How Do They Do It? Investigating Dialogue Moves within Dialogue Modes in Expert Human Tutoring

Blair Lehman[1], Sidney D'Mello[2], Whitney Cade[1], and Natalie Person[3]

[1]Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152
{balehman|wlcade}@memphis.edu

[2]Departments of Psychology and Computer Science, University of Notre Dame
Notre Dame, IN 46556
sdmello@nd.edu

[3]Department of Psychology, Rhodes College, Memphis, TN 38112
person@rhodes.edu

Abstract. Expert human tutors are widely considered to be the gold standard for increasing student learning. While not every student has access to an expert tutor, it is possible to model intelligent tutoring systems after expert tutors. In an effort to achieve this goal, we have analyzed a corpus of 50 hours of one-to-one expert human tutoring sessions. This corpus was coded for speech acts (dialogue moves) and larger pedagogical strategies (dialogue modes). Using mixed-effects modeling, we found that expert tutors differentially used dialogue moves depending on the dialogue mode. Specifically, tutor posed questions, explanations, and motivational statements were predictive of different dialogue modes (e.g., Lecture, Scaffolding).

Keywords: expert tutoring, speech acts, dialogue, ITSs, pedagogical strategies

1 Introduction

Expert human tutors have widely been considered the gold standard for learning, with Bloom [1] reporting a 2 sigma (or approximately 2 letter grade) learning gain over traditional classroom instruction Novice human tutors typically only achieve a gain of 0.4 sigma [2], while intelligent tutoring systems (ITSs) produced a 1 sigma learning gain over traditional classrooms [3]. A recent meta-analysis by VanLehn [4], however, reported a more modest effect for expert tutors (d = .79). Interestingly, ITSs had a comparable impact on learning (d = .76). Despite this more modest learning effect, the pedagogical practices of expert tutors are still effective and there might be advantages associated with building ITSs that model the strategies of expert tutors.

So what exactly are these strategies that make expert tutors so effective? Unfortunately, many of the studies have relied on a small sample (N = 2) and the definition of expert status has varied widely (e.g., [5-6]). For example, college professors [5] and graduate students have been used as expert tutors [6]. These two groups may be considered experts, but there is a lack of consensus on what constitutes expertise.

Despite these concerns, past research has been able to identify some of the strategies that make expert tutors effective. Lepper and Woolverton [7] have suggested that there are three key elements to this effectiveness: individualization, immediacy, and interactivity. Through modeling and monitoring student knowledge, tutors have the ability to dynamically adapt to the needs of individual students [8]. Over the course of a tutoring session, tutors can target the specific knowledge deficits and misconceptions, and construct just-in-time interventions for each student.

Although these broad strategies hint at why expert tutors are effective, a more detailed analysis of tutoring strategies is needed. There is, however, a question about the level of analysis. Past studies have analyzed tutoring sessions at the speech act level [9-10], problem-solving episode [11], and the larger pedagogical context [9, 12]. If the goal is to develop an ITS based on the strategies of expert tutors, it is necessary to understand these strategies at both a fine grained level and a more global level.

There have been few studies that investigated the interplay between different levels of tutorial dialogue [9, 13]. One study attempted to extract the larger pedagogical context from the speech acts of students and novice tutors using Hidden Markov Models [9]. In another study that took a more theory-driven approach, Cade et al. [12] created a dialogue mode coding scheme based on both learning theory (e.g., Modeling-Scaffolding-Fading paradigm, [14]) and observations from a corpus of expert tutoring sessions. Although this study yielded important insights into the strategies of expert tutors, it only considered tutorial dialogue at the mode level.

The present study addresses this issue via a multi-level analysis of tutorial dialogue in a corpus of 50 hours of one-to-one expert tutoring sessions. Previously, this corpus was coded at the dialogue move [10] and dialogue mode levels [12]. In the present paper we investigated the distribution of moves within each mode. Specifically, we sought to answer the question: How are dialogue modes manifested in dialogue moves? We will also test whether it is possible to discriminate between dialogue modes using dialogue moves as features.

