, domain differences do not impact the delivery of positive or negative feedback (see Figure 1A and 1B).

However, there was a significant *answer* $\times$ *domain* interaction for neutral feedback, $F(4, 148) = 3.59$, $p = .008$, partial $\eta^2 = .088$. Bonferroni post-hoc tests indicated that there was a significant domain difference in how tutors provided feedback to partially-correct answers but not any of the other answer types. It appears, that tutors are more likely to provide neutral feedback to partially-correct answers in science than math ($L_{MATH} = .013$, $L_{SCI} = .114$, $p = .001$, $d = 1.04$).

General Discussion

Our fine-grained analysis of the feedback strategies of presumably the largest expert tutoring corpus indicates that expert tutors' feedback is *direct*, *immediate*, and *discriminating* at least when students provide error-ridden, partially-correct, and correct answers. Their feedback is direct because they primarily provide negative feedback to incorrect answers, positive feedback to correct answers, and both neutral and positive feedback to partially-correct answers. Their feedback is immediate because compared to any other move it is negative feedback that immediately follows error-ridden answers and positive feedback that

immediately follows correct answers. Furthermore, about half the time positive feedback immediately follows partially-correct answers. Their feedback is discriminating because it is sensitive to differences in answer types (see Figure 1); a strategy not adopted by novice tutors (Person



et al., 1994).

Figure 1. Answer × Domain Interaction

It appears that the feedback strategies of expert tutors transcend domain differences for negative and positive feedback. However, they are more likely to provide neutral feedback to partially-correct answers in science compared to math. This finding is intuitively plausible because when compared to math, science answers are fuzzier because the distinction between correct and partially-correct answers is more subtle.

The expert tutors do not provide feedback when students provide vague answers or do not provide an answer altogether. Expert tutoring feedback is discriminatory and evaluative, hence, one would not expect feedback when the student does not provide an answer or hedges and provides a vague answer. Although not highlighted in this paper, it appears that expert tutors respond to both vague and no answers by either (a) providing the correct answer, (b) simplifying the problem, or (c) providing a hint.

We are currently in the process of developing a tutoring system (Guru) for high school biology based on the tactics, actions, and dialogue of expert human tutors. The pedagogical and motivational strategies of Guru are informed by a detailed computational model of expert human tutoring. The computational model transcends various levels of granularity from tutorial modes (e.g., lectures, modeling, scaffolding), to collaborative patterns of dialogue moves within individual modes (e.g., information-elicitation, information-transmission), to individual dialogue moves (e.g., direct instruction, positive feedback, solidarity statement), to the language, facial expression, intonation, and gestures of tutors. Understanding how expert tutors are direct, immediate, and discriminating with their feedback will guide Guru's feedback strategies. Whether a direct and immediate approach towards feedback will enhance learning compared to more indirect and delayed strategies awaits further technological development and empirical testing.

Acknowledgements

This research was supported by the by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080594 and by the U. S. Office of Naval Research Grant N00014-05-1-0241. The opinions expressed are those of the authors and do not represent views of the funding agencies.

References

- Aronson, J. (2002). *Improving academic achievement: Impact of psychological factors on education*. San Diego: Academic Press.
- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13(2), 111-127.
- Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The Instructional-Effect of Feedback in Test-Like Events. *Review of Educational Research*, 61(2), 213-238.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Bloom, B. (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.
- Cohen, P., Kulik, J., & Kulik, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19(2), 237-248.
- Corbett, A. (2001). *Cognitive computer tutors: Solving the two sigma problem*. Paper presented at the Eighth International Conference on User Modeling.

