Tutorial Dialogue Patterns: Expert vs. Non-expert Tutors

Xin Lu, Barbara Di Eugenio, Trina C. Kershaw, Stellan Ohlsson, Andrew Corrigan-Halpern
University of Illinois at Chicago, Chicago IL, USA
{xlu4,bdieugen,tkersh1,ahalpe1,stellan}@uic.edu

Abstract

Studies of one-on-one tutoring have found that expert tutoring is more effective than non-expert tutoring, but the reasons for its effectiveness are relatively unexplored. Since tutoring involves deep natural language interactions between tutor and student, we study the tutorial dialogue patterns by comparing an expert tutor to non-expert tutors. Our results show what behaviors constitute expertise and provide a basis for modelling effective tutorial language in intelligent tutoring systems.

1 Introduction

To enhance interactive learning in Intelligent Tutoring Systems (ITSs), natural language interfaces are used to deliver instructional feedback. With this interface, researchers try to make the ITSs act like real human tutors, especially like expert tutors. Tutors with different levels of expertise may behave differently and have different effects on learning. Some recent research (Chae et al., 2005; Di Eugenio et al., 2006) shows that expert tutors engender much better learning outcomes than non-expert tutors. But it is not yet well understood what makes expert tutoring more effective and which features of tutoring dialogues should be included in interfaces to ITSs. There are two possible reasons why those issues are still under investigation: there are no comprehensive comparisons between expert and non-expert tutors; expert tutors tend to use more complex strategies and language (Glass et al., 1999). Our research aims at exploring the difference between expert tutors and non-expert tutors. In this paper, we focus on the comparison of tutorial interaction patterns between expert and non-expert tutors.

Our tutoring domain concerns extrapolating complex letter patterns (Kotovsky and Simon, 1973), which is a well known task for analyzing human information processing in cognitive science. Students are taught how to solve some problems called "Sequence Extrapolation Problems". This type of problem is composed of a sequence of letters that follow a particular pattern. The student's task is to find the pattern and recreate a sequence with a given starting letter, so the new sequence follows that same pattern. For example, the pattern of the sequence "ABM-CDM" is: "M" as a chunk marker separates the whole sequence into two chunks of letters progressing according to the alphabet. Then with a starting letter "E", to maintain this pattern, the student needs to finish the sequence as "EFMGHM". Only knowledge of the alphabet is required in this domain. We collected dialogues in this domain. During the training session, each student goes through a curriculum of 13 problems of increasing complexity. The training will improve the student's ability in solving letter pattern problems. To test the performance, each student also needs to solve two post-test problems, each with a pattern which is 15 letters long, via a computer interface.

We collected tutoring dialogues with three tutors, one expert, one novice, and one lecturer who is experienced in teaching, but not in one-on-one tutoring. Comparison of the student's performance

showed that the expert tutor was significantly more effective than the other two tutors. We analyzed the individual tutor and student moves independently (Di Eugenio et al., 2006) and found that some behaviors of our tutors do not support the predictions from literatures (Chi et al., 2001). Tutoring is an interaction between tutor and student so tutor moves and student moves are not independent. Our next step was to compare the expert tutor to the non-expert tutors in interaction patterns.

In this paper we first introduce our previous work in study of human tutors including data collection and annotation, and our initial analysis of dialogue moves. Then we study the interaction patterns by comparing expert and non-expert tutors. At last we conclude and discuss future work.

2 Our Previous Work

To investigate the effectiveness of expert tutors, we ran experiments in the letter pattern domain with three different tutors: the expert tutor with years of experience in one-on-one tutoring; the lecturer with years of experience in lecturing but little experience in one-on-one tutoring; the novice tutor with no experience in teaching or tutoring. We also have a control group of students with no tutoring at all. Figure 1 reports the post-test performance of the four groups of student. It shows that the expert tutor is significantly more effective than the other two tutors and control (no tutoring) on both post-test problems. The post-test performance is the average number of letters correct out of a total of 90 letters (in 6 trials, each trial starts from a new letter) for each problem per subject.

