

Noticing Relevant Feedback Improves Learning in an Intelligent Tutoring System for Peer Tutoring

Erin Walker¹, Nikol Rummel², Sean Walker³, Kenneth R. Koedinger³

¹School of Computing, CIDSE, Arizona State University, USA

²Institute of Psychology, Ruhr-Universität Bochum, Germany

³Human-Computer Interaction Institute, Carnegie Mellon University, USA
erin.a.walker@asu.edu, nikol.rummel@rub.de, walker.sean.m@gmail.com,
koedinger@cmu.edu

Abstract. Intelligent tutoring techniques can successfully improve student learning from collaborative activities, but little is known about why and under what contexts this support is effective. We have developed an intelligent tutor to improve the help that peer tutors give by encouraging them to explain tutee errors and provide more conceptual help. In previous work, we have shown that adaptive support from this “tutor” tutor improves student learning more than randomly selected support. In this paper, we examine this result, looking more closely at the feedback students received, and coding it for relevance to the current situation. Surprisingly, we find that the amount of relevant support students receive is not correlated with their learning; however, there is a positive correlation with learning and students noticing relevant support, and a negative correlation with learning and students ignoring relevant support. Designers of adaptive collaborative learning systems should focus not only on making support relevant, but also engaging.

Keywords: intelligent tutoring, computer-supported collaborative learning, adaptive collaborative learning systems, peer tutoring

1 Introduction

Intelligent tutoring systems (ITSs) successfully improve domain learning by tracking problem-solving progress, providing tailored help and feedback, and selecting problems that target misconceptions [1]. However, many of the successful ITSs have been in domains that have well-defined rules such as math and physics (e.g., [2]). Early ITSs were criticized for over-constraining student problem-solving, overemphasizing shallow procedural knowledge, and thus not properly addressing higher-order skills like collaboration, critical thinking, and creativity. In recent years, several ITSs have been developed in response to these criticisms, focusing on metacognition [3], affective modeling and detection [4], and interpersonal interaction [5]. This new wave of ITSs represents an important step towards personalization at all levels of learning: cognitive, metacognitive, motivational, and social [6].

We contribute to this effort by improving the abilities of ITSs for providing adaptive support to collaborative learning. Students benefit from group work, but only

when they exhibit productive behaviors [7]. In theory, adaptive collaborative learning support (ACLS) would be an improvement over nonadaptive forms of support for collaboration, which overstructure the activity for some students while providing insufficient support for others [8, 9]. Indeed, early empirical results suggest that ACLS is an improvement over fixed support and no support at all [10, 11]. However, it is not yet clear why and when ACLS is effective at improving learning. Our work takes a step towards understanding the conditions under which ACLS is effective.

In [12], we proposed two hypotheses for why adaptive support may be effective: 1) Students benefit from receiving relevant support that they can apply to their interaction; and 2) Students who believe support is adaptive feel more accountable for their collaborative actions. To test these hypotheses, we developed an intelligent tutor that assists peer tutors in giving more correct help and higher quality help. In a controlled study, using pre-post measures of learning and surveys of student perceptions, we found evidence that it is the actual adaptivity of support that matters, rather than whether students perceive support as adaptive. However, our conclusions were limited because our analysis did not include process data.

This paper examines why ACLS is effective by looking directly at the *relevance* of each feedback message peer tutors received from the computer, and at the way peer tutors reacted to each message. There have been several ACLS systems that have not been tested in a classroom, but have been evaluated by verifying the validity of the collaborative model used [13], or the applicability of the feedback given [14]. The construction and evaluation of these systems rest upon two hypotheses: Adaptive support systems increase the amount of relevant support given to collaborating students (*H1*), and the more relevant support students receive, the more they will learn (*H2*). Further, research on individual learning from ITSs suggests that it's important that students pay attention to support at the right moments [4]. One reason why relevant support on its own may not be effective is if students fail to notice and engage with the support. Thus, we also examine the relationship between peer tutors' *noticing* of feedback given by the ITS and their domain learning, by including opportunities in the interface for peer tutors to rate support. We hypothesize that peer tutors who notice more relevant support will learn more (*H3*; see Table 1).

