How Adaptive Is an Expert Human Tutor?

Michelene T.H. Chi[1] and Marguerite Roy[2]

[1] Arizona State University
[2] Medical Council of Canada
Michelene.Chi@asu.edu

Abstract. In examine the tutoring protocols of one expert human tutor tutoring 10 students in solving physics problems, four analyses reveal that he tutored the five good learners in different ways than the five poorer learners, resulting also in greater adjusted gains for the good learners. This opens up the question of whether the tutor is non-optimally adaptive. We introduce a new conceptual framework and a new perspective in our coding analyses in order to examine how adaptive an expert tutor is.

1 Introduction

In this paper, we address a specific question, by analyzing protocol data that were collected in a 2008 study of an expert human Tutor tutoring 10 Tutees in solving physics problems [1]. Studying a single tutor tutoring 10 students allowed us to examine tutor variability as a function of the tutees. The study and its major learning results from the perspective of the bystander observers were published in [1]. Here, we report other analyses of the raw protocols to address specifically the question of whether an expert tutor is adaptive.

The common assumption among tutoring researchers is that tutoring is beneficial to all students in part because tutors are adaptive to the needs of the tutees. Adaptiveness can be defined in many ways, but the general idea is that a human tutor is adaptive in the sense that she tailors her instruction to the needs of her tutee. Tailoring can be defined in a macro way as selecting the appropriate next problem for a student to solve [2], such as a more difficult problem if a tutee successfully solved the current problem. Using this criterion, we had also examined the choice of problems our Tutor had posed to the 10 Tutees. Although the number of problems from which our Tutor could have selected were few (4 problems), they nevertheless did vary in difficulty. Overall, as we showed in our prior study [1, Pp. 334-335], there were no significant differences in whether the Tutor selected and posed the more difficult problems to the better tutees, suggesting that the Tutor was not adaptive in the macro level sense. Furthermore, the overall finding that the bystander observers could learn as well as the Tutees [1] even though the Tutor could not have tailored their instruction toward the observers, made us wonder whether tutors are in fact as adaptive as commonly believed.

From a micro perspective, tutoring adaptiveness is usually identified as choosing the appropriate next solution step for the student to work on, whether to give a

V. Aleven, J. Kay, and J. Mostow (Eds.): ITS 2010, Part I, LNCS 6094, pp. 401–412, 2010.

proactive hint or scaffold before the step [3], or deciding on the specificity of the hint to pose [4], contingent upon the success of the student in solving the prior step. Intelligent tutoring systems generally take this finer-grained approach in defining adaptiveness. In short, whether at a more macro problem level or at a more micro solution-step level, both approaches to defining adaptiveness view it as a choice of what materials to present by the *tutor* to the tutees.

In this work, we define adaptiveness from the perspective of the *tutee*. Instead of looking at what content the tutor chose to present to a tutee, we examined instead what kind of pedagogical move the tutor made to elicit productive learning activities from the tutees. Accordingly, we also attempted to modify our coding from the perspective of the tutor to the perspective of the tutee.

Our perspective is derived from a framework we outlined in [5] to define "active learning." To improve learning, it has been widely proposed in many areas of literature that students engage in "active learning" as opposed to "passive learning." For example, in educational psychology "active learning" has been broadly defined as encouraging learners to pay "attention to relevant information, organizing it into coherent mental representations, and integrating representations with other knowledge" [6]. In engineering education, "active learning" has been defined as "engaging students in the learning process [through] activities that are introduced into the classroom" [7].

We have differentiated "active learning" into three different kinds of student activities—*active, constructive, interactive*--that can be observed overtly, and defined the cognitive processes corresponding to each kind of activities. For example, *active* activities might include copying a solution from a whiteboard, underlying a sentence in a text, or clip-and-pasting a sentence. *Constructive* activities mean producing some new knowledge that was not presented in the instructional materials, such as drawing a diagram or a concept map, comparing-and-contrasting two examples or self-explaining a worked-out example. In these cases, a student is producing something beyond what was contained in the instructional materials: such as a diagram, similarities and differences, or self-explanation inferences. Finally, *interactive* activities involve directly interacting with a peer or a tutor, such that both partners can further elaborate, elucidate, scaffold, provide feedback, and so on, to each other. Our framework explains why being "active" promotes more learning than being "passive," which was operationalized as not doing anything overtly. Moreover, we had hypothesized that participating in *interactive* activities is often (but not always) better for learning than participating in *constructive* activities, which in turn is better than participating in *active* activities, which in turn is better than being *passive* [5]. We are not saying that one must engage in a specific kind of overt activities in order to learn. Rather, we are simply proposing that in general, students are more likely to learn more by engaging in one kind of learning activities over another kind, and the ordering ranks as follows: *interactive>constructive>active>passive*.