The dialogues on two specific problems in the curriculum were transcribed and annotated from the videotapes which recorded the tutors' interaction with the students. For each tutor, six students' dialogues were transcribed and annotated with the tutor and student moves by utterance. The annotation scheme is based on the literature (Chi et al., 2001; Litman et al., 2004). The tutor moves include four high level categories, reaction, initiative, support, conversation. Tutor reaction and initiative are also subcategorized.

- Reaction: the tutor reacts to something the student says or does, which is subcategorized as follows:

  Answering: answering a direct question from the student

  Evaluating: giving feedback about what the student is doing

  Summarizing: summarizing what has been done so far

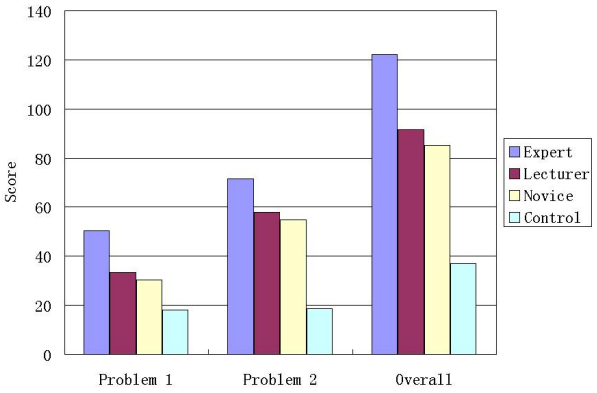

Figure 1: Post-Test Performance

- Initiative is subcategorized as follows:

  Prompting: prompting the student into some kind of activity, further subcategorized as:
  - **General:** laying out what to do next
    Why dont you try this problem
  - **Specific:** trying to get a specific response from the student
    What would the next letter be?

  Diagnosing: trying to determine what the student is doing
  Why did you put a D there?

  Instructing: providing the student with information about the problem. Further subcategorized as:
  - **Declarative:** providing facts about the problem
    Notice the two Cs here? They are separating different parts of the problem
  - **Procedural:** giving hints or tricks about how to solve problem
    Start by counting the number of letters in each period

  Demonstrating: showing the student how to

solve the problem.
Watch this. First I count the number of letters between the G and J here.

- **Support:** the tutor encourages the student in his/her work without referring to particular elements of the problem
- **Conversation:** acknowledgments, continuers, and small talk

Corresponding to the tutor moves, there are six categories in our student moves:

- **Explanation:** explaining what the student said or did, reasoning, or thinking aloud
and see I put them like together.
- **Questioning:** asking the tutor a question
- **Reflecting:** evaluating one's own understanding
I don't really understand about the whole c thing.
- **Reaction:** reacting to something the tutor says, further subcategorized:
  - **Answering:** directly answering a tutor's question
  - **Action Response:** performing some action (e.g., writing down a letter) in response to the tutor's question or prompt
- **Completion:** completing a tutor's utterance
(Tutor: that's right but if you think of these c s as +/) Student: separators right right.
- **Conversation:** same as the one for tutor moves — acknowledgments, continuers, and small talk

Two independent groups, each group with two annotators, coded the tutor moves and the student moves on all the dialogues. The Kappa coefficient is used to evaluate agreement (Carletta, 1996; Di Eugenio and Glass, 2004). After several rounds of annotation, the inter-coder agreement on most of the categories reached an acceptable level (perfect agreement $0.8<\text{Kappa}\leq 1$, or substantial agreement $0.6<\text{Kappa}\leq 0.8$). Table 1 reports the Kappa values for each category of student move. Only the category "completion" is not very reliable because there are only a few cases. The tutor moves are more difficult to be annotated. Some of the categories are not so reliable as those of student moves. The detail Kappa values for tutor moves can be found in (Di Eugenio et al., 2006).

