What Level of Tutor Interaction is Best?

Leena RAZZAQ, Neil T. HEFFERNAN and Robert W. LINDEMAN
Computer Science Department, Worcester Polytechnic Institute
100 Institute Road, Worcester, Massachusetts, USA

Abstract. Razzaq and Heffernan (2006) showed that scaffolding compared to hints on demand in an intelligent tutoring system could lead to higher averages on a middle school mathematics post-test. There were significant differences in performance by condition on individual items. For an item that proved to be difficult for all of the students on the pretest, an ANOVA showed that scaffolding helped significantly ($p < 0.01$). We speculated that the scaffolding had a greater positive effect on learning for this item because it was much more difficult for the students than the other items. We thought that this result warranted a closer look at the link between the difficulty of an item and the effectiveness of scaffolding. In this paper, we report on an experiment that examines the effect of math proficiency and the level of interaction on learning. We found an interesting interaction between the level of interaction and math proficiency where less-proficient students benefited from more tutor interaction and more-proficient students benefited from less interaction.

Keywords. Intelligent tutoring, scaffolding, feedback, interactive help, math tutoring

1. Introduction

Several studies in the literature have argued that human tutors that are more interactive lead to better learning and can achieve greater learning gains. In a comparison of Socratic and didactic tutoring strategies, Core, Moore and Zinn [1] found that the more interactive (based on words produced by students) Socratic tutorial dialogs had a greater correlation with learning. Katz, Connelly and Allbritton [2] found that students learned more when they participated in post practice dialogs with a tutor than students who did not. Chi, Siler, Jeong, Yamauchi and Hausmann [3] found that students who engaged in a more interactive style of human tutoring were "able to transfer their knowledge better than the students in the didactic style of tutoring."

We are interested in the role of tutor interaction in intelligent tutoring systems (ITS) and similar results have been found in other studies of interactive ITS. Razzaq and Heffernan [4] found a positive effect on learning when students worked on solving equations in an ITS called E-tutor which incorporated tutorial dialogs. E-tutor is a model-tracing tutor that is able to carry on a coherent dialog that consists of breaking down problems into smaller steps and asking new questions about those steps, rather than simply giving hints. Tutorial dialogs were chosen from transcripts of human tutoring sessions and were incorporated in E-tutor. E-tutor does not have a hint button and when students make errors they are presented with a tutorial dialog if one is available. The student must respond to the dialog to exit it and return to solving the original problem. Students stay in the dialog loop until they respond correctly or the tutor has run out of dialog. When the tutor has run out of dialog, the last tutorial response presents the student with the correct action and input similar to the bottom-out hint in a hint sequence. A close mapping between the human tutor dialog and the E-tutor dialog was attempted. E-tutor was compared to a control version that did not engage in dialog, but did have a hint button to supply hints to students when they asked for them. E-tutor with dialog led to better learning and represents a more interactive tutor than the "*hints on demand*" control condition.

It seems that a positive relationship between learning and tutor interaction exists, and we would expect students to learn more whenever they engage in interactive tutoring conditions than in less interactive conditions such as reading text. There is, however, evidence that this is not always the case. VanLehn, Graesser, Jackson, Jordan, Olney, & Rose [5] reviewed several studies that hypothesize that the relationship between interactivity and learning exists, as well as a few studies that failed to find evidence for this relationship. VanLehn et al [5] found that when students found the material to be difficult, tutoring was more effective than having the students read an explanation of how to solve a problem. However, this was not the case when the students found the material to be at their level: interactive tutoring was not more effective than canned-text. We found a similar effect in a study in Razzaq and Heffernan [6].

We used the ASSISTment system [7], a web-based tutoring system that blends assisting students with assessing their knowledge, in this study. The system tutors students on 7^{th}, 8^{th} and 10^{th} grade mathematics content that is based on Massachusetts Comprehensive Assessment System (MCAS) released test items. There are currently over 1000 students using the ASSISTment system as part of their mathematics class.