2 The Adaptive Peer Tutoring Assistant

Our system builds on the Cognitive Tutor Algebra (CTA), a successful intelligent tutoring system for high school mathematics [2], and allows students to tutor each other using the same interface. The Adaptive Peer Tutoring Assistant (*APTA*) is modeled after traditional novice peer tutoring scripts, where one student tutors another student of the same ability. These scenarios have been successful in classroom environments [e.g., 15], primarily because students of all abilities benefit from giving help [16]; peer tutors engage in *reflective* processes, where they reflect on their partners' errors and notice their own misconceptions, and *elaborative* processes,

Table 1. Hypotheses investigated in this paper.

Name	Description
<i>H1</i>	An adaptive system increases the relevant support collaborating students receive.
<i>H2</i>	The more students receive relevant support, the more they will learn.
<i>H3</i>	The more students notice relevant support, the more they will learn.

where they build on their knowledge as they construct explanations [17]. *APTA* encourages peer tutors to engage in these processes, focusing on three skills:

Skill 1: Necessary help. Peer tutors respond to tutee errors and requests for help. This skill leads peer tutors to reflect on the errors and requests.

Skill 2: Targeted help. Peer tutors ask tutees to self-explain and directly address tutee misconceptions in dialogue. Again, this skill leads peer tutors to reflect on misconceptions.

Skill 3: Conceptual help. Peer tutors give conceptual help, prompting them to engage in elaborative behaviors as they construct explanations.

Because help-giving is an important component of many collaborative scenarios [7], we believe that testing our hypotheses within *APTA* will generalize to other *ACLS*.

In the learning environment, students are given a problem like “Solve for y,” for an equation like “ $ay + by + m = n$ ”. They are grouped into pairs and are seated at different computers at opposite sides of the same classroom. For the remainder of this paper, we refer to the student acting as the tutor in the learning activity as the *peer tutor*, and the student being tutored as the *tutee*. Tutees solve the problem using menus, selecting options like “Subtract from both sides” and typing in a term like m . For some problems, the computer performs the operation; for other, more advanced problems, the student must type in the result of the operation themselves. Peer tutors