Category	Kappa
Explanation	**0.64**
Questioning	**0.89**
Reflecting	**0.65**
Answering	**0.80**
Action Response	**0.97**
Completion	0.43
Conversation	**0.71**

Table 1: Kappa Values of Student Moves

Category	Novice	Lecturer	Expert
Answering	**10.1**	5.4	1.4
Evaluating	16.4	12.9	7.8
Summarizing	**6.9**	16.7	16.6
General Prompting	4.4	3.3	4.1
Specific Prompting	17.6	**27.7**	13.9
Diagnosing	2.5	3.3	3.3
Declarative Instructing	**22.6**	6.2	4.0
Procedural Instructing	0.6	4.4	**17.2**
Demonstrating	6.3	0.0	11.1
Support	0.6	0.6	5.4
Conversation	9.4	16.9	10.5

Table 2: Percentages of Tutor Moves, by Tutor

Category	Novice	Lecturer	Expert
Explanation	**7.5**	26.3	19.8
Questioning	**18.3**	8.4	6.8
Reflecting	14.2	16.5	13.9
Answering	25	27.1	35.4
Action Response	12.5	10.4	9.7
Completion	0	0.8	0.8
Conversation	22.5	10.6	13.5

Table 3: Percentages of Student Moves, by Tutor

Table 2 and Table 3 report the percentages of tutor and student moves by tutor. After analyz-

ing both the tutor and student moves independently, we found that some behavior of our tutors supports the predictions on effective tutoring from the literatures (Landsberger, 2005; Chi et al., 2001):

- the expert tutor and the lecturer summarize more than the novice;
- students with the expert tutor and the lecturer do more explanations than the students with the novice tutor.

However, some behaviors of the expert tutor are different from the predictions. Compared to the lecturer, the expert tutor does less specific prompting and his students explain less. This contradicts the claim that students learn best when they construct knowledge by themselves, and that as a consequence, the tutor should prompt and scaffold students, and leave most of the talking to them (Chi et al., 2001). This led us to look for other aspects that make the expert tutor more effective. Interestingly, we found that the expert tutor does much more procedural instructing, demonstrating and supporting than the non-expert tutors. Consistently, the novice tutor does much more declarative instructing. Declarative instructing provides facts about the problem. Procedural instructing gives hints to the student about how to solve the problem rather than just providing information. So these moves will be the most interesting features which we are going to look into deeply.

3 Study of Tutorial Interaction Patterns

In order to distinguish the expert tutor from the non-expert tutors, our study of interaction patterns focuses on the following two issues:

Tutor-Student Interaction Patterns: What's the difference between each group of students' behaviors after each type of tutor move?

Student-Tutor Interaction Patterns: How do the expert tutor and the non-expert tutors respond differently to each type of student move?

Table 4 presents a fragment from a transcript of the expert's tutoring. A pair of moves which appear in sequence is an interaction pattern. For example, after the tutor's diagnosing in line 38, the student gives an answer in line 39. This forms a tutor-student interaction pattern — "T–diagnosing + S–answering". Then the tutor does a specific prompting, so line 39 and line 40 form a student-tutor interaction pattern — "S–answering + T–specific prompting". The student's explanations in line 42 and line 45 show that he is explaining his answer in line 39. Totally there are 72 possible types of tutor-student pattern and 72 possible types of student-tutor pattern, which are the combinations of 12 categories of tutor move and 6 categories of student move (For the moment, we left out "Conversation"s in tutor move and student move, since some of them are not so related to expert tutoring.)

