





“Can You Clarify What You Said?”: Studying the Impact of Tutee Agents’ Follow-Up Questions on Tutors’ Learning

Tasmia Shahriar^(✉)  and Noboru Matsuda^(✉) 

North Carolina State University, Raleigh, NC 27695, USA
{tshahri, noboru.matsuda}@ncsu.edu

Abstract. Students learn by teaching others as tutors. Advancement in the theory of learning by teaching has given rise to many pedagogical agents. In this paper, we exploit a known cognitive theory that states if a tutee asks deep questions in a peer tutoring environment, a tutor benefits from it. Little is known about a computational model of such deep questions. This paper aims to formalize the deep tutee questions and proposes a generalized model of inquiry-based dialogue, called the constructive tutee inquiry, to ask follow-up questions to have tutors reflect their current knowledge (aka knowledge-building activity). We conducted a Wizard of Oz study to evaluate the proposed constructive tutee inquiry. The results showed that the constructive tutee inquiry was particularly effective for the low prior knowledge students to learn conceptual knowledge.

Keywords: Learning by teaching · Deep questions · Teachable agents · Tutor learning · Knowledge-building · Wizard of Oz

1 Introduction

Students often learn by teaching others. This type of learning is called *tutor learning* [1–10]. Although learning by teaching is impactful for the tutors, some researchers argue about the consistency and effect size of the tutor learning [3]. Roscoe [11] sought to understand how the tutors learn in the learning by teaching environment to leverage the benefits of tutoring more consistently. In one study, Roscoe *et al.* [12] found that tutor learning happens when students (who are *tutors*) engage in instructional activities like reflecting on their understanding [13], revisiting the concepts, providing correct and complete explanations to make sense of solution steps [14, 15], and recovering from their misconceptions or knowledge gaps [15] while tutoring. These instructional activities are called knowledge-building activities. When students perform knowledge-building activities to answer a question, those answers are known as knowledge-building responses that elicit learning [12]. One of our motivations for the current study is to evaluate the impact of knowledge-building responses on tutor learning in an Intelligent Tutoring System (ITS) setting.

Students infrequently engage in knowledge-building activities when they act as tutors in a peer-tutoring environment. Instead of developing their knowledge, students seem

more inclined to deliver what they already know or directly dictate solution steps with little elaboration to their tutees [11, 16, 17]. Researchers found that *deep* tutee questions made tutors generate more knowledge-building responses while tutoring [16–23]. Although different types of *deep* tutee questions have been observed, they all agreed that not all tutee questions are equally helpful for tutor learning. Despite the advancement in developing effective teachable agents [24–29], a computational model of *deep* tutee questions beneficial for tutor learning is yet to be developed.

In the current study, we develop a synthetic tutee that asks questions to promote tutors' knowledge-building responses. We hypothesize that letting the synthetic tutee ask follow-up questions to remind the tutor to provide knowledge-building responses facilitates tutor learning. For example, the synthetic tutee may ask tutors to elaborate their shallow response. We shall call this kind of question asking the *Constructive Tutee Inquiry* (CTI). We propose a model of CTI for a tutee to ask follow-up questions. This is the first attempt to model a synthetic tutee's follow-up questions to facilitate tutor learning to the best of our knowledge. As an initial step, we evaluate the effectiveness of CTI through a Wizard of Oz study (WoZ). In the study, a synthetic tutee embedded into an artificial peer learning environment called APLUS was controlled by a researcher who manually typed the tutee questions as if a machine generated them. The study participants were told that the synthetic tutee was artificial intelligence.

In the traditional APLUS (which does not have CTI), students interactively teach a synthetic tutee named SimStudent [30] to solve linear algebraic equations. In WoZ, an extended version of the traditional APLUS, called AskMore APLUS, was used. In AskMore APLUS, a synthetic tutee named Gabby instantiated follow-up questions according to the CTI. The result showed that the proposed CTI model helped the low prior students generate more knowledge-building responses, facilitating their learning of conceptual knowledge. Our contributions are summarized as follows: (1) We propose a domain-independent Constructive Tutee Inquiry model that encourages tutors to provide more knowledge-building responses. (2) We present rigorous analysis to understand how and why Constructive Tutee Inquiry facilitates tutor learning.