Since tutorial dialogues involve a tutor conversing with a tutee and expects a response from a tutee, should we consider all tutorial dialogues as naturally interactive? We propose that if we examine tutorial dialogues from the perspective of a tutee's contributions, then we can clearly differentiate a tutee's contribution as either *passive, active, constructive,* or *interactive*, so that not all tutee responses should automatically be considered to be interactive. In particular, if a tutee responds to a tutor's comment with a continuer, such as "ok," or "uh-huh," then we can consider a continuer type of responses as an *active* response only. However, if the tutee responds to a tutor's comment with a content-relevant follow-up, then we can consider it a *constructive* response. We further apply the criterion that a tutee's response is *interactive* if it initiates some new topic, new direction, and so forth.

Using this operational definition of tutee's responses as a way to differentiate *active, constructive,* and *interactive,* our assumption is that some tutor moves are more likely to promote one kind of responses than another. In Fig. 1 below, we re-ploted the three largest categories of tutor moves—explaining, giving feedback, and scaffolding, averaged across 11 tutors, taken from data reported in our 2001 study [8, Fig. 3], in terms of proportion rather than frequency. Fig. 1 shows that the tutors' explanations elicited the largest proportion of continuer type of *active* responses, and a smaller proportion of shallow *constructive* type of follow-up responses. In contrast, the tutors' scaffolding moves elicited the smallest proportion continuer type of responses and the largest proportion of shallow follow-ups. Feedback moves also elicited proportionately more shallow follow-ups than continuers, but the difference was not as pronounced as for scaffolding moves. Overall, all tutor moves elicited comparable and minimal deep follow-ups. Comparing explaining and scaffolding moves only in this paper, this suggests that scaffolding was a better tutor move than explaining, because scaffolding moves often elicited some *constructive* responses from the tutees whereas explaining moves were more likely to elicit *active* responses.

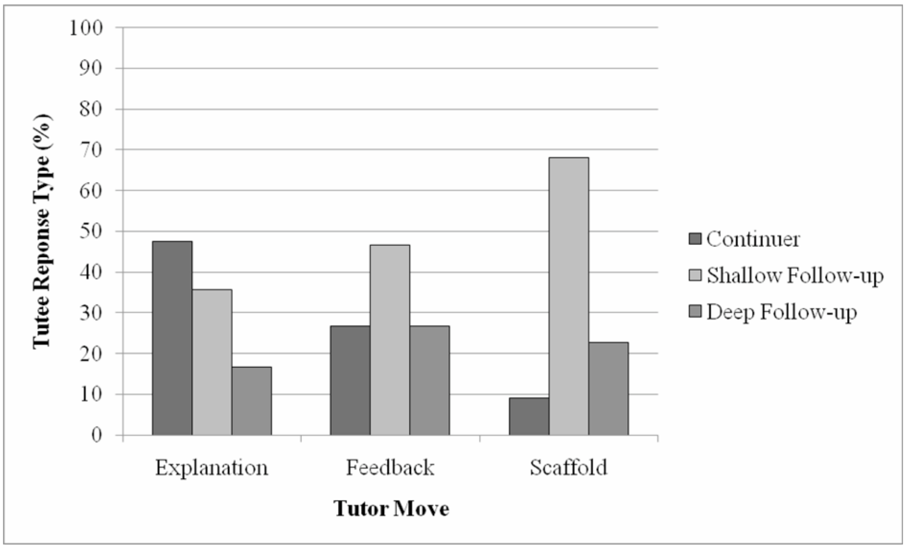

Fig. 1. The proportion of tutees' continuer, shallow, and deep follow-up responses to tutors' explanations, feedback and scaffolding moves (taken from Chi, et al. 2001 data)

2 The Data of the Current Analyses

The data analyzed for this paper consisted of an experienced teacher who had taught college physics for over 30 years. Moreover, he was in a lab that developed intelligent tutoring systems and knew that a good tutor ought to scaffold tutees, thus we consider him to be an expert tutor. In the 2008 study [1], this Tutor was asked to tutor 10 different Tutees. A pre-test was administered to the 10 Tutees after they had learned the materials in Chapter 5 of a physics text on their own, without feedback. Essentially the pre-test measured how well the students could learn the content of Chapter 5 on their own, after having learned the first four chapters to mastery. Thus, the pre-test was not a test of their prior knowledge about physics, but more of an assessment of whether a tutee is a good or a poor learner, when they had to learn unguided. After the pre-test, then these 10 students were individually tutored by the Tutor on solving four problems pertaining to Chapter 5 content.