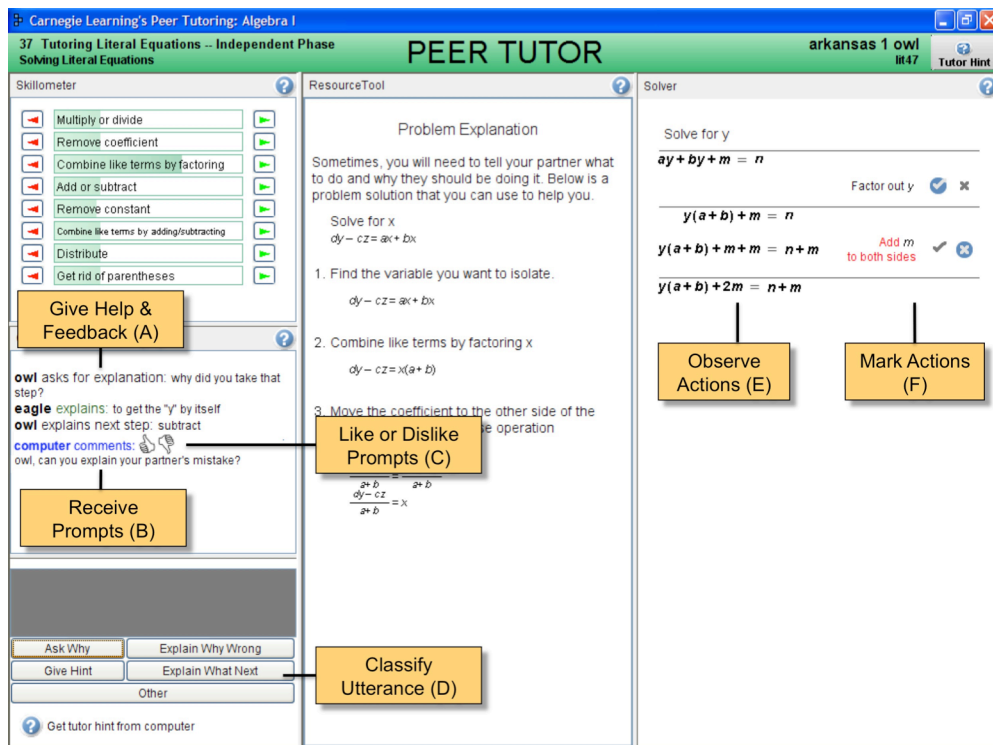


Figure 1. Peer tutor’s interface in *APTA*. The peer tutor watches the tutee take problem-solving actions (E), and marks the actions right or wrong (F). Students can talk in the chat window (A), where they receive prompts from the computer (B), and can choose to like them, dislike them, or ignore them (C).

can see their peer tutee's actions, but cannot solve the problem themselves (see *E* in Figure 1). Instead, they mark the peer tutee's actions right or wrong (*F* in Figure 1), and receive feedback from the cognitive tutor on whether their marks are correct (described more in [12]). Peer tutors can also interact with tutees in a chat tool, where they give help and feedback (*A* in Figure 1). We augmented the chat tool with sentence classifiers (*D* in Figure 1), asking peer tutors to label their utterances prior to submitting them. Encouraging students to use sentence classifiers correctly was an additional system goal (*Skill 4: Use of Classifiers*).

APTA supports peer tutors in giving better help using reflective prompts visible to both students in the chat window (*B* in Figure 1). For example, after peer tutor instrumental help like "subtract x ", the computer might say "[Tutor], why do you say that? Can you explain more?" The reflective prompts were adaptive in terms of content and timing, based on knowledge tracing of the four skills described above (necessary help, targeted help, conceptual help, and use of classifiers). In response to each relevant peer tutor or tutee action, *APTA* followed a four-step process. First, in Step 1, the problem state was assessed based on the tutee problem-solving action, a machine classification of the peer tutor chat, and a self-classification of the tutor chat. Next, in Step 2, *APTA* used a 20-rule production model to classify the action as effective, somewhat effective, somewhat ineffective, or ineffective, as it related to each of the skills relevant to the particular action. The system assessments of each relevant skill were adjusted using Bayesian Knowledge Tracing. In Step 3, all skills whose assessments had been adjusted based on the previous action were compared to pre-defined thresholds related to the rules that had been fired, to determine if feedback should be given based on the skill. Each threshold had a priority, and the activated threshold with the highest priority was selected as a candidate for feedback. Finally, in Step 4, a feedback message was selected randomly from all possible messages associated with a given threshold. Table 2 displays sample positive and negative feedback related to each skill. We did not give positive feedback for use of classifiers because we considered it to be more distracting than valuable.

Table 2. Positive and negative feedback messages for the four skills traced by *APTA*. Positive feedback was given in response to firing of effective or somewhat effective rules, while negative feedback was given in response to firing of ineffective or somewhat ineffective rules.

Skill	Positive Feedback	Negative Feedback
Necessary help	Keep at it! When your partner asks for help, it's a good chance to explain how to solve the problem.	[Tutor], if you don't know how to help your partner ask the computer for a hint.
Targeted help	Good work! Remember, exploring what your partner is doing wrong can help them not make the same mistake on future problems.	[Tutor], can you explain your partner's mistake?
Conceptual help	Keep it up! Talking about concepts behind the problems can help you to understand them better.	[Tutor], when you explain a step to your partner tell them why they should be doing the step.
Use of classifiers	None	[Tutor], think about whether "ask why", "explain why wrong", "hint", or "explain next step" best describes what you last said.