2 Related Work

Many researchers have tested the effectiveness of varying tutee prompts in facilitating tutor learning. Prior studies revealed a controversial impact of *explanation prompt* on tutor learning. While some studies found asking for explanation effective [31, 32], some studies showed that asking a tutor to explain at all times has a detrimental effect on tutor learning [23, 33, 34]. In another study, Rittle-Johnson *et al.* [22] argued that despite the general effectiveness of explanation prompt for improved learning, explanatory questions divide tutors' attention to different types of information, negatively impacting tutor learning. These studies highlight that *deep* tutee questions are more than just asking for explanations from the tutor.

More recently, Baker *et al.* [35] found a positive impact of asking *contrasting prompts* to draw tutors' attention towards identifying similarities or dissimilarities between contradictory scenarios on tutor learning. On the contrary, Sidney *et al.* [36] claimed that contrasting questions alone on their own was not beneficial for tutor learning; instead, it

was both the *explanation* and *contrasting prompts* that together facilitated tutor learning [36]. This analysis indicates that any question type alone is insufficient for tutor learning.

Looi *et al.* [37] investigated the design of the synthetic tutee’s question prompts to engage the tutors in knowledge-building activities on an online learning environment named Betty’s Brain. These questions were one-sided questions with no follow-up regime in case of failure to engage tutors in knowledge-building activities. We hypothesize that *deep* questions are not just a single one-sided question; rather, it is an accumulation of subsequent follow-up questions on a certain topic. On the other hand, their proposed method requires the synthetic tutee to be equipped with a complete expert model, making it a domain-dependent question generation model. Our proposed model is domain-independent because no path of the CTI algorithm requires specific knowledge about the domain to generate the subsequent follow-up questions. Our model also does not assume that the synthetic tutee needs to be more knowledgeable than the tutor to engage them in knowledge-building activities. Therefore, our model arguably captures the naturalistic scenarios for tutee inquiries in a classroom setting.

3 The Traditional APLUS with SimStudent

Our study extends the SimStudent project [38], which we call traditional APLUS. Traditional APLUS is an intelligent learning environment where students (who are *tutors*) teach SimStudent how to solve linear algebraic equations. SimStudent learns through demonstration. When the tutor teaches a step, SimStudent learns generalized production rules that look like, “if [*preconditions*], perform [*transformation*]”. In APLUS, *transformation* allows four basic math operations: *add*, *subtract*, *divide*, and *multiply* by a *number*. SimStudent keeps adding or modifying the production rules in its knowledge base according to the tutor’s feedback. Besides demonstrating solution steps, tutors can interact with SimStudent and give textual explanations using the chat panel. In traditional APLUS, SimStudent only asks why questions in some particular scenarios. For example, “*Why did you perform [transformation] here?*” or “*I performed this [transformation] due to a previously taught production rule. Why do you think I am wrong?*”. SimStudent never follow-up after the tutor’s response to the why questions. Additionally, tutors can quiz SimStudent anytime they want to check its knowledge status. Quiz topics include equations with one-step, two-steps, variables on both sides, and a final challenge containing variables on both sides. Traditional APLUS also has resource tabs like problem bank, unit overview, and worked-out examples for tutors to review at any time.

4 Overview of the Constructive Tutee Inquiry

4.1 Motivation

Prior studies have highlighted that tutors learn most effectively when they engage in *knowledge-building activities* [12, 16, 39]. In this paper, we call the tutors’ responses to the tutee’s questions that required them to engage in knowledge-building activities the *knowledge-building responses* (KBR). As an example, suppose the tutee asked, “*Why do we perform a transformation on both sides?*”. Tutor’s reply, “*An equation is like a*

balanced weight scale. You do the same thing on both sides to keep it balanced at all times,” is KBR because the tutor provided a complete and correct explanation to make sense of a solution step, which is a *knowledge-building* activity.

Roscoe *et al.* [12] grouped KBR broadly into two categories: elaboration and sense-making (Table 1). The elaboration response provides extended explanations or novel examples to clarify a concept of interest. For example, the tutor’s answer “*one needs to undo all the operations performed on the variable term to solve an equation*” when asked how to know an equation is solved is KBR because the tutor provides an extended explanation to clarify the concept of equation solving. On the other hand, the sense-making response reflects that the tutor realized their errors or realized new inferences based on their prior knowledge. For example, the tutor may mention that he has just learned that “*subtracting a positive number is the same as adding a negative number.*”

Table 1. A summary of knowledge-building response category and its sub-categories. The types of prompt Gabby uses as follow-up tutee inquiry is also shown.