In order to make our codings more manageable and more meaningful, we focus here on dialogue protocols segmented into episodes about "critical" nodes. Critical nodes were those nodes in a problem space that we thought were more important in terms of requiring the solver to generate a specific equation, solution step, or about main concepts and principles. The problem space of all possible nodes were identified initially by transcribing how our expert Tutor solved each of the problems that he was to tutor. All utterances pertaining to a critical node were counted in the node's episode. Thus, the tutorial dialogues of all participants have approximately the same number of episodes, because the Tutor usually made sure that all critical nodes were covered. All the analyses to be reported below used "episode" as the unit of analyses.

3 Results

Before describing four analyses to give a view of how adaptive our expert Tutor was, we first assert that although tutoring is often considered to be the best instructional technique in helping students learn, nevertheless, tutoring is not equally effective with all students. We can see tutoring's differential effectiveness easily in multiple ways. For example, the 10 Tutees in our sample varied in their pre-test scores significantly, ranging from a low score of around 30% to a high pre-test score of around 60%. And, as is the case with many other kinds of interventions, there was a significant correlation between pre-test scores and post-test scores. If we use the data of all 69 participants across all five treatment groups reported in [1], the correlation between pre-test and post-test scores were significant at the $p<.0005$ level ($r=0.648$). In focusing here on the tutoring condition only, Figure 2 below divides the Tutees into three groups (to show the incremental variability): the 3 Low Tutees obtained a pre-test scores between 30-40%, the 3 Medium Tutees obtained scores between 40-50%, and the 4 High Tutees obtained scores between 60-70% (all pre-test scores are shown in the dark bars). The results show that Tutees learned different amounts, depending on whether they had more or less knowledge coming into the tutoring situation after having learned the materials in Chapter 5 on their own.

The point of Figure 2 is simply to show that the Tutor was not equally effective with all Tutees; in fact, he was least effective with the poorest Tutees. This suggests that there is room for improvement in terms of what tutors can do to guide a poor tutee's learning. For example, was the Tutor adaptive in making the Tutees more *constructive* rather than merely *active?* The next four analyses examine the Tutor's adaptiveness from a tutee's perspective.

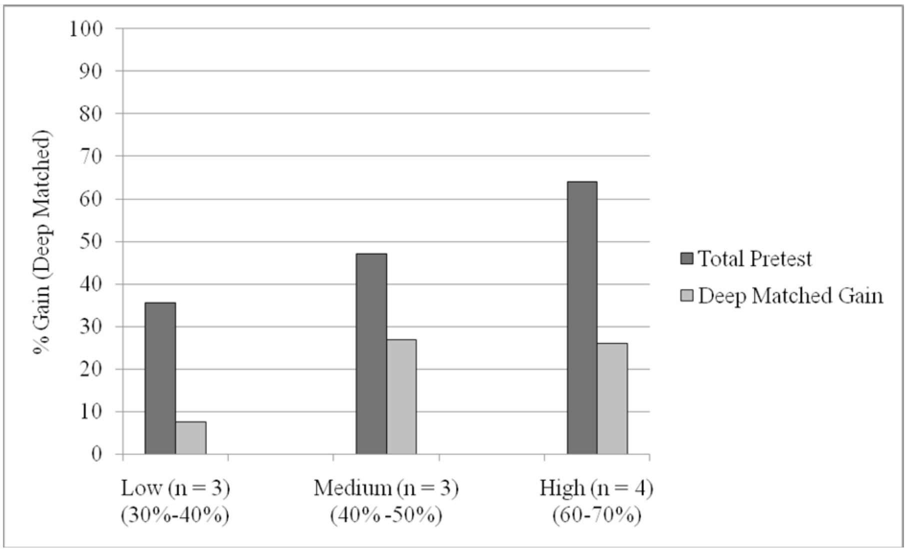

Fig. 2. Dark bars show the pre-test scores of the Tutees, divided into Low, Medium, and High, and the lighter-colored bars show how much they improved after tutoring on the same matched pre- and post-tests, scored for correct deep solution steps only

3.1 Does an Expert Tutor Jointly Explain or Take over the Coverage of a Critical Node?

Given that we only have 10 Tutees, our analyses henceforth will divide the Tutees into two groups: Good versus Poor Tutees. Good Tutees were defined as the five Tutees who gained more (on average 25% from pre- to post-test) and made fewer errors (on average 56 error steps across 4 problems); and Poor Tutees were defined as the five Tutees who gained less (16%) and made many more errors (89 error steps). Further details about the Good versus Poor Tutees split are described in [1].

