Expert vs. Non-expert Tutoring: Dialogue Moves, Interaction Patterns and Multi-Utterance Turns

Xin Lu[1], Barbara Di Eugenio[1], Trina C. Kershaw[2], Stellan Ohlsson[1], Andrew Corrigan-Halpern[1]

[1] University of Illinois at Chicago, Chicago IL, USA
{xlu4,bdieugen,stellan,ahalpe1}uic.edu
[2] University of Massachusetts Dartmouth, North Dartmouth MA,USA
tkershaw@umassd.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 explore the differences between an expert and non-expert tutors through the analysis of individual dialogue moves, tutorial interaction patterns and multi-utterance turns. Our results are a first step showing what behaviors constitute expertise and provide a basis for modeling effective tutorial language in intelligent tutoring systems.

1 Introduction

It has been widely reported that natural language is important to learning. Fox[1] observed that one-on-one tutoring involves a collaborative construction of meaning, a process that arises from a natural language interaction or dialogue between individuals. To enhance interactive learning in Intelligent Tutoring Systems (ITSs), natural language interfaces are used to deliver instructional feedback. With such an 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[2][3] shows that expert tutors engender better learning outcomes than non-expert tutors. This means that a computational model of expert tutoring will improve the effectiveness of ITSs. 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[4]. Our research aims at exploring the difference between expert tutors and non-expert tutors from the natural language point of view.

Our tutoring domain concerns extrapolating complex letter patterns[5], which is a well known task for analyzing human information processing in cognitive science. Given a sequence of letters that follows a particular pattern, the student is asked to find the pattern and create a new sequence from a new starting letter. For example, the pattern of the sequence "ABMCDM" 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 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[3] and found that some behaviors of our tutors do not support the predictions from literatures[6]. Tutoring is an interaction between tutor and student so tutor moves and student moves are not independent. And also tutors are likely to use more than one utterance in a single turn. Our next step was to compare the expert tutor to the non-expert tutors in interaction patterns and multi-utterance turns.

In this paper we first introduce our previous work in study of human tutors including data collection and annotation, and our analysis of dialogue moves. Then we study the interaction patterns and multi-utterance turns 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. There are 11 students in each group, who are all psychology majored freshmen and native speakers in English. Comparing the post-test performance of the four groups of student shows that the expert tutor is significantly more effective than the other two tutors and control (no tutoring) on both post-test problems[3]. 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 subjects. For each tutor, six subjects' dialogues were transcribed and anno-

tated with the tutor and student moves by utterance. The annotation scheme is based on the literature[6][7]. 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
- 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 not 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
- **Conversation:** 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[8][9]. After several rounds of annotation, the inter-coder agreement on most of the categories reached an acceptable level (perfect

Table 1. Kappa Values and Percentages of Student and Tutor Moves by Tutor

Student Move	Kappa	Novice	Lecturer	Expert	Tutor Move	Novice	Lecturer	Expert
Explanation	**0.64**	**7.5**	26.3	19.8	Answering	**10.1**	5.4	1.4
Questioning	**0.89**	**18.3**	8.4	6.8	Evaluating	16.4	12.9	7.8
Reflecting	**0.65**	14.2	16.5	13.9	Summarizing	**6.9**	16.7	16.6
					General-Prompting	4.4	3.3	4.1
Answering	**0.80**	25	27.1	35.4				
Action-Response	**0.97**	12.5	10.4	9.7	Specific-Prompting	17.6	**27.7**	13.9
Completion	0.43	0	0.8	0.8	Diagnosing	2.5	3.3	3.3
					Declarative-Instructing	**22.6**	6.2	4.0
Conversation	**0.71**	9.4	16.9	10.5				
					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

agreement 0.8<Kappa≤1, or substantial agreement 0.6<Kappa≤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 detail Kappa values for tutor moves can be found in[3]. Table 1 reports the percentages of student and tutor moves by tutor. After analyzing both the tutor and student moves independently, we found that some behavior of our tutors supports the predictions on effective tutoring from the literatures[6][10]:

- the expert tutor and the lecturer summarize more than the novice;
- subjects with the expert tutor and the lecturer do more explanations than the subjects 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 [6]. 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. 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:

Table 2. A Transcript Fragment from the Expert's Tutoring

Line	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

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 2 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.)

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 2, 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 2, 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 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. The left part of Table 3 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:

Table 3. Interaction Patterns of the Expert Tutor

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

- **Explanation:** the novice tutor summarizes 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 2, after the student's explanation in line 45 the expert tutor evaluates it 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.

The right part of Table 3 summarizes the student-tutor interaction patterns in which the expert tutor is different from the non-expert tutors.

4 Study of Multi-Utterance Turns

While we were studying the interaction patterns, we observed that not all of tutor's specific prompting are immediately followed by any student move: 35.6% of the expert tutor's specific prompting is not immediately followed by any student move, which is much higher than that of the lecturer's (21.5%) and the novice's (25%). For example, in Table 2, the expert tutor does specific prompting in line 40 but this specific prompting is followed by another specific prompting, instead of a student turn. This may be because most of the time the expert tutor does specific prompting in multi-utterances. This phenomenon also appears for other tutor moves, like from line 46 to line 48: in this single turn, the expert tutor uses three utterances, two categories of move.