KBR category	KBR Sub-category	Follow-up Tutee Inquiry prompts
Elaboration	Providing extended clarification for a concept of interest	Explanation prompt [Pumping] “ <i>Can you elaborate?</i> ”
	Providing more examples for clarification	Example prompt [Pumping] “ <i>I know little about [x], tell me more?</i> ”
Sense-making	Realizing own errors or misconceptions	Error realization prompt [Splicing] “ <i>I acted according to the rule I have learnt from you. Why am I wrong?</i> ”
	Realizing new inferences based on prior knowledge	Inference prompt [Pumping] “ <i>How to convert the unfamiliar form to the familiar one you’ve taught me before?</i> ”

We hypothesize that when a tutee’s inquiry did not induce the tutors’ KBR, having the tutee ask a follow-up question will increase the tutors’ chance of committing KBR. We, therefore, propose Constructive Tutee Inquiry (CTI) as a sequence of follow-up inquiries to guide the tutor to KBR. CTI consists of an initial tutee inquiry followed by a chain of follow-up inquiries based on tutors’ responses to the previous inquiry. The initial inquiries are the same as the ones that SimStudent asks in traditional APLUS.

4.2 Mechanism of the Follow-Up Tutee Inquiry

A follow-up tutee inquiry can be one of the following prompts: (1) explanation, (2) example, (3) error realization, and (4) inference, as shown in Table 1.

The mechanism of the follow-up tutee inquiry is inspired by how the teacher and students jointly improve the quality of an answer in the classroom proposed by Graesser *et al.* [10]. Teachers follow *pumping*, *splicing* and, *summarization* techniques introduced

in [10] to improve the quality of their students’ answers. Teacher asking for more information when students’ answer is not informative or vague is *pumping*. If students’ answer is error-ridden, a teacher may split the answer into correct and incorrect parts and ask for information about the incorrect part, called *splicing*. Finally, a teacher summarizes the gathered information to the students, called *summarization*.

CTI implements pumping, splicing and, summarization to operate the follow-up tutee inquiries. The follow-up tutee inquiry starts after tutors’ response to an initial tutee inquiry and ends with the tutee’s *summarization*. The following algorithm operates the subsequent follow-up tutee inquiries:

```

IF [tutor’s response is vague] THEN [use the explanation prompt (which is pump)]
ELSE IF [tutor’s response contradicts with already perceived knowledge] THEN [use the error realization prompt (which is splice)]
ELSE IF [tutor’s response agrees with already perceived knowledge] THEN [use example generation prompt (which is pump)]
ELSE IF [tutor’s response reveals tutor is stuck] THEN [use the inference generation prompt (which is pump)]
ELSE [summarize and move to the next scenario]
    
```

Figure 1 shows an example of CTI. It starts with an initial tutee inquiry, like traditional APLUS. Since the tutor provided a vague answer to the initial tutee inquiry, Gabby pumps using the *explanation prompt*. Gabby ends the follow-up tutee inquiry by summarizing because tutor’s response to the *explanation prompt* was not vague and did not contradict any perceived knowledge.

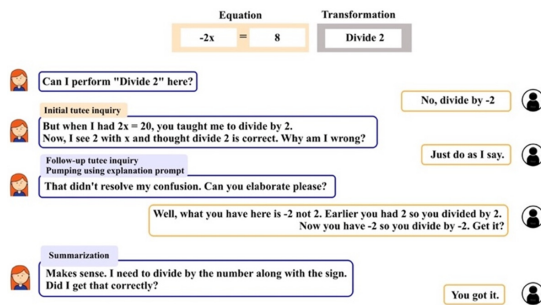


Fig. 1. An example of Constructive Tutee Inquiry

5 Method

This paper explores the following research questions: (1) Does Constructive Tutee Inquiry (CTI) inspire tutors to generate more knowledge-building responses while tutoring? (2) If so, do increased knowledge-building responses facilitate tutor learning?

We conducted a Wizard of Oz experiment (WoZ) to gather early-stage evidence of our proposed Constructive Tutee Inquiry (CTI) effectiveness as a randomized controlled

trial with two conditions: AskMore APLUS as the experimental and Traditional APLUS as the control condition. In WoZ, a human researcher controlled Gabby in AskMore APLUS and manually typed questions according to the CTI model as if a machine generated them. To save the cost of system implementation, AskMore APLUS was built as a chat app running on a standard Web browser by mimicking the appearance of traditional APLUS. We used the same interface to simulate traditional APLUS to avoid any confounding issue due to the difference in the appearance of the interface and system interaction. When participants were using simulated traditional APLUS, the human researcher controlled SimStudent and manually typed questions the same way SimStudent would ask in traditional APLUS. Participants were redirected to another screen containing traditional APLUS whenever they clicked the quiz tab in the chat application. A session manager fed all the tutor demonstrated steps to traditional APLUS to ensure that the quiz reflects the synthetic tutee's knowledge status.