Each episode comprised of either a single turn by either the Tutor or the Tutee, or multi-turns by both. When it consists of multi-turns, and if both the Tutor and the Tutee made substantive contributions, then the node is considered "jointly-covered." However, if only one person (either the Tutor or the Tutee) made substantive contributions in covering (i.e., explaining or solving) a critical node, then we consider that node to be independently covered. Thus, while tutoring, a critical node can be covered, whether successfully or unsuccessfully, either by a tutee alone, by a tutor alone, or by both of them jointly, when only substantive contributions are considered. Based on our framework described above, when the Tutor covered a node alone, then the Tutee seems *passive.* But when the Tutee covered a node alone, then the Tutee was *constructive.* However, when the Tutor-and-Tutee jointly covered a node, then the Tutee was likely *interactive.* As our *interactive>constructive>active>passive* hypothesis suggested above, *interactions* should facilitate learning more so than being *constructive,* which is better than being *passive.*

On average, the majority of the critical nodes (55) were covered by Tutor-and-Tutee jointly, whereas 32 were covered independently by the Tutor and 16 independently by the Tutees. Figure 3 shows a breakdown of how the nodes were covered for Good and Poor Tutees, and an interesting pattern of differences emerge. It is not surprising to find that Good Tutees were more able to explain/solve a node independently than Poor Tutees ($F(1,8)=50.oo$, $p<.0005$, $d=4.21$). However, the Tutor covered the critical nodes independently more often for the Poor Tutees than the Good Tutees ($F(1,8)=98.04$, $p<.0005$, $d=5.219$), whereas they covered the nodes jointly more often with the Good versus the Poor Tutees ($F(1,8)=21.00$, $p=.002$, $d=2.892$).

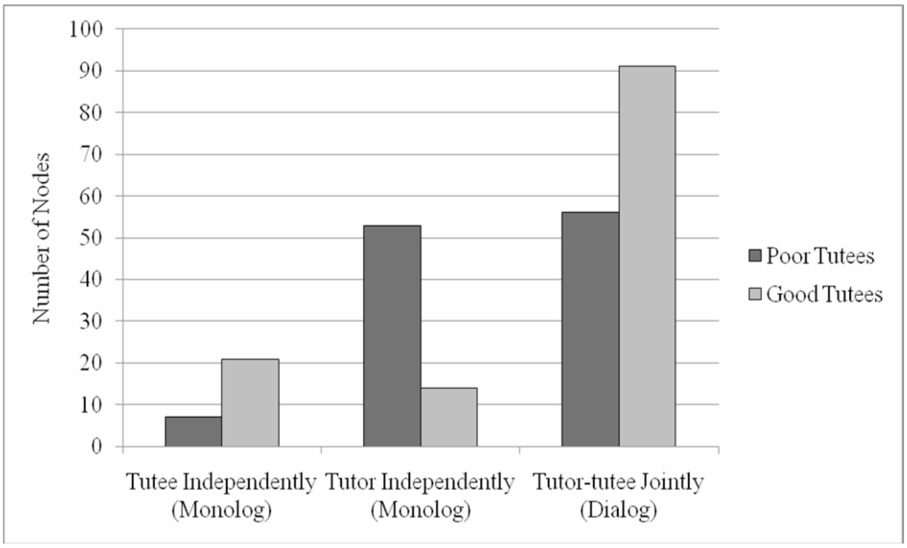

Fig. 3. The number of critical nodes covered by the Tutees alone, by the Tutor alone, or jointly by both, for Poor (dark bars) and Good (light bars) Tutees

In our framework, to maximize learning, a tutor should instead cover the nodes less frequently alone, since independent coverage by a Tutor essentially means that the Tutor explains didactically to the Tutees, allowing the Tutee to be *passive.* Moreover, the Tutor should encourage more joint coverage with Poorer Tutees than the Good Tutees, since joint coverage would encourage the tutees to be *interactive*. In short, the

Tutor was not optimizing his adaptiveness, from the perspective of differentially eliciting *active, constructive,* or *interactive* responses from the Good and the Poor Tutees.

To verify whether being *interactive* (during joint coverage) and being *constructive* (when Tutees cover a node independently) facilitated learning, there ought to be significant correlations between the frequency of joint coverage with Tutees learning (r=0.457, p=.043), and Tutees' independent coverage with Tutees' learning (r=0.640, p=.046), but not when the Tutor covered a node alone (n.s.), since the Tutees were most likely *passive.* In fact, the correlation results do support our predictions.