First we compared the total number of tutor-student patterns and student-tutor patterns and the

Line No.	Utterances	Annotation
38	**Tutor:** how'd you actually get the *n* in the first place?	Diagnosing
39	**Student:** from here I count from *c* to *g* and then just from *n* to *r*.	Answering
40	**Tutor:** okay so do the *c* to *g*.	Specific Prompting
41	**Tutor:** do it out loud so I can hear you do it.	Specific Prompting
42	**Student:** *c d e f.*	Explanation
43	**Student:** so it's three spaces.	Answering
44	**Tutor:** okay so it's three spaces in between.	Summarizing
45	**Student:** *n o p q* and *r*.	Explanation
46	**Tutor:** okay.	Evaluating
47	**Tutor:** you obviously made a mistake the first time.	Evaluating
48	**Tutor:** one of the more obvious methods would be like just count backwards and double-check everything.	Procedural Instructing
...	...	
56	**Tutor:** questions?	Diagnosing
57	**Student:** uh not really.	Reflecting
...	...	

Table 4: A Transcript Fragment from the Expert's Tutoring

number of pattern types. Table 5 reports the number of interaction patterns and pattern types. Numbers in boldface refer to significant differences (we use Chi-square [1] for the significant test.). We found that in the tutoring dialogues from the novice tutor there are many fewer types of interaction patterns than from the other two tutors; the expert tutor has similar number of pattern types in much fewer interactions than the lecturer. This supports the finding that expert tutors tend to use more complex tutorial strategies and language than novices (Glass et al., 1999).

Interaction Pattern	Novice	Lecturer	Expert
Tutor-Student			
Types	**22**	37	39
Frequency	49	206	128
Ratio	0.45	**0.18**	0.30
Student-Tutor			
Types	**16**	31	38
Frequency	50	205	127
Ratio	0.32	**0.15**	0.30

Table 5: Number of Interaction Patterns and Types, per Tutor

3.1 Tutor-Student Interaction Patterns

We ran Chi-square on the frequencies of all tutor-student interaction patterns. Across all patterns, there are significant differences in student's reactions to tutor moves between the novice tutor and the other two tutors ($p < 0.01$). In each type of pattern that started with a specific tutor move, each group of students reacts significantly differently ($p < 0.05$) to each type of tutor move with the exception of specific prompting. More specifically, we found:

- **Answering:** the novice tutor's answer is followed by student's questioning, not for the other two tutors;
- **Evaluating:** the lecturer's evaluating leads to much more student's explanation but much less reflecting than the expert and novice tutor;
- **Summarizing:** with the novice tutor students almost never react to summarizing; the lecturer's summarizing leads to more student's reflecting; on the contrary, the expert tutor's leads to more student's explanation (e.g. in Table 4, the expert tutor summarizes in line 44 and then in line 45 the student does explanation);
- **General Prompting:** the students with the expert tutor never have questions after his general prompting, but they do with the non-expert tutors;
- **Specific Prompting:** the specific prompts from the expert tutor and the lecturer lead the students to explain much more than for the novice tutor (e.g. in Table 4, the expert tutor does specific prompting in line 41 and then in line 42 the student does explanation); to the tutor's specific prompting, the students with the novice tutor respond with many more questions than with the other tutors;
- **Procedural Instructing:** the lecturer's procedural instructing leads to more reflecting (i.e. assessing one's own understanding); the expert tutor's leads to more explanation;
- **Demonstrating:** with the novice tutor and the lecturer, students hardly react to demonstrating; on the contrary, the expert tutor's demonstrating leads to any kind of student move.
- **Support:** with the novice tutor and the lecturer, students hardly react to support; on the contrary, the expert tutor's support leads to any kind of student move.

Comparing the expert tutor with the lecturer, although he does specific prompting significantly less than the lecturer and his students do less explanation than the lecturer's students, he tends to use more varied strategies to have the students self-explain, instead of just specific prompting. Comparing the expert with the other two tutors, the expert's answering, general and specific prompting must be clearer

[1] Chi square is a non-parametric test of statistical significance. Typically, the hypothesis tested with chi square is whether or not two different samples are different enough in some characteristic or aspect of their behavior that we can generalize from our samples that the populations from which our samples are drawn are also different in the behavior or characteristic.