5.1 Structure of the Study

30 middle school students (11 male and 19 female) of 6th–8th grade from various middle schools participated in the study. Participants visited our lab individually and received \$15/h as monetary compensation for their participation.

Participants took a pre-test for 15 min before the intervention. Then they were randomly assigned to one of the conditions ensuring an equal balance of the average pre-test scores between two conditions. 16 participants were assigned to AskMore and 14 to the Traditional APLUS condition. A two-tailed unpaired t-test on the pre-test score confirmed no condition difference between the mean pre-test scores; $M_{AskMore} = 9.9 \pm 5.4$ vs. $M_{Traditional} = 10.6 \pm 4.1$; $t(28) = -0.44$, $p = 0.66$. Participants watched a 10-min video tutorial on the given intervention before using the assigned app. In the video, participants were informed that their goal was to have SimStudent / Gabby pass the quiz. A single intervention session took about 90 min, depending on how quickly the participants achieved their goals. Most of our participants (23 out of 30) came in for the second-day intervention session. Only 18 out of 30 participants met the goal of having their synthetic tutee pass the quiz: 10 for AskMore and 8 for Traditional APLUS. Our data analysis considered all 30 participants irrespective of their synthetic tutee pass the overall quiz. Upon completion of the intervention, participants took 15 min post-test. The tutoring sessions were audio and video recorded, and students were asked to think aloud in both conditions. All interface actions taken by participants and the tutee inquiries and participant responses were logged.

5.2 Measures

The pre- and post-test questions were isomorphic. Both pre- and post-tests consist of 2 tests: (1) The Conceptual Knowledge Test (CKT) contains 9 questions - 2 multiple choice questions and 7 single-choice questions that address various misconceptions of linear algebraic equation solving. An example of a single-choice question that addresses the misconception of zero while solving equations is “The equation $7x + 14 = 0$ is same as $7x = 14$ because the RHS is 0” with Agree/Neutral/Disagree as options. (2) The Procedural Skill Test (PST) contains 10 questions – solving one-step (1 question),

two steps (3 questions), and with variables on both sides (6 questions) equations. An example of two steps equation is “Solve for x: $4x + 15 = 3$ ”. The answers were scored as either 1 for overall correctness or 0 for incorrect or incomplete answers. The highest score any participants could achieve in CKT is, therefore, 9 and in PST is 10.

The participants entered a total of 4605 responses while tutoring the synthetic tutee. Two human coders categorized those responses into “knowledge-building” and “non-knowledge-building” responses. Based on Cohen’s Kappa coefficient, the inter-coder reliability for this coding showed $\kappa = 0.81$.

6 Results

6.1 Test Scores

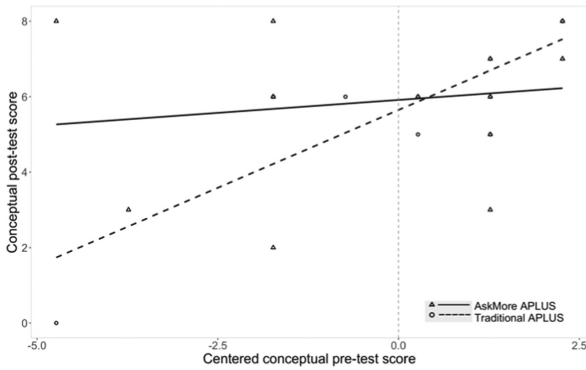
Table 2 shows a summary of average test scores—overall, procedural skill test (PST), and conceptual knowledge test (CKT)—both for Traditional and AskMore APLUS.

We ran a mixed design analysis for both CKT and PST with test-time (pre vs post) as within subject variable and condition (AskMore vs Traditional) as a between subject variable. There was no condition difference for post-test scores in CKT ($M_{AskMore} = 5.9 \pm 1.8$, $M_{Traditional} = 6.0 \pm 2.0$, $F(1,28) = 0.30$, $p = 0.60$) and PST ($M_{AskMore} = 6.3 \pm 3.8$, $M_{Traditional} = 7.0 \pm 2.2$, $F(1, 28) = 0.14$, $p = 0.71$). However, test-time (pre vs. post) was a main effect for both CKT ($M_{Pre} = 4.7 \pm 2.1$, $M_{Post} = 6.0 \pm 2.0$, $F(1,28) = 9.56$, $p < 0.01$) and PST ($M_{Pre} = 5.5 \pm 3.3$, $M_{Post} = 6.6 \pm 3.1$, $F(1, 28) = 5.53$, $p < 0.05$). *Tutors in both conditions showed an equal amount of learning from pre to post tests on both CKT and PST.*

Table 2. Average pre- and post-test scores in AskMore and Traditional APLUS condition.