Multi-utterances usually mean that in a single turn the tutor or the student make a sequence of moves (more than one) successively without being interrupted. The number of utterances in a single turn is called the "length" of the multi-utterance turn. The utterances are segmented based on the CHILDES transcription manual[11], which the transcribers used. So the first question is: what is the difference between the expert tutor and the non-expert tutors in lengths and frequencies of tutor multi-utterance turns and student multi-utterance turns? To answer this question, we counted the lengths and frequencies of tutor multi-utterance turns and student multi-utterance turns in each tutoring transcripts (for both problem 2 and problem 9 in the curriculum, three tutors, there are a total of 36 transcripts). Then we ran ANOVA on the counts to see whether there are significant differences between each two of tutors and between the two problems. (One-way ANOVA — analysis of variance, is a statistical procedure for testing the null hypothesis that several univariate data sets have the same mean. When significant, ANOVAs are followed by Games-Howell tests to determine which condition is significantly different from the others.)

Figure 1(a) shows the average lengths of multi-utterance tutor and student turns per problem. There is a significant difference in the average length of multi-utterance student turns between problem 2 and problem 9 ($p < 0.03$). Problem 9 is much more complex than problem 2 so the students use more utterances in a single turn.

Figure 1(b) shows the average lengths of multi-utterance tutor and student turns per tutor. The average length of the expert tutor's multi-utterance turn is significantly greater than the non-expert tutors' ($p < 0.005$). This means that the expert tutor talks more in each turn. The length of the expert tutor's multi-utterance turn varies from 1 to 22, but the maximum length of the Lecturer's is 9 and only two turns of the novice tutor have a length greater than 7. We

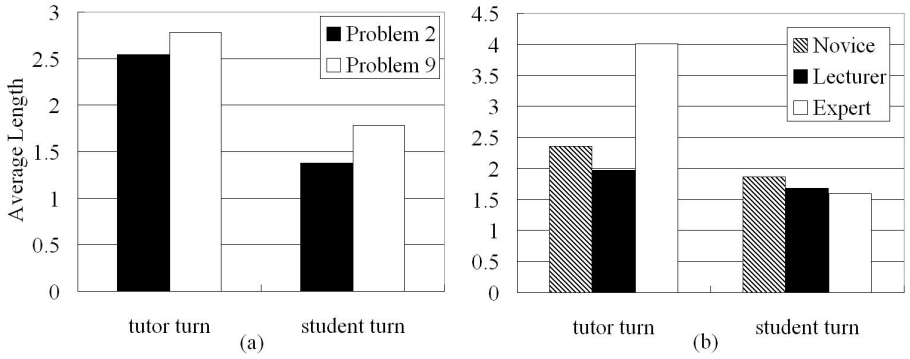

Fig. 1. Average Length of Multi-Utterance Tutor and Student Turns, per Problem(a) and per Tutor(b)

Table 4. Percentages of Each Category of Tutor Move Followed by Another Tutor Move, per Tutor

Tutor Moves	Novice(%)	Lecturer(%)	Expert(%)
Answering	5	7.212	0.743
Evaluating	13.75	**22.6**	9.653
Summarizing	**11.25**	30.77	22.77
General Prompting	5	3.365	4.455
Specific Prompting	8.75	14.9	7.673
Diagnosing	3.75	0.962	2.723
Declarative Instructing	**40**	10.58	5.198
Procedural Instructing	1.25	7.212	**23.02**
Demonstrating	11.25	**0**	15.59
Support	0	0.481	**6.188**

ran Chi-square on the length distributions of the three tutors' turns and there are significant differences between tutors in length 1, length 3 and length 4 ($p < 0.05$). The expert tutor's turns with only one utterance are significantly fewer than the non-expert tutors, but his 3-utterance and 4-utterance turns are significantly more than the novice tutor. It supports that the expert tutor tends to talk more in each single turn. The next question is how differently the expert tutor organizes his turn from the non-expert tutors. We analyzed the multi-utterance patterns of tutor turns with regards to how the tutors follow up differently each particular tutor move. First we looked at the differences between tutor as concerns which categories of tutor move are more likely followed by another tutor move. We ran Chi-square on the data in Table 4 (Numbers in boldface refer to significant differences). We found:

- the novice tutor has significantly fewer summarizing, but many more declarative instructing followed by another move than the expert tutor and the lecturer ($p < 0.003$ in both cases);

– the expert tutor has significantly more procedural instructing and support followed by another move than the non-expert tutors ($p < 0.004$ in both cases);
– the lecturer has much more evaluating but no demonstrating followed by another move than the novice and expert tutors ($p < 0.03$ in both cases);

Procedural instructing teaches the student how to solve a problem procedurally so it can seldom be completed by one single utterance. So we speculate that the expert tutor likes to use completed procedural instructing to help students. Before continuing the tutoring, the expert tutor also likes to encourage his student by support which would push students to move forward.