Tutor Move	Student Move
Summarizing	Explanation
Procedural Instructing	Explanation
Demonstrating	Explanation
Demonstrating	Reflecting
Support	Answering

Table 6: Tutor-Student Interaction Patterns of the Expert Tutor

to the students, since the students have no questions. Also demonstrating and support are the most interesting strategies that make the expert tutor different from the other tutors. Table 6 summarizes the tutor-student interaction patterns in which the expert tutor is different from the non-expert tutors.

3.2 Student-Tutor Interaction Patterns

From the ITS point of view, how the tutor reacts to a student move is more helpful for building a tutorial model. There are significant differences ($p < 0.02$) in tutor's reactions to student moves between all the tutors. Further we analyze the student-tutor interaction patterns in the following two directions:

1. how the tutors react differently to each type of student move;
2. using each type of tutor move, which student moves the tutors react to.

In the first direction we found:

- **Explanation:** the novice tutor uses summarizing much less than the expert tutor and the lecturer; in response to a student's explanation, the lecturer uses specific prompting much more than the other moves and the other tutors;
- **Questioning:** the expert tutor does not answer immediately or directly, but the non-expert tutors do;
- **Reflecting:** the expert tutor uses much more procedural instructing, demonstrating and general prompting;
- **Answering:** the novice uses many fewer specific prompts but much more evaluating and declarative instructing — she immediately delivers the knowledge or the solution;
- **Action Response:** the expert tutor uses much more summarizing and procedural instructing — actions involve procedures, so summarizing and procedural instructing may be more appropriate.

In the second direction (using each type of tutor move, which student moves the tutors react to), we found:

- **Evaluating:** the expert tutor and the lecturer evaluate the student's explanation more than the student's answer and reflecting (e.g. in Table 4, after the student's explanation in line 45 the expert tutor does evaluating in line 46);
- **Summarizing:** the expert tutor and the lecturer summarize more after a student's explanation, reflecting and action response — those involve more information to be summarized;
- **Specific Prompting:** the lecturer does specific prompting after any kind of student move instead of just in response to answering like what the novice and expert tutor do;
- **Diagnosing:** the expert tutor diagnoses after any kind of student move, not just the student's reaction moves (answering and action response);
- **Declarative Instructing:** the expert tutor mostly does declarative instructing after the student's reflecting — only does it when the student directly expresses lack of some concepts;
- **Procedural Instructing:** the expert tutor and the lecturer do more procedural instructing after the student's reflecting;
- **Demonstrating:** the expert tutor does more demonstrating after the student's reflecting, the lecturer never does demonstrating — in this particular domain, demonstration is more useful.

Table 7 summarizes the student-tutor interaction patterns in which the expert tutor is different from the non-expert tutors.

Student Move	Tutor Move
Explanation	Diagnosing
Summarizing	Diagnosing
Reflecting	General Prompting
Reflecting	Declarative Instructing
Reflecting	Procedural Instructing
Reflecting	Demonstrating
Action Response	Summarizing
Action Response	Procedural Instructing

Table 7: Student-Tutor Interaction Patterns of the Expert Tutor

4 Conclusions and Future Work

Our analysis of tutorial dialogue moves, interaction patterns provides plenty of information to distinguish expert from non-expert tutors. The expert tutor is much more effective than the non-expert tutors because of the following behaviors and natural language features:

1. Instead of delivering information directly, demonstrates or models the process for solving the problem (demonstrating, procedural instructing);

2. Finds success, and reinforces effort, in even minor accomplishment (support)— although there are not so many supports in the tutoring dialogues, the expert tutor does it in various situations and much more frequently than the non-expert tutors;

3. Summarizes and reviews (summarizing);

4. Assesses the situation not only after a student's answer or action (diagnosing);

While we were studying the interaction patterns, we observed that not all of tutor's specific prompting are immediately followed by any student move. For example, in Table 4, the expert tutor does specific prompting in line 40 but this specific prompting is followed by another specific prompting, instead of a student turn. We are currently study the difference between expert and non-expert tutors in patterns of multi-utterance turns. This study will enhance our investigation of expert tutoring versus non-expert tutoring.