2 Expert Tutoring Corpus

The corpus consisted of 50 tutoring sessions between ten expert tutors and 40 students [10, 12]. Expert status was defined as licensed to teach at the secondary level, five or more years of tutoring experience, employed by a professional tutoring agency, and recommended by local school personnel. The students were in middle or high school and having difficulty in a science or math course. All tutor-student pairs were working together prior to the study. Each session lasted approximately one hour.

Tutor-student dialogue was coded at two levels: dialogue moves [10] and dialogue modes [12]. Dialogue moves varied in length from one-word acknowledgements to lengthy explanations. Dialogue modes were longer, pedagogically distinct phases that consisted of both tutor and student contributions over multiple dialogue turns. The dialogue move (*kappa* = .88) and dialogue mode coding schemes (*kappa* = .87) were developed and coded independent of each other.

Tutor dialogue move coding scheme. The 26-item tutor dialogue move coding scheme [10] was divided into groups based on similar functions within the tutoring session: *direct instruction* (example, counterexample, preview, summary, provide correct answer, direct instruction), *question* (new problem, simplified problem, prompt, pump, hint, forced-choice), *feedback* (positive, neutral, negative), *motivational statement* (humor, attribution, general motivation, solidarity), *conversational "Okay"*, and *off-topic*.

Student dialogue move coding scheme. The 16-item student dialogue move coding scheme was divided into eight groups based on the function of each move: *answer* (correct, partially-correct, vague, error-ridden, none), *question* (common ground, knowledge deficit), *misconception*, *metacomment*, *work-related action* (think aloud, read aloud, work silently), *socially motivated action* (social coordination, acknowledge), *gripe*, and *off-topic*.

Dialogue mode coding scheme. An 8-category coding scheme was used to code dialogue modes [12]. The coding scheme for dialogue modes consisted of *Introduction*, *Lecture*, *Clarification*, *Modeling*, *Scaffolding*, *Fading*, *Off-topic*, and *Conclusion*. *Lecture*, for example, involved the tutor explicitly delivering information to the student with fewer student responses, while *Scaffolding* involved collaborative problem solving between the tutor and student.

The present paper focused on *Lecture*, *Clarification*, *Modeling*, *Scaffolding*, and *Fading* because these modes have predominantly pedagogical functions, whereas the remaining dialogue modes (*Introduction*, *Conclusion*, *Off-topic*) involved social and rapport building dialogue [15].

3 Results & Discussion

3.1 Dialogue Moves Predicting Dialogue Modes

Mixed-effects logistic regressions [16] were used to investigate whether dialogue move groups (e.g., feedback) and individual dialogue moves (e.g., positive feedback) could predict the presence (1) or absence (0) of each dialogue mode. Mixed-effects modeling is the recommended analysis for the present data set because of the repeated and nested structures in the data (e.g., moves embedded within modes). There were a total of 47,318 observations (dialogue moves) in the corpus.

In each model, the random effects were the tutor, student, domain (math or science), and order of the dialogue move within the tutoring session. The fixed effects were either move groups or individual moves. Separate models were constructed for tutor and student moves to isolate their independent contributions. For each mode five models were tested: random effects only, move groups (tutor or student), and individual moves (tutor or student). The lme4 package in R [17] was used to perform the requisite computation.

For all modes, models with fixed effects fit the data significantly better than the random effects only models ($p < .001$). Table 1 shows the pattern of significant ($p < .05$) predictors, using move groups as fixed effects. However, instances in which individual moves differed from move groups are discussed below.