Our results show that students are learning 8^{th} grade math at the computer by using the system [7], but we were not certain if this is due to students getting more practice on math problems or more due to the intelligent tutoring that we created. Students are forced to participate in this intelligent tutoring if they get a problem incorrect. In Razzaq and Heffernan [6], we conducted a simple experiment to see if students learned on a set of 4 problems if they were forced to do the scaffolding questions, which would ASK them to complete each step required to solve a problem, compared with being given hints on demand, which would TELL them the same information without expecting an answer to each step. In the study, the "*scaffolding + hints*" condition represents a more interactive learning experience than the "*hints on demand*" condition.

The results of the Razzaq and Heffernan [6] experiment showed that *scaffolding + hints* led to higher averages on a post-test than *hints on demand*, although it was not statistically significant. When we compared scores on particular post-test items that students had seen as pretest items, we found significant differences by condition. For one item, which concerned finding the y-intercept from an equation, the ANOVA showed a statistically significant ($p < 0.01$) with an effect size of 0.85. This item on finding the y-intercept from an equation proved to be a difficult problem for all of the students on the pretest and scaffolding helped significantly. We speculated that the *scaffolding + hints* had a greater positive effect on learning for the first pretest item because it was much more difficult for the students than the second pretest item. We thought that this result warranted a closer look at the link between the difficulty of an item and the effectiveness of scaffolding.

In this study, we look at three different conditions, in the ASSISTment system, which have varying levels of tutor interaction. The first two conditions are the same as in [6] (*scaffolding + hints* and *hints on demand*). The third condition is a *delayed feedback* condition where students get no feedback from the tutor until they finish all of the problems in the experiment, whereupon they receive worked out solutions to all of the problems. Both immediate and delayed feedbacks have been shown to be helpful to students (Mathan and Koedinger) [8]. In Razzaq and Heffernan [4], our human tutor provided immediate feedback to student errors, on most occasions, keeping students on the correct solution path. There was one out of 26 problems where the human tutor gave delayed feedback to a student error. In this instance, the tutor allowed the student to continue where she had made an error and then let her check her work and find the error seemingly promoting evaluative skills. This happened with the student who was taking a more advanced algebra class and was slightly more advanced than the other students in the study. However, for the other 25 problems, the tutor provided immediate feedback to promote the development of generative skills. This agrees with McArthur et al [9] in their examination of tutoring techniques in algebra where they collected one and a half hours of videotaped one-on-one tutoring sessions. "In fact, for every student error we recorded, there was a remedial response. At least the tutors we observed were apparently not willing to let students explore on their own and perhaps discover their own errors…teachers may believe that such explorations too frequently lead to unprofitable confusion for the student."

The purpose of this experiment was to determine which level of interaction worked best for students learning math: *scaffolding + hints*, *hints on demand* or *delayed feedback*, and how their math proficiency influenced the effectiveness of the feedback provided.

2. The ASSISTment System

Limited classroom time available in middle school mathematics classes requires teachers to choose between time spent assisting students' development and time spent assessing their abilities. To help resolve this dilemma, assistance and assessment are integrated in a web-based system called the ASSISTment [1] System which offers instruction to students while providing a more detailed evaluation of their abilities to teachers than is available under most current approaches. Many teachers use the system by requiring their students to work on the ASSISTment website for about 20 minutes per week in their schools' computer labs. Each week when students work on the website, the system "learns" more about the students' abilities and thus, it can hypothetically provide increasingly accurate predictions of how they will do on a standardized mathematics test [10]. The ASSISTment System is being built to identify the difficulties individual students - and the class as a whole – are having. It is intended that teachers will be able to use this detailed feedback to tailor their instruction to focus on the particular difficulties identified by the system. Unlike other assessment systems, the ASSISTment technology also provides students with intelligent tutoring assistance while the assessment information is being collected. The hypothesis is that ASSISTments can do a better job of assessing student knowledge limitations than practice tests by taking the amount and nature of the assistance that students receive into account.