3.2 Could Poor Tutees Be Helped with More Tutor Scaffolding?

One could dismiss the results shown in Fig. 3 by pointing out that of course the Tutor covered more nodes independently with Poor Tutees and engaged in more joint dialogues with the Good Tutees, because the Good Tutees were more capable of independent coverage and engaging in joint dialogues with the Tutor. Our point is that to be truly adaptive, a tutor could in principle be more responsive to poorer tutees' inability to respond in joint dialogues, initiate new comments, and cover nodes by themselves. But the question is, would it help the Poor Tutees if they did receive more scaffolding?

Our argument is indirectly supported by the results from our 2001 Study 2 [8]. In Study 2 [8], tutoring in a conceptual domain (the human circulatory system), the 11 tutors were suppressed from giving explanations. In fact, they were permitted only to scaffold the tutees in a restricted content-free way, with scaffolding prompts such as "What does this mean?" The tutees in Study 2 learned just as much from the 11 tutors when they were scaffolded, as the tutees in Study 1 when the same tutors tutored more naturally. Although this is indirect evidence, it does suggest that all tutees (good and poorer ones) could learn when tutors were only permitted to scaffold them.

To address the same question here, we analyzed the proportion of tutee responses that were merely *active*, such as a continuer, versus more *constructive*, such as a shallow follow-up. Fig. 4a below shows that for the five Good Tutees, they generated proportionately more *active* continuer type of response to Tutor's explanations than to Tutor's scaffolding, whereas they generated more *constructive* follow-up responses to Tutor's scaffolding than Tutor's explanations. As before in the data collected in the 2001 study [8] and shown in Fig. 1, they did not generate many deep follow-ups.

The very same pattern of tutee responses hold for the Poor Tutees as well, as shown in Fig. 4b. They responded to explanations with more continuers than to scaffolding, whereas they responded more with shallow follow-ups to scaffolding than to explanations. This pattern of results again suggests that a tutor move such as scaffolding is advantageous to both Good and Poor Tutees. Nevertheless, the Tutor is much more likely to explain to a Poor Tutee than to scaffold a Poor Tutee, and conversely, the Tutor is more likely to scaffold a Good Tutee than explain to a Good Tutee.

Even though scaffolding is beneficial to both Good and Poor Tutees, as shown in Fig. 4a and 4b, the Tutor gave significantly more explanation statements to the Poor Tutees than the Good Tutees (F(3,6)=8.281, p=.015), but in contrast, gave predominantly more scaffolding statements to the Good Tutees than the Poor Tutees F(3,6)=7.333, p=.020). In this sense, from the tutee's perspective, the Tutor was not very adaptive.