3 Method

In the study discussed in this paper, described more fully in [12], we compared three conditions. In the *real adaptive* condition, students received adaptive support and were told it was adaptive. In the *real nonadaptive* condition, students received nonadaptive support and were told it was nonadaptive. In the *told adaptive* condition, students received nonadaptive support but were told it was adaptive. As we noticed from previous studies that much nonadaptive support was still plausible feedback that could be applied to the interaction context, the inclusion of the *told adaptive* condition was, in part, to evaluate if students who believed support was adaptive would benefit from nonadaptive support that they received. If students thought the system was adapting to their behaviors, they may be more likely to attend to the support and apply it to their interaction. The *real adaptive* condition used APTA, as described above, while the two nonadaptive conditions received prompts selected as follows. Every time students would have received a reflective prompt were they in the *real adaptive* condition, they did not receive a prompt in the nonadaptive conditions. However, they received a prompt within the next three turns, thus *yoking* the nonadaptive prompt to the adaptive prompt. We randomly selected the content of the nonadaptive prompt, with one exception: we did not choose content related to the skill addressed in the yoked adaptive prompt. Nevertheless, there were many situations where the randomly selected prompt could be perceived as relevant.

Participants were 130 high school students (49 males, 81 females) from one high school, currently enrolled in Algebra 1, Geometry, or Algebra 2. The study was run at the high school, either immediately after school or on Saturdays. Students participated in sessions with up to eight other students (M group size = 7.41, SD = 1.35). Each session was randomly assigned to condition, and then within each pair, students were randomly assigned to the role of tutor or tutee. For the most part, students came with partners they had chosen. For ease of scheduling, we sometimes assigned an extra student to a given session, and 8 students worked alone. 1 dyad was excluded due to a logging error with the computer prompts. Thus, 120 students participated in the collaborative activity. Since our goal was to improve the help that peer tutors give, our discussion in this paper focuses on the 60 students who were assigned the role of peer tutor. An analysis of tutee learning is presented in [12].

Students first took a 20-minute domain pretest, and then spent 20 minutes working individually using the CTA to prepare for tutoring. They were then assigned either the tutor or tutee role. Students spent 60 minutes in a tutoring phase, with one student tutoring another student. Finally, students took a 20 minute domain posttest. Pre- and posttests were counterbalanced, and assessed knowledge of literal equation solving.

We used process data from the study to measure two variables: relevance of computer support and peer tutor noticing of support. First, we coded each instance of support delivered by the computer tutor for whether it was relevant to the current context, as defined by the tutee-tutor interactions spanning the last instance of tutee dialogue, tutor dialogue, and tutee problem step. To be relevant, negative feedback had to meet three criteria:

1. *Not contradict the current situation.* E.g., feedback that referred to an error contradicts the situation if tutees had not made an error.

2. *Refer to something students were not currently doing.* E.g., feedback that prompted for more conceptual help would only be relevant if students were not giving conceptual help.
3. *If students were to follow the help, their interaction would be improved, based on the four skills.* E.g., feedback that tells the tutor to give help would improve the interaction if the tutee had asked for help and not received it.

For positive feedback to be relevant, students had to be doing something to merit positive feedback, and then the advice given by the feedback had to meet the above criteria #1 and #3. To calculate interrater reliability, two raters independently coded 30% of the data, with a kappa of 0.70. Conflicts were resolved through discussion.

The second construct, peer tutor noticing of support, came from an interface feature we added to allow students to give us feedback on the computer prompts. As each prompt was given in the chat window, students could choose to rate the feedback (by clicking thumbs up or thumbs down, see C in Figure 1), or ignore it completely. Students were told that this action would help us determine which feedback was useful. We coded students as noticing the feedback if they rated the feedback, suggesting that they had read and reflected on the feedback. Not rating the feedback gave us no information on their response. We further discuss the implications of this measure in the discussion.