It is easy to carry out randomized controlled experiments in the ASSISTment System [11]. Items are arranged in modules in the system. The module can be conceptually subdivided into two main pieces: the module itself, and sections. The

[1] The term ASSISTment was coined by Kenneth Koedinger and blends Assisting and Assessment.

module is composed of one or more sections, with each section containing items or other sections. This recursive structure allows for a rich hierarchy of different types of sections and problems. The section component is an abstraction for a particular listing of problems. This abstraction has been extended to implement our current section types, and allows for future expansion of the module unit. Currently existing section types include "Linear" (problems or sub-sections are presented in linear order), "Random" (problems or sub-sections are presented in a pseudo-random order), and "Choose One" (a single problem or sub-section is selected pseudo-randomly from a list, the others are ignored).

3. Experimental Design

Problems in this experiment addressed the topic of interpreting linear equations. Figure1 shows an item used in the experiment. The item shows the different feedback that students can receive once they have answered a question incorrectly. (We call this top-level question the original question.)

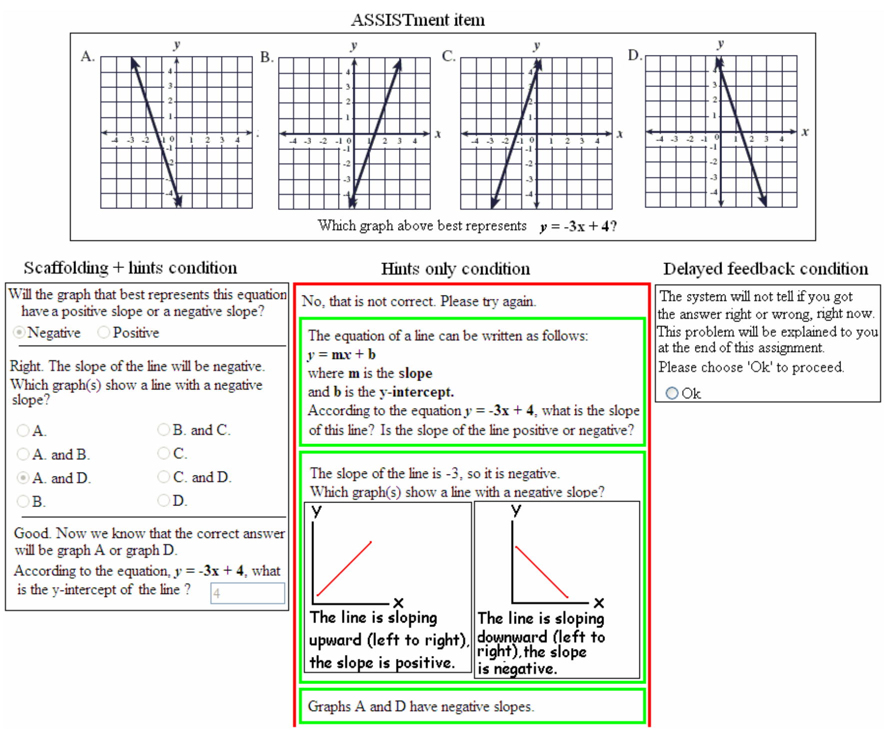


Figure 1. An ASSISTment item showing 3 different levels of interaction.

A student in the *scaffolding + hints* condition is immediately presented with the first scaffolding question. Students must answer a scaffolding question correctly to proceed and receive the next scaffolding question (or finish the problem). Students can ask for hints on the scaffolding questions, but not on the original question. They cannot go back and answer the original question, but rather are forced to work through the problem.

Students in the hints condition receive a message, outlined in red, of “No, that is not correct. Please try again.” The hints, outlined in green, appear when the student requests them by pressing the Hint button. Students do not see the hints unless they ask for them. Figure 1 shows a sequence of three hints to solving the problem. The full sequence has seven hints in total, with a bottom-out hint at the end of the sequence. The bottom-out hint gives the student the answer to the problem.