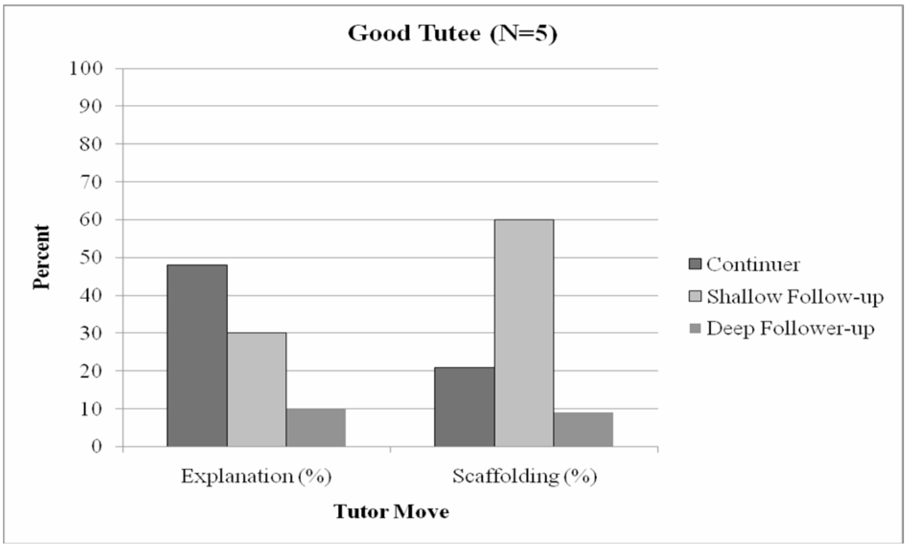

Fig. 4a. Good Tutee's responses to different types of tutor moves

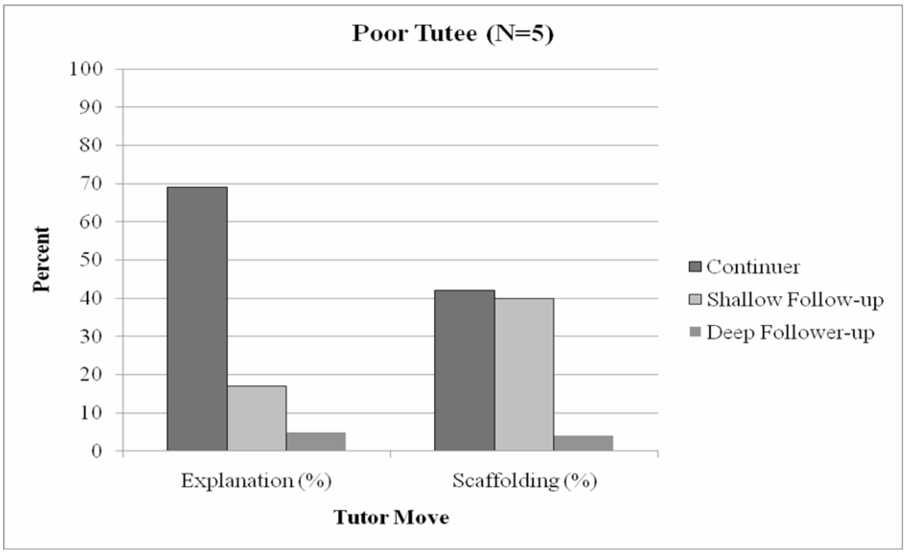

Fig. 4b. Poor Tutee's responses to different types of tutor moves

3.3 Examples of Tutor Explaining Moves versus Scaffolding Moves

In this analysis, we want to give specific contrasting examples of how explaining versus scaffolding can be coded. One way to code explaining is to determine whether the Tutor gave statements that were more telling and directing versus those that made open-ended requests. Open-ended requests are more scaffolding-like moves in that they are more likely to elicit *constructive* responses from the tutees. Figure 5 below shows that the Tutor gave didactic telling and directing statements more often to Poor Tutees than the Good Tutees, and conversely, requested fewer open-ended goals and explanations from the Poor Tutees than the Good Tutees. We recognize that the Tutor asked more open-ended requests from the Good Tutees in part because they are more able to answer such requests. So the Tutor is adapting, but not in a way that is optimal for all the Tutees' learning. That is, the Tutor should be doing the reverse, asking the Poor Tutees more open-ended questions so the Poor Tutees can be more *constructive.*

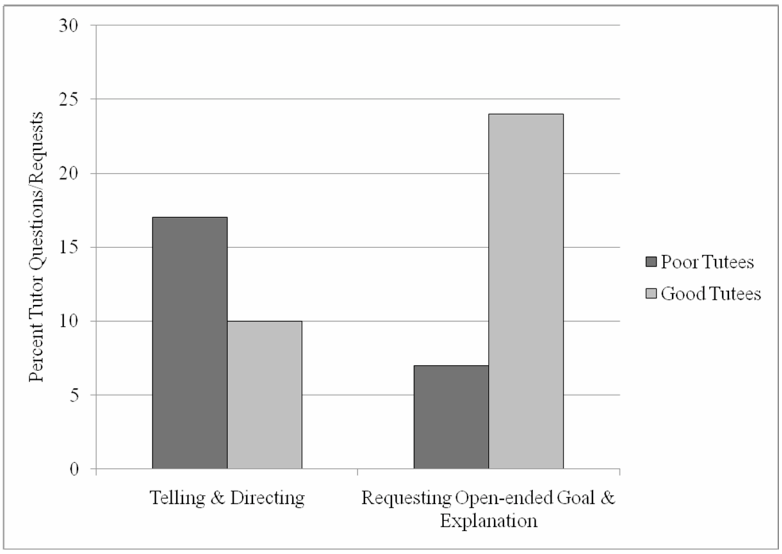

Fig. 5. Telling and directing statements versus open-ended requests made by the Tutor to Poor (dark bars) and Good (light bars) Tutees

3.4 Coding from the Tutees' Perspective: Was the Tutor Adaptive in Optimizing Tutees' Interactive Responses During Joint Dialogues?