Condition	Overall (pre)	Overall (post)	PST (pre)	PST (post)	CKT (pre)	CKT (post)
AskMore	9.9 ± 5.4	12.2 ± 5.3	5.4 ± 3.7	6.3 ± 3.8	4.4 ± 2.2	5.9 ± 1.8
Traditional	10.6 ± 4.1	13.0 ± 3.5	5.6 ± 3.0	7.0 ± 2.2	5.1 ± 2.0	6.0 ± 2.0

An aptitude treatment interaction (ATI) found for the conceptual knowledge test (CKT). Figure 2 shows a scattered plot with the centered conceptual pre-test score on the x-axis and conceptual post-test score on the y-axis. In the plot, *among those who scored below average on the pre-test, AskMore tutors outperformed Traditional tutors on the post-test.* We ran a two-way ANOVA with the conceptual post-test as the dependent variable, conceptual pre-test and condition (AskMore vs. Traditional APLUS) as the independent variables. The interaction between conceptual pre-test and condition was statistically significant; $F(1, 26) = 5.70$, $p < 0.05$. However, an ATI was not observed for the procedural skill test. The same two-way ANOVA, as shown above, did not show a statistically significant interaction term among the procedural pre-test score and condition; $F(1, 26) = 0.05$, $p = 0.83$. *Tutors with lower prior competency (below-average pre-test score) learned more conceptual knowledge when using AskMore APLUS than Traditional APLUS.*



16 datapoints in triangle represent AskMore APLUS tutors and 14 datapoints in circle represent Traditional APLUS tutors. There are many overlapping datapoints due to more than one tutors achieving the same test scores. For tutors with the below-average pre-test score (left on the dotted vertical line), AskMore tutors outperformed Traditional tutors on the post-test.

Fig. 2. Centered pre- vs. post-test score for CKT between AskMore vs. Traditional tutors.

6.2 Effect of Responding to the Follow-Up Tutee Inquiries

We hypothesized that simply answering more tutee inquiries (regardless it induced knowledge-building responses or not) facilitated conceptual learning. A simple regression model with the normalized learning gain on CKT, i.e., $(\text{posttest} - \text{pretest}) / (1 - \text{pretest})$, as a dependent variable and the number of inquiries tutor answered as an independent variable did not reveal the number of inquiries answered as a reliable predictor for the learning gain on CKT; $F(1,14) = 0.21, p = 0.66$. *Simply answering more tutee inquiries did not predict conceptual learning.*

6.3 Learning to Generate Knowledge-Building Responses

We investigated if generating more KBR promoted conceptual learning. A regression analysis fitting the conceptual post-test with the conceptual pre-test and normalized KBR count confirmed that normalized KBR is a reliable predictor; $F(1, 27) = 14.11, p < 0.01$. The regression model suggests that if *tutors committed one more knowledge-building response than the average, they would have ended up with a 1.0% increase in their conceptual post-test score*. However, the equivalent regression analysis suggests that *committing to more knowledge-building responses did not help procedural learning*; $F(1, 27) = 1.55, p = 0.22$.

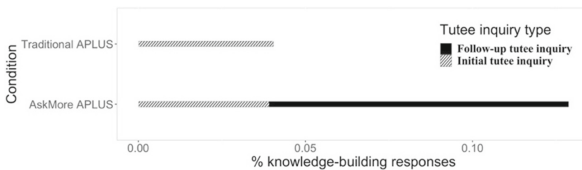


Fig. 3. Boxplot of average % knowledge-building responses generated by tutors due to initial tutee inquiries and follow-up tutee inquiries in both conditions.

We also found a significant positive correlation between the number of follow-up tutee inquiries asked and KBR generated by tutors; $r = 0.51, p < 0.01$. *About 24% of the follow-up tutee inquiries yielded knowledge-building responses.*

To further understand the process of generating KBR between the conditions, we divided total tutee inquiries into two parts (1) initial tutee inquiries (ITI) and (2) follow-up tutee inquiries (FTI). ITI was available in both conditions where FTI was only available in AskMore APLUS. We tagged the KBR generated by AskMore APLUS tutors into two categories (1) KBR due to initial tutee inquiries (KBR-ITI) and (2) KBR due to follow-up tutee inquiries (KBR-FTI).