After highlighting what makes the tutoring expertise, we will be able to model the expert tutoring. With all the dialogues, we will then use machine learning techniques to learn tutorial rules for generating effective natural language feedback in ITSs. We have already developed an baseline ITS to solve the letter pattern problems and did some experiments on the baseline system with different kinds of simple feedback messages (Di Eugenio et al., 2006). The baseline ITS engendered better learning outcomes than the control (no tutoring) but its performance is still far below the expert tutor. So we will embody the tutorial rules in the final version of the letter pattern ITS which is able to deliver more effective feedback.

Finally, our findings on the effectiveness of the expert tutor and features of his tutoring are based on a small dataset, and on one single tutor. They clearly need to be repeated in a larger data set, or with different tutors and / or in different domains. We are transcribing more dialogues in this letter pattern extrapolating domain and also collecting tutoring dialogues in another domain — basic data structure and algorithms. In this introductory computer science domain, students take a pretest, then interact with one of two tutors, then take the post-test. One tutor, the expert, is a retired Math and Computer Science college professor with many years of experience in one-on-one tutoring; the other, the novice, is a senior in Computer Science, with just a few hours under his belt as a volunteer tutor for some introductory classes. Like for the letter pattern domain, we will again compare expert and non-expert tutoring so that we will have a very comprehensive study of expert tutoring. This study will contribute to computationally modelling expert tutoring in ITSs.

5 Acknowledgements

This work is supported by awards CRB S02 and CRB S03 from University of Illinois at Chicago, and by grant N00014-00-1-0640 from the Office of Naval Research.

References

Jean Carletta. 1996. *Assessing agreement on classification tasks: The Kappa statistic*. Computational linguistics, 22(2): 249–254.

Hyeun Me Chae, Jung Hee Kim and Michael Glass. 2005. *Effective Behaviors in a Comparison Between Novice and Expert Algebra Tutors*. Proceedings of Sixteenth Midwest AI and Cognitive Science Conference, Dayton : 25–30.

Michelene TH Chi, Stephanie A. Siler, Heisawn Jeong, Takashi Yamauchi and Robert G. Hausmann. 2001. *Learning from human tutoring*. Cognitive Science, 25(4): 471–533.

Barbara Di Eugenio and Michael Glass. 2004. *The Kappa statistic: a second look*. Computational linguistics, 30(1): 95–101.

Barbara Di Eugenio, Trina C. Kershaw, Xin Lu, Andrew Corrigan-Halpern and Stellan Ohlsson. 2006. *Toward a computational model of expert tutoring*. 19th International conference of Florida Artificial Intelligence Research Society, Melborne Beach, Florida (to appear).

Michael Glass, Jung Hee Kim, Martha W. Evens, Joel A. Michael and Allen A. Rovick. 1999. *Novice vs. expert tutors: a comparison of style*. Tenth Midwest Artificial Intelligence and Cognitive Science Conference, Bloomington, IN.

K. Kotovsky and H. Simon. 1973. *Empirical tests of a theory of human acquisition of information-processing analysis*. British Journal of Psychology, 61: 243–257.

Joe Landsberger. 2005. *Feedback to improve Study Guides*. http://www.studygs.net

Diane J. Litman, Carolyn P. Rose, Kate Forbes-Riley, Kurt Vanlehn, Dumiszwe Bhembe and Scott Silliman. 2004. *Spoken versus typed human and computer dialogue tutoring*. Proceedings of 7th International Conference on Intelligent Tutoring Systems, Alagoas, Brazil.

Claus Zinn and Johanna D. Moore and Mark G. Core. 2002. *A 3-Tier Planning Architecture for Managing Tutorial Dialogue*. Intelligent Tutoring Systems 6th. Intl. Conference, Biarritz, France: 574–584.