The prior analyses, as many others have done in the literature, typically coded from the perspective of the Tutor, in terms of starting out with what the Tutor said. This is understandable given that tutors typically lead the dialogues. However, coding from a tutor's perspective cannot give us a true sense of a tutee's *interactivity,* since a tutee is obliged to give a response. Moreover, *interactivity* in the prior coverage analysis as shown in Fig. 3, was operationalized merely as both the Tutor and the Tutees having made substantive contributions. In order to operationalize *interactiveness* independent of Grice's conversational obligation and using a strict criterion, in this analysis, we coded from the Tutees' perspective, in that we coded the joint dialogues starting with what a Tutee initiated. Initiate means that the tutees commented on a new idea, a new topic, or a question that were not a follow-up, and so forth. We operationalized such tutee initiating moves as being *interactive.* Using this strict criterion and perspective, we coded four types of *interactive* dialogues, using the Tutee as the starting point.

a) Tutee initiate, Tutor revoice
b) Tutee initiate, Tutor scaffold
c) Tutee initiate with a question, Tutor answers
d) Tutee makes a meta-comment, Tutor responds.

The only tutor move that is novel in the above analysis is tutor "revoicing." In the literature, "revoicing" is defined as *a paticular kind of re-utterance of a teacher's contribution by a student* or as *a teachers' redecoration of a students response* [9]. Here, we treat revoicing as a repetition of parts of a tutee's utterances, but not necessarily verbatim. And such revoicing moves are typically undertaken by the Tutor when a Tutee makes a correct move. In essence, revoicing is a positive feedback move that is not discussed in the tutoring literature. Here are two examples:

Tutee: First the gravity is pulling down [This is a tutee-initiated statement.]
Tutor: Pulling it down. [Tutor revoiced.]

Tutee: Weight is..the mass times..acceleration due to gravity and that's force.
Tutor: Right. Right. [Tutor giving correct feedback.]
Tutee: Ok.
Tutor: So weight is the force. [Tutor revoiced.]

In contrast, feedback moves are typically given to errors. In the tutoring literature, feedback is identified either as a negative response ("that was incorrect" or "no"), a corrective positive response, that is giving the correct answer or equation (such as "it should be F=ma"), or it could be an elaborated corrective response (such as giving a reason for the answer) [1]. In general, about 80%-90% of both Good and Poor Tutees' errors are responded to with either a negative feedback, a corrective feedback, or an elaborated feedback. Elaborative feedback obviously is the best kind since it gives more information and justification with respect to the feedback. Here is an example of an error, followed by both a corrective and elaborative feedback:

Tutee: FN would be…would FN be mass of A plus mass of B? Or?
Tutor: Again you…a force cannot be mass. [Corrective feedback].
These are two distinct quantities. [Elaborative feedback].

In our prior analysis [1, Table 5] we found the Poor Tutees received proportionately more corrective than elaborative feedback, whereas the Good Tutees received proportionately more elaborative feedback than corrective feedback. We had interpreted this difference to suggest that the Tutor was not as adaptive as one would like, since Poorer Tutees could have benefitted more from elaborative feedback.

In the current coding from the tutee's perspective, we found a significant overall difference between the percentage of critical nodes episodes that contained Tutee initiated statements for the Good Tutees (27.5%) than the Poor Tutees (6.43%, $F(1,7)=29.851$, $p=.011$). Moreover, there was a significant overall difference between Good and Poor Tutees for the proportion of their initiatives that were revoiced (18% for Good versus 5% for Poor, $F(1,7)=11.87$, $p=.011$). Although it is not surprising that Good Tutees can initiate more (since they know more), what is surprising is that the Tutor revoiced Good Tutees' initiatives more often than the Poor Tutees' initiatives. We surmise that the Good Tutees' initiations were more correct than the Poor Tutees' initiations, therefore the Good Tutees' initiations were more likely to be revoiced. Nevertheless, this is an important feedback move that is subtle, overlooked by the tutoring literature, and could potentially be overwhelmingly beneficial to the Good (as well as the Poor) Tutees. See Fig. 6, first pair of columns.

In addition to revoicing, Good Tutees' initiations were also followed significantly more often by scaffolding moves, consistent with the results reported above for 11 tutors of the prior study [1]. There were no differences between the Good and Poor Tutees in the frequency of the Tutor's responses to Tutees' questions or meta-comments they initiated. In short, dialogues with Good Tutees were more *interactive* largely because the Good Tutees initiated more frequently, and their initiations were taken-up by the Tutor, whereas Good and Poor Tutees' questions and meta-comments were responded to equally appropriately by the Tutor. The question is how should a tutor encourage more initiations from a poor tutee.