The boxplot in Fig. 3 shows that 12% of the total responses by the AskMore tutors were KBR, whereas, for the Traditional tutors, it is only 4%. A *t*-test confirmed that AskMore tutors had a higher ratio of generating KBR than the Traditional tutors; $t(21) = -4.89, p < .01$. The boxplot also revealed that the percentage of KBR-ITI was fairly equal in both conditions. A *t*-test confirmed that there was no difference between the average percent KBR ($M_{AskMore} = .04, M_{Traditional} = .04$) due to initial tutee inquiry between conditions $t(28) = .09, p = .93$.

The above observations suggest that the follow-up tutee inquiry resulted in increased knowledge-building responses (i.e., KBR-FTI) generated by AskMore tutors, which further facilitated conceptual learning.

Table 3 shows that the AskMore low prior tutors were asked 26 initial tutee inquiries on average, whereas Traditional low prior tutors were asked 25. A *t*-test revealed low prior tutors in both conditions ($M_{AskMore} = 2, M_{Traditional} = 3$) generated equal number of KBR-ITI on average; $t(9) = 1.60, p = 0.14$. However, AskMore low prior tutors were additionally asked 49 follow-up tutee inquiries on average that resulted in an additional 5 more KBR (i.e., KBR-FTI) during the entire tutoring session.

Table 3. Average number of initial tutee inquiries (ITI) and follow-up tutee inquiry (FTI); and the average number of resultant KBR due to ITI (KBR-ITI) and KBR due to FTI (KBR-FTI)

		Total inquiries		Total KBR	
Prior	Condition	ITI	FTI	KBR-ITI	KBR-FTI
Low	AskMore	26	49	2	5
	Traditional	25	–	3	–
High	AskMore	21	30	2	4
	Traditional	28	–	2	–

In sum, it was the follow-up tutee inquiries in AskMore APLUS that assisted the low prior tutors to generate more knowledge-building responses, which in turn facilitated their conceptual learning.

However, the same finding does not apply to the high prior tutors even though they generated 6 KBR on average while tutoring. We hypothesized that different types of KBR (Table 1) affect tutor learning differently. However, the current study was not designed to collect data to test this hypothesis. A future investigation is needed.

6.4 Learning to Review Resources

Our study revealed that reviewing the resources positively impacts conceptual learning regardless of the condition. A regression analysis with conceptual gain as the dependent variable and resource usage as the independent predictor confirmed that resource usage is a reliable predictor; $F(1, 28) = 4.39, p < 0.05$.

Did follow-up tutee inquiries inspire the low prior tutors to review the resources more, and did more frequent resource reviews result in better conceptual learning? Follow-up tutee inquiries inspired the AskMore low prior tutors to review the resource tabs 11.9 times, whereas Traditional low prior tutors reviewed the resources 8.8 times on average while tutoring. However, the t-test revealed no condition difference in resource review count; $t(13) = -1.01, p = .33$. A qualitative analysis of dialog data disclosed that the low prior tutors in AskMore APLUS tended to review the resources more often to better answer Gabby's inquiries. Figure 4 shows an example conversation in which an AskMore low prior tutor reviewed the resources while answering follow-up tutee inquiries. The tutor reviewed *UNIT OVERVIEW*, one of the resources available for tutors in APLUS. Although *UNIT OVERVIEW* does not contain a direct answer to Gabby's inquiry, it contains information like "To solve an equation you need to do mathematical transformations until at the end you have " $x = a \text{ number}$ " The tutor elaborated on this information to come up with a better answer to Gabby's inquiry.

7 Discussion

Our current data showed an interesting ATI—generating more knowledge-building responses (KBR) in AskMore APLUS facilitated learning conceptual knowledge only for the low prior tutors. Our hypothesis that CTI encouraged low prior tutors to generate more KBR was supported. *About 24% of the follow-up tutee inquiries (asked by Gabby) yielded knowledge-building responses (by the tutors).*

Our data also revealed that conceptual learning is highly correlated with reviewing the resources in the application. We further hypothesized that CTI inspired the low prior tutors to review the resources more often, which is another reason for the ATI we found. This hypothesis was not supported. However, conversational data showed low prior tutors frequently reviewed the resources to come up with a better answer to the follow-up tutee inquiry.

All these findings apply to the low prior tutors only. The number of KBR did not have a notable impact on the high prior tutors' learning. We hypothesize that different types of KBR (Table 1) have different contributions to learning. The current data do not allow us to conduct rigorous analysis on this hypothesis due to the limited number of participants. We aim to investigate this hypothesis at length in our future work.