Like for interaction patterns, it is more meaningful to find out that after each category of tutor move, how the expert tutor differs in the following move from the non-expert tutors. We ran Chi-square on the frequencies of all the multi-utterance patterns of the tutors. Across all patterns, there are significant differences in the following moves to each category of tutor move between all the tutors ($p \approx 0$). More specifically, we found:

– **Answering:** the expert tutor does specific prompting much more than the non-expert tutors after answering — this shows our expert tutor does often prompt and scaffold students but normally after his answering to students' question;
– **Evaluating:** the expert tutor and the lecturer do specific prompting much more than the novice tutor after evaluating; the expert tutor does procedural instructing much more than the non-expert tutors;
– **Summarizing:** the expert tutor does summarizing in multiple utterances much more than the non-expert tutors;
– **General Prompting:** the expert tutor does much fewer specific prompting than the non-expert tutors after general prompting;
– **Specific Prompting:**
  • the expert tutor and the lecturer do procedural instructing much more than the novice tutor after specific prompting;
  • all the three tutors do specific prompting in multiple utterances;
– **Diagnosing:** the expert tutor does much more procedural instructing and support than the non-expert tutors after diagnosing;
– **Declarative instructing:** the expert tutor does much more procedural instructing and demonstrating, but much fewer specific prompting than the non-expert tutors, after declarative instructing;
– **Procedural Instructing:** the expert tutor does procedural instructing in multiple utterances much more than the non-expert tutors; he also does much more demonstrating, but much fewer specific prompting than the non-expert tutors, after procedural instructing;
– **Demonstrating:** the lecturer never does demonstrating but the novice and expert tutors do demonstrating in multiple utterances;
– **Support:** the expert tutor does almost any kind of tutor move after support.

Table 5. Patterns of Multi-Utterance Turns of the Expert Tutor

Tutor Move	Tutor Move
Answering	Specific Prompting
Evaluating	Procedural Instructing
Summarizing	Summarizing
Diagnosing	Procedural Instructing
Diagnosing	Support
Declarative Instructing	Procedural Instructing
Declarative Instructing	Demonstrating
Procedural Instructing	Procedural Instructing
Procedural Instructing	Demonstrating
Support	Summarizing
Support	Procedural Instructing
Support	Support

Comparing the novice tutor with the expert tutor and the lecturer, she does declarative instructing after almost any kind of tutor move much more than the other two tutors. This supports our finding that the novice tutor tends to give out the information or tell the solution directly. These findings above hint at why the expert tutor is much more effective than the non-expert tutors even though he prompts less, talks more and leaves less talking to students comparing to the lecturer: the expert tutor summarizes more completely, does procedural instructing and demonstrating more effectively and encourages students by support before moving on. Table 5 summarizes the patterns of multi-utterance turns in which the expert tutor is different from the non-expert tutors.

5 Conclusions and Future Work

Our analysis of tutorial dialogue moves, interaction patterns and multi-utterance turns 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. Before moving on, 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);
5. Uses questions to enhance problem solving (prompting).

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 a baseline ITS to solve the letter pattern problems and did some experiments on the baseline system with different kinds of simple feedback messages[3]. 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 dataset, 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. For this introductory computer science 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.

6 Acknowledgements

This work is supported by grant N00014-00-1-0640 from the Office of Naval Research.

References

1. Fox, B.: The human tutorial dialogue project. Lawrence Erlbaum Associates (1993)
2. Chae, H. M., Kim, J. H., and Glass, M.: Effective behaviors in a comparison between novice and expert algebra tutors. Proceedings of Sixteenth Midwest AI and Cognitive Science Conference (2005) 25–30
3. Di Eugenio, B., Kershaw, T. C., Lu, X., Corrigan-Halpern, A., and Ohlsson, S. : Toward a computational model of expert tutoring: a first report. 19th International conference of Florida Artificial Intelligence Research Society (2006) 503–508
4. Glass, M., Kim, J. H., Evens, M. W., Michael, J. A., and Rovick, A. A.: Novice vs. expert tutors: a comparison of style. Proceedings of Tenth Midwest Artificial Intelligence and Cognitive Science Conference (1999)
5. Kotovsky, K. and Simon, H.: Empirical tests of a theory of human acquisition of information-processing analysis. British Journal of Psychology, **61** (1973) 243–257
6. Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., and Hausmann, R. G.: Learning from human tutoring. Cognitive Science, **25** (2001) 471–533
7. Litman, D. J., Rose, C. P., Forbes-Riley, K., Vanlehn, K., Bhembe, D., and Silliman., S.: Spoken versus typed human and computer dialogue tutoring. Proceedings of 7th International Conference on Intelligent Tutoring Systems (2004)
8. Carletta, J.: Assessing agreement on classification tasks: The kappa statistic. Computational linguistics, **22** (1996) 249–254
9. Di Eugenio, B. and Glass, M.: The kappa statistic: a second look. Computational linguistics, **30** (2004) 95–101
10. Landsberger, J.: Feedback to improve Study Guides (2005)
11. MacWhinney, B.: The CHILDES project.Tools for analyzing talk: Transcription Format and Programs, volume 1, 3rd edition. Lawrence Erlbaum (2000)