4 Results

For the purposes of this paper we focus on an analysis of how relevant feedback and noticing feedback influenced peer tutor learning. As reported in [12], we conducted a one-way ANCOVA to examine the effects of condition on peer tutor learning, with posttest score as the dependent measure, condition as a between subjects variable, and pretest score as a covariate (see Table 3). Condition had a significant effect on posttest score ($F[2,56] = 4.10, p = 0.022$), and pretest was also significantly predictive of posttest score ($F[1,56] = 31.49, p < 0.001$). We found that providing real adaptive support led peer tutors to learn more. According to an ANOVA, total feedback did not differ between the three conditions ($F[2,57] = 0.591; p = 0.557$; see Table 3), suggesting that the nature of the feedback led to the improvement.

We first examined *H1*, to verify using the process data that the implementation of the adaptive support condition indeed had the intended effect, in that the amount of relevant feedback differed between adaptive and nonadaptive conditions (see Table 3 for means). We conducted a linear regression with relevance as the dependent variable. We included two dummy coded condition variables in the regression, one representing the told adaptive condition and one representing the real nonadaptive condition. We also controlled for total feedback given by the computer, adding it as a predictor variable, and including the two interaction terms between each dummy coded condition variable and total feedback. Because we included interaction terms, we centered the total feedback variable by subtracting the mean. We found that the model that included the interaction terms was a better fit for the data ($F \text{ Change } [2,54] = 20.62, p < 0.001$). The results of the regression are presented in Table 4. The model was significant ($R^2 = 0.902, F[5,54] = 99.95, p < 0.001$). All variables entered

Table 3. Mean pretest scores, posttest scores, and amounts of total feedback given by the computer, relevant feedback given by the computer, and attended feedback given by the computer. Standard deviations are in parentheses.

Condition	Pretest	Posttest	Total Feedback	Relevant Feedback	Noticed Feedback
Real Adaptive	0.27 (0.17)	0.39 (0.18)	15.53 (11.28)	12.84 (10.83)	7.16 (6.56)
Told Adaptive	0.24 (0.12)	0.27 (0.14)	17.50 (9.37)	7.45 (5.50)	4.73 (5.91)
Real Nonadaptive	0.30 (0.15)	0.29 (0.18)	14.26 (8.00)	5.68 (4.44)	4.21 (4.12)

Table 4. Regression results comparing the effects of condition and total feedback on relevant feedback given by the computer.

Variable	β	$t(55)$	p
Told Adaptive	-0.402	-8.04	<0.001
Real Nonadaptive	-0.400	-7.97	<0.001
Total Feedback	1.149	17.62	<0.001
Total Feedback *Told Adaptive	-0.327	-5.66	<0.001
Total Feedback *Real Nonadaptive	-0.263	-4.98	<0.001

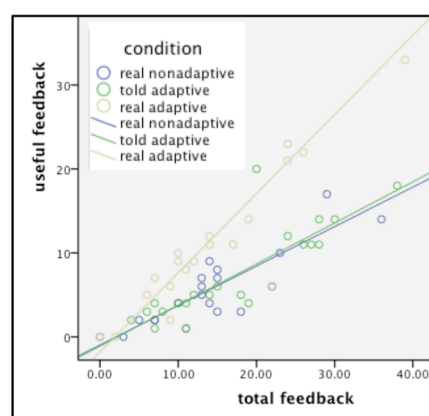


Figure 2. Graph representing the interaction between total feedback given by the computer, useful feedback, and condition.

were significant in the model. When all else is held constant, the real adaptive condition was responsible for significantly more instances of relevant feedback (76%) than the told adaptive (41%) and real nonadaptive conditions (40%). The interaction terms show that the more total instances of feedback, the greater the difference between the real adaptive condition and other conditions (see Figure 2).

We then examined *H2*, looking at whether total relevant feedback was related to learning. We conducted a linear regression, with posttest as the dependent variable, and told adaptive, real nonadaptive, pretest, and relevant feedback as predictor variables ($R^2 = 0.44$; $F(4,55) = 10.97$; $p < 0.001$). While as before condition and pretest were significantly predictive of learning, the total amount of relevant feedback was not ($\beta = -0.180$, $t(54) = -1.65$, $p = 0.104$). Despite the real adaptive condition containing more relevant help, this alone did not explain learning gains found.