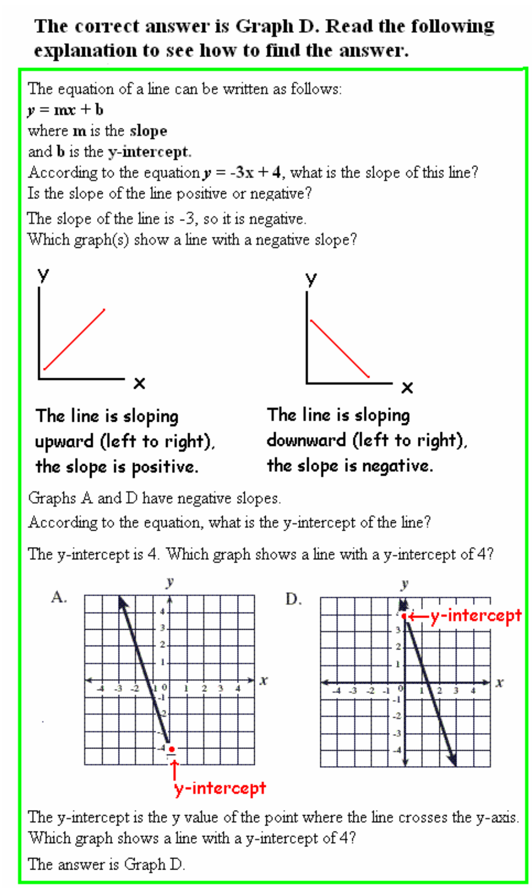


Figure 2. The delayed feedback condition provided explanations at the end of the assignment.

Students in the *delayed feedback* condition did not receive any feedback on the problems that they did until they had finished all of the problems. At that time, the students were presented with the answers and explanations of how to solve the problems. Figure 2 shows the explanation that students in the *delayed feedback* condition received for the item shown in Figure 1.

Based on the results of Razzaq and Heffernan [6], we hypothesized that less-proficient students would need more interaction and benefit more from the scaffolding than more-proficient students.

For this experiment, the number of problems was held constant, but students took as much time as they needed to finish all of the problems. Students were presented with two pretest problems, four experiment problems and four post-test problems that addressed the topic of interpreting linear equations. There were three versions of the experiment problems, one for each condition. Two of the pretest problems were repeated in the post-test.

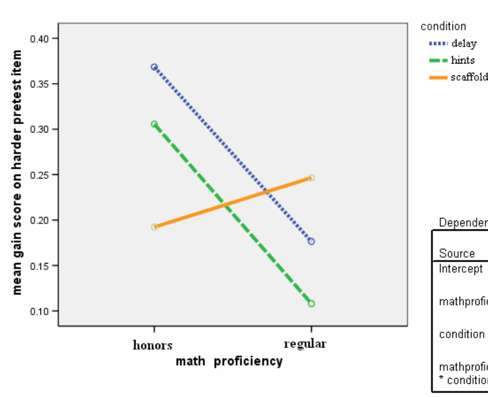
The ASSISTment system randomly assigned students to the *scaffolding + hints*, *hints on demand* or *delayed feedback* conditions with equal probability. There were 366 eighth grade students from the Worcester Public Schools in Worcester, Massachusetts who participated in the experiment: 131 students were in honors level classes and 235 were in regular math classes. There were 119 students in the *scaffolding + hints* condition, 124 students in the *hints on demand* condition and 123 students in the *delayed feedback* condition. The students worked on the problems during their regular math classes.

4. Analysis

We excluded students who got all of the pretest problems correct from the analysis because we assumed that they knew the material. Fifty-one students got perfect scores on the pretest and were excluded. We first checked to make sure that the groups were not significantly different at pretest by doing an ANOVA on pretest averages by condition. There was no significant difference between groups at pretest ($p = 0.556$). Students learned overall from pretest to post-test ($p = 0.005$).