Table 1. Dialogue move group patterns for dialogue modes

	Lecture	Clarify	Model	Scaffold	Fade
Tutor Dialogue Move Groups					
Direct Instruction	+	+	+	-*	-
Question	-			+	-
Feedback	-		-	+	+
Motivational Statement			+	-	
Comprehension Gauging Question	+	+	+	-	-
Conversational OK	+		-*	-	
Off-Topic	-	-	-	-	
Student Dialogue Move Groups					
Answer Quality	-	-	-	+	+
Misconception				+	
Metacomment	+			-	
Question	-*			+	
Work-Related Action	-		-	+	+
Socially Motivated Action	+	+	+	-	-
Gripe	-	+*	-		
Off-Topic	-	-	-	-	

*+ = positive predictor; - = negative predictor; blank = non-significant predictor; * = p < 0.1*

Overall, a contrast between a transmission model of learning [18] and a more collaborative interaction was revealed. Specifically, in *Lecture*, *Clarification*, and *Modeling* the tutor provided the majority of information and requested little information from the student. A different pattern emerged for *Scaffolding* and *Fading*. Tutors supplied less information and instead asked questions and provided feedback. Similarly, students asked and answered questions during *Scaffolding*. This profile of *Scaffolding* suggests that students were engaged in problem solving with the guidance of the tutor.

During *Fading*, tutor transmission of information became almost non-existent. Although tutor questions were a negative predictor of *Fading*, posing new problems was a significant positive predictor. For student moves, *Fading* was predicted by answers and work related actions. This suggests that during *Fading,* tutors took on a passive role and allowed students to apply their knowledge. Overall, these findings suggest that there is a connection between these two levels of tutorial dialogue.

3.2 Discriminating between Dialogue Modes

Next, we attempted to discriminate between *Lecture*, *Clarification*, *Modeling*, and *Scaffolding* with dialogue move groups. *Clarification* and *Modeling* were collapsed into one category due to similar pedagogical functions (referred to as *Modeling*). To account for unequal distributions, we downsampled to create more equal mode distributions (*Lecture* = .352; *Modeling* = .322; *Scaffolding* = .325).

Twelve models were tested using dialogue move groups and tutoring session context to discriminate between dialogue modes. Each model was trained and evaluated using discriminant function analyses. Classification accuracy (correct) and kappa scores (see Table 2) were computed using leave-one-out cross validation.

Table 2. Classification results

	Dialogue Move Groups					
	Tutor		Student		Tutor + Student	
Context	*Correct*	*Kappa*	*Correct*	*Kappa*	*Correct*	*Kappa*
None	39.1%	.090	39.5%	.072	43.6%	.147
Move Order	44.5%	.171	43.9%	.154	46.7%	.201
Domain	63.6%	.450	63.2%	.444	63.6%	.451
Domain + Move Order	66.7%	.497	66.7%	.497	67.0%	.503

The results indicated that models combining tutor and student move groups (*kappa* = .147) were the most effective at classifying modes (Tutor + Student). When the context of the tutoring session was added, classification accuracy improved (*kappa* = .503). In particular, inclusion of the tutoring session domain improved performance the most. These findings suggest that the tutoring session context, particularly the domain, is an important element to consider when generating tutorial dialogue.

4 Conclusion

There have been a number of studies investigating the strategies of expert tutors [5-8], but tutorial dialogue has rarely been analyzed at different levels within a single study. In the present paper we examined tutorial dialogue at two levels. While the patterns found were expected based on theories of learning and pedagogy (e.g., [7, 14]), it is important to find evidence that expert tutors actually use these practices. This paper confirmed that some of these 'ideal tutorial strategies' (e.g., Modeling-Scaffolding-Fading) are indeed implemented by more accomplished human tutors.

It is important to briefly consider the implications of our findings for ITSs. ITSs already manage tutorial dialogue at both a local and global level [19] and are effective in achieving learning gains at rates comparable to human tutors [4]. However, the dialogue of most ITSs is informed by learning theories or the practices of novice human tutors, not *expert* human tutors. The present findings can inform ITS dialogues in several important ways. First, expert tutors seem to use a balance of information transmission and collaborative problem solving. Second, the patterns of moves can be used to detect when transitions between modes should occur. Although the present analyses do not address transitions between modes, this has been previously analyzed [12]. Finally, the content of the tutoring session (i.e., domain) seems to have an impact on tutorial dialogue. Future research will need to further examine how strategies differ and under what circumstances different strategies should be deployed to further improve learning.