- Cromley, J., & Azevedo, R. (2005). What Do Reading Tutors Do?: A Naturalistic Study of More and Less Experienced Tutors in Reading. *Discourse Processes*, 40(2), 83-113.
- D'Mello, S., Taylor, R., & Graesser, A. (2007). Monitoring affective trajectories during complex learning. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 203-208). Austin, TX: Cognitive Science Society.
- Dodds, P., & Fletcher, J. (2004). Opportunities for new "smart" learning environments enabled by next-generation web capabilities. *Journal of Educational Multimedia and Hypermedia*, 13(4), 391-404.
- Evens, M., Spitkovsky, J., Boyle, P., Michael, J., & Rovick, A. (1993). *Synthesizing tutorial dialogues*. Paper presented at the Proceedings of the 15th Annual Conference of the Cognitive Science Society, Boulder.
- Fox, B. (1991). Cognitive and interactional aspects of correction in tutoring. In P. Goodyear (Ed.), *Teaching knowledge and intelligent tutoring*. Norwood, NJ: Ablex.
- Fox, B. (1993). *The human tutorial dialogue project*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Glass, M., Kim, J., Evens, M., Michael, J., & Rovick, A. (1999). *Novice vs. expert tutors: A comparison of style*. Paper presented at the Midwest Artificial Intelligence and Cognitive Science Conference, Bloomington, IN.
- Graesser, A., Person, N., Harter, D., & Group, T. T. R. (2000). *Teaching tactics in Autotutor*. Paper presented at the ITS 2000 Proceedings of the Workshop on Modeling Human Teaching Tactics and Strategies, Montreal, Canada.
- Graesser, A., Person, N., & Magliano, J. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9(6), 495-522.
- Hay, P., & Katsikitis, M. (2001). The 'expert' in problem-based and case-based learning: necessary or not? *Medical Education*, 35(1), 22-26.
- Jordan, P., & Siler, S. (2002). *Student initiative and questioning strategies in computer mediated human tutoring dialogues*. Paper presented at the ITS Workshop on Empirical Methods for Tutorial Dialogue Systems.
- Kulik, J., & Kulik, C. (1988). Timing of Feedback and Verbal-Learning. *Review of Educational Research*, 58(1), 79-97.
- Lajoie, S., Faremo, S., & Wiseman, J. (2001). Tutoring strategies for effective instruction in internal medicine. *International Journal of Artificial Intelligence and Education*, 12, 293-309.
- Lepper, M., Aspinwall, L., Mumme, D., & Chabay, R. (1990). Self-perception and social perception processes in tutoring: Subtle social control strategies of expert tutors. In J. Olson & M. Zanna (Eds.), *Self inference and social inference: The Ontario symposium* (Vol. 6, pp. 217-237). Hillsdale, NJ: Erlbaum.
- Lepper, M., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.
- Littman, D., Pinto, J., & Soloway, E. (1990). The knowledge required for tutorial planning: An empirical analysis. *Interactive Learning Environments*, 1, 124-151.
- McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring Techniques in Algebra. *Cognition and Instruction*, 7(3), 197-244.
- McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human Computer Interaction*, 5, 381-413.
- Merrill, D., Reiser, B., Ranney, M., & Trafton, J. (1992). Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *The Journal of the Learning Sciences*, 2(3), 277-305.
- Person, N., Graesser, A., Magliano, J., & Kreuz, R. (1994). Inferring What the Student Knows in One-to-One Tutoring - the Role of Student Questions and Answers. *Learning and Individual Differences*, 6(2), 205-229.
- Person, N., Lehman, B., & Ozburn, R. (2007). *Pedagogical and Motivational Dialogue Moves Used by Expert Tutors*. Paper presented at the 17th Annual Meeting of the Society for Text and Discourse, Glasgow, Scotland.
- Psotka, J., Massey, D., & Mutter, S. (1988). *Intelligent tutoring systems: Lessons learned*. Lawrence Erlbaum Associates.
- Putnam, R. (1987). Structuring and Adjusting Content for Students - a Study of Live and Simulated Tutoring of Addition. *American Educational Research Journal*, 24(1), 13-48.
- Shute, V. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153-189.
- Sleeman, D., & Brown, J. (Eds.). (1982). *Intelligent tutoring systems*. New York: Academic Press.
- VanLehn, K., Graesser, A., Jackson, G., Jordan, P., Olney, A., & Rosé, C. (2004). Natural language tutoring: A comparison of human tutors, computer tutors and text. *Submitted to Cognitive Science*.
- VanLehn, K., Graesser, A., Jackson, G., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31(1), 3-62.
- Wisher, R., & Fletcher, J. (2004). The case for advanced distributed learning. *Information & Security: An International Journal*, 14, 17-25.