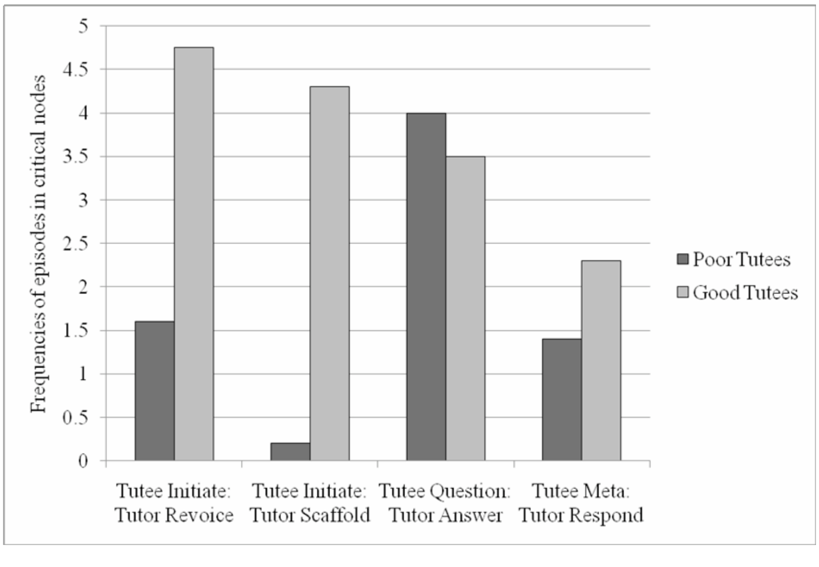

Fig. 6. The Tutor's responses to Good and Poor Tutees' initiatives

4 Discussion and Future Work

An implicit assumption among tutoring researchers and ITS developers is that tutoring is effective because in part a tutor is adaptive. Adaptiveness has been defined from the perspective of the tutor, in terms of a tutor's appropriate selection of a problem or a hint that is tailored to the tutee. We examined adaptiveness instead in terms of the kind of moves a tutor makes from the perspective of whether a move elicited *passive, active, constructive* or *interactive* responses from the tutees. We consider a tutoring move to be a "good" kind of moves if they elicit more *active, constructive* or *interactive* responses. We found that our expert Tutor tended to provide a greater number of good tutoring moves (such as scaffolding, asking open-ended questions, revoicing) to the Good Tutees than the Poor Tutees, and conversely, provided a greater number of less-effective tutoring moves (such as explaining, telling and directing) to the Poor Tutees than the Good Tutees. This pattern of tutor moves suggests that the Tutor was not optimizing the poorer tutees' learning, therefore, the Tutor was basically maladaptive. Granted that poorer tutees were incapable of offering more initiatives and responses to scaffoldings, our position is that our evidence suggests that they could if tutors gave them more guidance in doing so. We surmise that because tutors have a bias in wanting to get the correct knowledge or solutions out there, they have the inclination of telling and directing the tutees when they struggle, rather than help them get through their struggling. The results reported here suggest that future analyses may benefit from taking the perspective of the tutees, in order to understand their contributions in enhancing learning from tutoring.

Acknowledgments. The authors are grateful for funding provided by the National Science Foundation grant number 0325054.

References

1. Chi, M.T.H., Roy, M., Hausmann, R.G.: Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. Cognitive Science 32(2), 301–341 (2008)
2. Shute, V.J.: A macroadaptive approach to tutoring. Journal of Artificial Intelligence in Education 4(1), 61–93 (1993)
3. Murray, R.C., VanLehn, K., Mostow, J.: Looking ahead to select tutorial actions: A decision-theoretic approach. International Journal of Artificial Intelligence in Education 14(3/4), 235–278 (2004)
4. Wood, D.J., Wood, H., Middleton, D.: An experimental evaluation of four face-to-face teaching strategies. International Journal of Behavioral Development 1, 131–147 (1978)
5. Chi, M.T.H.: Active-constructive-interactive: A conceptual framework for differentiating learning activities. Topics in Cognitive Science 1, 73–105 (2009)
6. Mayer, R.: Learning and Instruction, p. 17. Pearson Education, Inc., New Jersey (2008)
7. Prince, M.: Does active learning work? A review of the research. Journal of Engineering Education 93(3), 223–231 (2004)
8. Chi, M.T.H., Siler, S., Jeong, Y.H., Yamauchi, T., Hausmann, R.: Learning from human tutoring. Cognitive Science 25, 363–387 (2001)
9. O'Connor, M., Michaels, S.: Shifting participant framework: orchestration thinking practices in group discussion. In: Hicks, D. (ed.) Discourse, learning, and schooliing, pp. 63–103. Cambridge University Press, Cambridge (1996)