When we look at performance on the post-test by condition, the difference is not significant; however there is an interesting trend when we separate students by math proficiency. There is a significant interaction ($p = 0.045$) between condition and math proficiency on the post-test average. The regular students seem to benefit more from the *scaffolding + hints* condition, while honors students seem to benefit more from the *delayed feedback* condition. We also ran an ANOVA on the gain score of the two pretest items. Again, there is no statistically significant difference by condition, but there is a significant interaction between math proficiency and condition ($p = 0.036$), where once more, honors students do best in the *delayed feedback* condition and regular students do best in the *scaffolding + hints* condition.

We decided to take a closer look at the item that proved most difficult for students. The problem concerned finding the y-intercept from an equation and was presented to students in the pretest and again in the post-test. We did a one-way ANOVA using math proficiency as a covariate. An interaction between condition and math proficiency is evident ($p = 0.078$); honors students performed best when they received *delayed feedback* and the regular students performed best when they received *scaffolding + hints*.



Descriptive Statistics

Dependent Variable: gainscore_y_intercept

mathproficiency	condition	Mean	Std. Deviation	N
honors	d	.3684	.48885	38
	h	.3056	.46718	36
	s	.1923	.40192	26
	Total	.3000	.46057	100
regular	d	.1765	.38405	68
	h	.1081	.31264	74
	s	.2466	.43400	73
	Total	.1767	.38234	215

Tests of Between-Subjects Effects

Dependent Variable: gainscore_y_intercept

Source		Type III Sum of Squares	df	F	Sig.
Intercept	Hypothesis	14.522	1	124.897	.007
	Error	.237	2.039		
mathproficiency	Hypothesis	.835	1	1.968	.295
	Error	.853	2.010		
condition	Hypothesis	.232	2	.272	.786
	Error	.852	2		
mathproficiency * condition	Hypothesis	.852	2	2.577	.078
	Error	51.099	309		

Figure 3. There are significant interactions between condition and math proficiency.

5. Discussion

We interpret the low p-values on the interaction term to mean that there are different rates of learning on the single items based upon the interaction between the level of math proficiency and condition. Students who come in with less knowledge benefit more from the *scaffolding + hints* than students who come in with more knowledge. Students who come in with more knowledge benefit from the *delayed feedback* more than the other groups.

The results of this experiment were surprising. We did find evidence to support the interaction hypothesis for regular students. The regular students performed best in the *scaffolding + hints* condition, which is the most interactive condition. We did not expect students in the *delayed feedback* condition to learn more than in other groups, however, the honors students did better in this condition than in the *scaffolding + hints* or *hints on demand* conditions.

One possible explanation is that less-proficient students benefit from more interaction and coaching through each step to solve a problem while more-proficient students benefit from seeing problems worked out and seeing the big picture. Another possible explanation, put forth by one of the eighth grade teachers, is that honors students are often more competitive and like to know how they do on their work. The *delayed feedback* group had to wait until the end of the assignment to see whether they got the questions right or wrong. Perhaps the honors students ended up reading through the explanations more carefully than they would have read the scaffolding questions or hints because they were forced to wait for their results.

Chi [12] found a difference in the way that students used worked examples based on their proficiency in problem-solving. "… we find that the Good students use the examples in a very different way from the Poor students. In general, Good students, during problem solving, use the examples for a specific reference, whereas Poor students reread them as if to search for a solution." Although we did not present worked examples to the students in the *delayed feedback* condition during the experiment, the "worked solutions" to the experiment problems may have behaved as worked examples to the post-test problems.

The students in the *hints on demand* condition did not perform as well as the *delayed feedback* groups for both more-proficient and less-proficient students. One possible explanation is the more proactive nature of the *delayed feedback* explanations. Murray and VanLehn [13] found that proactive help was more effective for some students. "Proactive help when a student would otherwise flounder can save time, prevent confusion, provide valuable information at a time when the student is prepared and motivated to learn it, and avoid the negative affective consequences of frustration and failure." In the ASSISTment system, students only see hints if they ask for them and they are less likely to ask for hints on multiple choice questions when they can guess more easily.