8 Conclusion

We found that Constructive Tutee Inquiry (CTI) helped low prior students provide more knowledge-building responses (KBR) by prompting them to elaborate the confusing explanations and motivating them to commit sense-making reflections. The data also

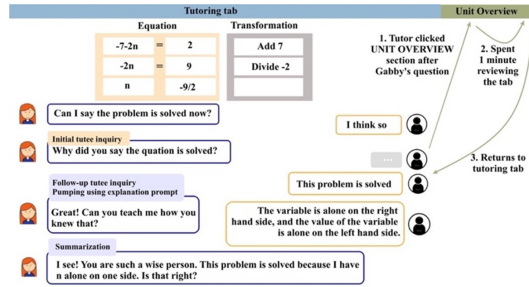


Fig. 4. Example of reviewing resources after follow-up tutee inquiries

showed that the KBR further facilitated learning conceptual knowledge. Understanding how CTI could assist in procedural skills learning is within the scope of our future work. Our focus of interest also goes to the high prior students who did not get the advantage of CTI as effectively as the low prior students. We believe that learning by teaching, with the advanced teachable agent technology, can offer rich learning opportunities to diverse students. Understanding how KBR types affect tutor learning differently would allow us to investigate how to make the CTI more effective.

Acknowledgment. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant No. R305A180319 and National Science Foundation Grant No. 1643185 to North Carolina State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

1. Annis, L.F.: The processes and effects of peer tutoring. *Hum. Learn.* **2**(1), 39–47 (1983)
2. Bargh, J.A., Schul, Y.: On the cognitive benefits of teaching. *J. Educ. Psychol.* **72**(5), 593–604 (1980)
3. Cohen, P.A., Kulik, J.A., Kulik, C.L.C.: Education outcomes of tutoring: a meta-analysis of findings. *Am. Educ. Res. J.* **19**(2), 237–248 (1982)
4. Renkl, A.: Learning for later teaching: an exploration of mediational links between teaching expectancy and learning results. *Learn. Instr.* **5**(1), 21–36 (1995)
5. Roscoe, R.D., Chi, M.T.H.: Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Rev. Educ. Res.* **77**(4), 534–574 (2007)
6. Chi, M.T.H., et al.: Learning from human tutoring. *Cogn. Sci.* **25**, 471–533 (2001)
7. Cohen, E.G.: Restructuring the classroom: conditions for productive small groups. *Rev. Educ. Res.* **64**(1), 1–35 (1994)
8. Devin-Sheehan, L., Feldman, R.S., Allen, V.L.: Research on children tutoring children: a critical review. *Rev. Educ. Res.* **46**(3), 355–385 (1976)
9. Gartner, A., Kohler, M., Riessman, F.: *Children Teach Children: Learning by Teaching*. Harper & Row, New York (1971)
10. Graesser, A.C., Person, N.K., Magliano, J.P.: Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Appl. Cogn. Psychol.* **9**(6), 495–522 (1995)