Acknowledgement. The research reported here was supported by the Institute of Education Sciences (R305A080594) and the U. S. Office of Naval Research (N00014-05-1-0241). The opinions expressed are those of the authors and do not represent views of the IES, the U.S. DoE, the ONR, or DoD.

References

1. Bloom, B.: The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4--16 (1984)
2. Cohen, P., Kulik, J., & Kulik C.: Educational outcomes of tutoring: A meta-analysis of findings. American Educational Research Journal, 19, 237--248 (1982)
3. Corbett, A., Anderson, J., Graesser, A., Koedinger, K., & VanLehn, K.: Third generation computer tutors: Learn from or ignore human tutors? In Proceedings of the 1999 Conference of Computer-Human Interaction (pp. 85--86). ACM Press, New York (1999)
4. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197--221 (2011)
5. Evens, M., Spitkovsky, J., Boyle, P., Michael, J., & Rovick, A.: Synthesizing tutorial dialogues. In Proceedings of the 15th Annual Conference of the Cognitive Science Society (pp. 137--142). Lawrence Erlbaum Associates, Hillsdale (1993)
6. Fox, B.: Cognitive and interactional aspects of correction in tutoring. In Goodyear, P. (Ed.) Teaching knowledge and intelligent tutoring (pp. 149--172). Ablex, Norwood (1991)
7. Lepper, M., & Woolverton, M.: The wisdom of practice: Lessons learned from the study of highly effective tutors. In Aronson, J. (Ed.) Improving academic achievement: Impact of psychological factors on education (pp. 135--158). Academic Press, Orlando (2002)
8. Chi, M., Siler, S., & Jeong, H.: Can tutors monitor students' understanding accurately? Cognition and Instruction, 22(3), 363--387 (2004)
9. Boyer, K., Phillips, R., Ingram, A., Ha, E., Wallis, M., Vouk, M., et al.: Investigating the relationship between dialogue structure and tutoring effectiveness: A Hidden Markov Modeling approach. International Journal of Artificial Intelligence in Education (in press)
10. Person, N., Lehman, B., & Ozbun, R.: 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 (2007)
11. VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W.: Why do only some events cause learning during human tutoring? Cognition and Instruction, 21(3), 209--249 (2003)
12. Cade, W., Copeland, J., Person, N., & D'Mello, S.: Dialogue modes in expert tutoring. In Woolf, B., Aimeur, E., Nkambou, R., & Lajoie, S. (Eds.) Proceedings of 9th International Conference on Intelligent Tutoring Systems (pp. 470--479). Springer, Berlin (2008)
13. D'Mello, S., Olney, A., & Person, N.: Mining collaborative patterns in tutorial dialogues. Journal of Educational Data Mining, 2(1), 1--37 (2010)
14. Rogoff, B., & Gardner, W.: Adult guidance of cognitive development. In Rogoff, B., & Lave, J. (Eds.) Everyday cognition: Its development in social context (pp. 96--116). Harvard University Press, Cambridge (1984)
15. Cade, W., Lehman, B., & Olney, A.: An exploration of off topic conversation. In Burstein, J., Harper, M., & Penn, G. (Eds.) NAACL-HLT 2010 Conference (pp. 669--672). Association for Computational Linguistics (2010)
16. Pinheiro, J., & Bates, D.: Mixed-effects models in S and S-PLUS. Springer Verlag, New York (2000)
17. Bates, D., & Maechler, M.: lme4: Linear mixed-effects models using S4 Classes [computer software]. Available from http://cran.r-project.org/ (2010)
18. Dillon, T.: Questioning and teaching: A manual of practice. Teachers College Press, New York (1988)
19. VanLehn, K.: The behavior of tutoring systems. International Journal of Artificial Intelligence in Education, 16(3), 227--265 (2006)