We believe the results of this experiment provide further evidence of the interaction hypothesis for less-proficient students. However, the interaction between condition and math proficiency presents a good case for tailoring tutor interaction to types of students to maximize their learning.

Acknowledgements

We would like to thank all of the people associated with creating the ASSISTment system listed at www.ASSISTment.org including investigators Kenneth Koedinger and Brian Junker at Carnegie Mellon. We would also like to acknowledge funding from the US Department of Education, the National Science Foundation, the Office of Naval Research and the Spencer Foundation. All of the opinions expressed in this paper are those solely of the authors and not those of our funding organizations.

References

[1] Core, M. G., Moore, J. D., and Zinn, C. The Role of Initiative in Tutorial Dialogue, the 10th Conference of the European Chapter of the Association for Computational Linguistics, Budapest, Hungary, April, 2003.

[2] Katz, S., Allbritton, D., Connelly, J. Going beyond the problem given: How human tutors use post-solution discussions to support transfer. International Journal of Artificial Intelligence in Education 13 (2003) 79-116.

[3] Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from tutoring. *Cognitive Science*, 25:471-533.

[4] Razzaq, L. & Heffernan, N. T. (2004) Tutorial dialog in an equation solving intelligent tutoring system. In J. C. Lester, R. M. Vicari, & F. Parguacu (Eds.) Proceedings of 7th Annual Intelligent Tutoring Systems Conference, Maceio, Brazil. Pages 851-853.

[5] VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. When is reading just as effective as one-on-one interactive human tutoring? In Proceedings of the 27th Annual Meeting of the Cognitive Science Society (pp. 2259-2264). Mahwah, NJ: Erlbaum. (2005) 2259-2264.

[6] Razzaq, L., Heffernan, N. T. (2006). Scaffolding vs. hints in the ASSISTment System. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. Springer-Verlag: Berlin. pp. 635-644. 2006.

[7] Razzaq, L., Heffernan, N. T., Koedinger, K. R., Feng, M., Nuzzo-Jones, G., Junker, B. Macasek, M. A., Rasmussen, K. P., Turner, T. E., & Walonoski, J. A. (2007). Blending Assessment and Instructional Assistance. In Nadia Nedjah, Luiza deMacedo Mourelle, Mario Neto Borges and Nival Nunesde Almeida (Eds). Intelligent Educational Machines within the Intelligent Systems Engineering Book Series. 23-49 Springer Berlin / Heidelberg.

[8] Mathan, S. & Koedinger, K. R. (2003). Recasting the Feedback Debate: Benefits of Tutoring Error Detection and Correction Skills. In Hoppe, Verdejo & Kay (Eds.), *Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies. Proceedings of AI-ED 2003* (pp. 39-46). Amsterdam, IOS Press.

[9] McArthur, D., Stasz, C., & Zmuidzinas, M. (1990) Tutoring techniques in algebra. *Cognition and Instruction*. 7 (pp. 197-244.)

[10] Feng, M., Heffernan, N. T., Koedinger, K. Predicting State Test Scores Better with Intelligent Tutoring Systems: Developing Metrics to Measure Assistance Required. Submitted to the 8th International Conference on Intelligent Tutoring Systems (2005).

[11] Nuzzo-Jones, G., Walonoski, J. A., Heffernan, N. T., Livak, T. The eXtensible Tutor Architecture: A New Foundation for ITS. In C.K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.) *Proceedings of the 12th Artificial Intelligence In Education*, Amsterdam: ISO Press (2005) 902-904

[12] Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145 – 182.

[13] Murray, C., VanLehn K. (2006) A Comparison of Decision-Theoretic, Fixed-Policy and Random Tutorial Action Selection. In Ikeda, Ashley & Chan (Eds.). Proceedings of the 8th International Conference on Intelligent Tutoring Systems. Springer-Verlag: Berlin. pp. 116-123. 2006.