Next, we looked at *H3*, examining whether the amount of relevant support students rated affected their learning. We had divided support into two categories:

support that peer tutors noticed (by pressing the like or dislike button), and support that peer tutors ignored. Given that we had also coded support for whether it was relevant or irrelevant, we then had four categories: noticed relevant support, ignored relevant support, noticed irrelevant support, and ignored irrelevant support (see Table 5 for means). We conducted a linear regression, with posttest as the dependent variable, and several predictor variables: pretest, noticed relevant feedback, ignored relevant feedback, noticed irrelevant feedback, and ignored irrelevant feedback. The overall model was significant ($R^2 = 0.512$, $F[5,54] = 11.32$, $p < 0.001$). Noticing relevant feedback was significantly positively related to learning, while ignoring relevant feedback was significantly negatively related to learning (see Table 6). On the other hand, student interactions with irrelevant feedback did not relate to learning.

Because noticing or ignoring relevant feedback related to learning, we explored how those variables differed between conditions. We conducted a MANCOVA with noticed relevant and ignored relevant feedback as dependent variables, and condition and total feedback as predictor variables. Condition significantly affected the amount of noticed relevant feedback ($F[2,56] = 7.10$, $p = 0.002$) and ignored relevant feedback ($F[2,56] = 3.46$, $p = 0.038$). This relationship was strongest for the noticed relevant variable, where post-hoc pairwise comparisons revealed that the real adaptive condition was significantly different from both the real nonadaptive condition ($p = 0.009$) and the told nonadaptive condition ($p = 0.003$). For ignored relevant feedback, the adaptive condition was marginally different from the told adaptive condition ($p = 0.06$) and not significantly different from the real nonadaptive condition ($p = 0.105$).

As noticing feedback played a role in tutor learning, we further examined whether students noticed different amounts of feedback across conditions. A one-way ANCOVA with noticed feedback as the dependent variable, condition as the independent variable, and controlling for total feedback, revealed that students did not notice different amounts of feedback across conditions ($F[2,56] = 1.78$, $p = 0.178$; means of noticed feedback are in Table 3). Students noticed similar numbers of feedback across conditions, but because there was more relevant feedback in the real adaptive condition, students noticed more relevant feedback in that condition.

Table 5. Means of variables relating to attention and relevant feedback. Standard deviations are in parentheses.

Condition	Noticed Relevant	Ignored Relevant	Noticed Irrelevant	Ignored Irrelevant
Real Adaptive	5.63 (5.33)	7.21 (9.02)	1.53 (1.98)	1.16 (1.54)
Told Adaptive	2.05 (3.65)	5.41 (4.53)	2.68 (3.31)	7.36 (5.18)
Real Nonadaptive	1.79 (2.10)	3.89 (4.46)	2.42 (2.48)	6.16 (5.48)

Table 6. Regression results for the effects of relevant and attended feedback on posttest score.

Variable	β	$t(55)$	p
Pretest	0.550	5.66	<0.001
# Noticed Relevant	0.324	2.81	0.007
# Ignored Relevant	-0.279	-2.64	0.011
# Noticed Irrelevant	-0.088	-0.84	0.407
# Ignored Irrelevant	-0.070	-0.61	0.543

5 Discussion and Conclusions

In this paper, we examined when adaptive collaboration support might be effective. We discovered that our adaptive system indeed provided students with more relevant support than a nonadaptive system, and this difference became more apparent the more feedback students received. However, relevant support alone was not related to student learning. Instead, students had to notice relevant support in order to benefit from the support. Students noticed support at similar rates across all three conditions, but because there was more relevant support in the adaptive condition, students noticed more relevant support when the system was adaptive.