11. Roscoe, R.D.: Opportunities and barriers for tutor learning: knowledge-building, metacognition, and motivation. University of Pittsburgh (2008)
12. Roscoe, R.D., Chi, M.: Tutor learning: the role of explaining and responding to questions. *Instr. Sci.* **36**(4), 321–350 (2008)
13. Butler, D.L.: Structuring instruction to promote self-regulated learning by adolescents and adults with learning disabilities. *Exceptionality* **11**(1), 39–60 (2003)
14. Hong, H.-Y., et al.: Advancing third graders' reading comprehension through collaborative Knowledge Building: a comparative study in Taiwan. *Comput. Educ.* **157**, 103962 (2020)
15. Cohen, J.: Theoretical considerations of peer tutoring. *Psychol. Sch.* **23**(2), 175–186 (1986)
16. Roscoe, R.D., Chi, M.T.H.: The influence of the tutee in learning by peer tutoring. In: Forbus, K., Gentner, D, Regier, T. (eds.) *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pp. 1179–1184. Erlbaum, Mahwah (2004)
17. Roscoe, R.D.: Self-monitoring and knowledge-building in learning by teaching. *Instr. Sci.* **42**(3), 327–351 (2013). <https://doi.org/10.1007/s11251-013-9283-4>
18. Chin, C., Osborne, J.: Students' questions: a potential resource for teaching and learning science. *Stud. Sci. Educ.* **44**(1), 1–39 (2008)
19. Watts, M., et al.: Prompting teachers' constructive reflection: pupils' questions as critical incidents. *Int. J. Sci. Educ.* **19**(9), 1025–1037 (1997)
20. Graesser, A.C., Person, N.K.: Question asking during tutoring. *Am. Educ. Res. J.* **31**(1), 104–137 (1994)
21. Van den Boom, G., Paas, F., Van Merriënboer, J.J.: Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learn. Instr.* **17**(5), 532–548 (2007)
22. Rittle-Johnson, B., Lochr, A.M.: Eliciting explanations: constraints on when self-explanation aids learning. *Psychon. Bull. Rev.* **24**(5), 1501–1510 (2016). <https://doi.org/10.3758/s13423-016-1079-5>
23. Matsuda, N., et al.: Studying the effect of tutor learning using a teachable agent that asks the student tutor for explanations. In: Sugimoto, M., et al. (eds.) *Proceedings of the International Conference on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL 2012)*, pp. 25–32. IEEE Computer Society, Los Alamitos (2012)
24. Biswas, G., et al.: Learning by teaching: a new agent paradigm for educational software. *J. Appl. Artif. Intell.* **19**(3 & 4), 363–392 (2005)
25. Bredeweg, B., et al.: DynaLearn - engaging and informed tools for learning conceptual system knowledge. In: Pirrone, R., Azevedo, R., Biswas, G. (eds.) *Cognitive and Metacognitive Educational Systems (MCES 2009)*, AAAI Fall Symposium, pp. 46–51. AAAI Press, Arlington (2009)
26. Ketamo, H., Kiili, K.: Conceptual change takes time: game based learning cannot be only supplementary amusement. *J. Educ. Multimed. Hypermedia* **19**(4), 399–419 (2010)
27. Nichols, D.M.: *Intelligent Student Systems: an Application of Viewpoints to Intelligent Learning Environments*. Lancaster University, Lancaster (1993)
28. Pareto, L., Arvemo, T., Dahl, Y., Haake, M., Gulz, A.: A teachable-agent arithmetic game's effects on mathematics understanding, attitude and self-efficacy. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *AIED 2011. LNCS (LNAI)*, vol. 6738, pp. 247–255. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21869-9_33
29. Reif, F., Scott, L.A.: Teaching scientific thinking skills: students and computers coaching each other. *Am. J. Phys.* **67**(9), 819–831 (1999)
30. Matsuda, N.: *Applying machine learning to cognitive modeling for cognitive tutors*. [Technical report]. Carnegie Mellon University. School of Computer Science. Machine Learning Dept. (2006). Carnegie Mellon University, School of Computer Science, Machine Learning Dept., Pittsburgh [15] p.

31. Chi, M.T.H., et al.: Eliciting self-explanations improves understanding. *Cogn. Sci.* **18**(3), 439–477 (1994)
32. Kwon, K., Kumalasari, C.D., Howland, J.L.: Self-explanation prompts on problem-solving performance in an interactive learning environment. *J. Interact. Online Learn.* **10**(2), 1–17 (2011)
33. Rosé, C.P., Torrey, C.: Interactivity versus expectation: eliciting learning oriented behavior with tutorial dialogue systems. In: *Proceedings of Interact 2005* (2005)
34. DeCaro, M.S., Rittle-Johnson, B.: Exploring mathematics problems prepares children to learn from instruction. *J. Exp. Child Psychol.* **113**(4), 552–568 (2012)
35. Baker, R.S., Corbett, A.T., Koedinger, K.R.: Learning to distinguish between representations of data: a cognitive tutor that uses contrasting cases. In: *Proceedings of the 6th International Conference on Learning Sciences* (2004)
36. Sidney, P.G., Hattikudur, S., Alibali, M.W.: How do contrasting cases and self-explanation promote learning? Evidence from fraction division. *Learn. Instr.* **40**, 29–38 (2015)
37. Looi, C.K., Wu, L.: Design of agent tutee’s question prompts to engage student’s role-playing as tutor in a learning-by-teaching agent environment. In: *Proceedings of the 8th International Conference on International Conference for the Learning Sciences-Volume International Society of the Learning Sciences* (2008)
38. Matsuda, N., et al.: Learning by teaching SimStudent – an initial classroom baseline study comparing with cognitive tutor. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *AIED 2011. LNCS (LNAI)*, vol. 6738, pp. 213–221. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21869-9_29
39. Berghmans, I., et al.: A typology of approaches to peer tutoring. Unraveling peer tutors’ behavioural strategies. *Eur. J. Psychol. Educ.* **28**(3), 703–723 (2013)