Our results depend heavily on our measure of relevance and our measure of noticing. The coding scheme we developed for feedback relevance took several iterations, and we found that many feedback messages could be interpreted as relevant in several different situations. The nonadaptive conditions had relatively high incidences of relevant help, even though we tried to select messages that were not relevant. It is possible that a carefully designed nonadaptive system may be able to mimic the performance of an adaptive support system. Our second measure tracked whether students liked or disliked particular feedback messages as an indication of whether students noticed feedback. This measure of noticing implies that students had read the feedback, and had potentially reflected on how it related to their interaction. This method has limitations; if students did not respond to a feedback message, it is impossible to be certain that they did not notice it. However, as a rough measure, it provided insight on how students reacted. Including these types of measures in other ITSs may provide useful online information on how students react to support.

One interpretation of the results is causal: The adaptive system led students to notice more relevant support, and students who noticed relevant support learned more. This interpretation might explain why students in the real adaptive condition learned the most. However, the adaptive system also caused students to *ignore* more relevant support (albeit to a lesser degree) and students who ignored more relevant support learned less. It is possible that students who ignored relevant support were struggling the most with the learning activity, and also learning less because of their difficulties. While we are limited in our ability to draw causal conclusions from this analysis, we do know that the amount of relevant support played a factor in student learning; noticing relevant support related to learning, while noticing irrelevant support did not. Encouraging students to notice more support, while continuing to work on making support more relevant, may be one key to maximizing the benefits of ACLS.

Thus, the next step in this work will be to examine why students notice support, and determine how to encourage more students to attend to and reflect on support. It is likely that individual differences affect the degree to which students notice help (although noticing relevant support was not correlated with pretest score). A promising approach might be to use data mining techniques to improve the design of feedback messages, improving student likelihood of noticing those messages. The timing of messages might have an influence: In initial exploration, we found that feedback messages that appeared when peer tutors were struggling and distracted were more likely to be ignored. Content might also have an influence: Feedback messages that were specific and easily implemented appeared to be more engaging.

Our work makes a contribution to the study of ACLS by showing that producing more relevant support alone is not sufficient for improving learning. Students who benefit from relevant support must notice the support. While this finding is intuitive, and has been discussed in individual learning, it had not previously been demonstrated in learning from collaborative systems or discussed in the design of ACLS. ACLS systems are often designed and evaluated with the ultimate goal of creating more relevant support. Future designers of such systems will have to explore how to improve student noticing of support in order to have a significant impact.

Acknowledgments. This work was supported by the Pittsburgh Science of Learning Center, NSF Grant #SBE-0836012, and a Computing Innovations Fellowship, NSF Grant #1019343. Thanks to Ruth Wylie and Kasia Muldner for their comments.

References

1. VanLehn, K. (2006). The behavior of tutoring systems. *IJAIED*, 16 (3), 227-265.
2. Koedinger, K., Anderson, J., Hadley, W., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43.
3. Muldner, K., Burleson, B., VanLehn, K. (2010). "Yes!": using tutor and sensor data to predict moments of delight during instructional activities. In *Proc. UMAP*, 159-170.
4. Roll, I., Aleven, V., McLaren, B.M., & Koedinger, K.R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, 21, 267-280.
5. Ogan, A., Aleven, V., Kim, J., & Jones, C. (2011). Persistent Effects of Social Instructional Dialog in a Virtual Learning Environment. In *Proc. 15th International Conference on AIED*, pp.238-246.
6. du Boulay, B., Avramides, K., Luckin, R., Martinez-Miron, E., Rebolledo-Mendez, G., & Carr, A. (2010). Towards Systems That Care: A Conceptual Framework based on Motivation, Metacognition and Affect. *International Journal of Artificial Intelligence in Education*, 20(3), 197-229.
7. Johnson, D. W. & Johnson, R. T. (1990). Cooperative learning and achievement. In S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 23-37). NY: Praeger.
8. Dillenbourg, P. (2002). Over-scripting CSCL: The risk of blending collaborative learning with instructional design. In Kirschner, P. A. (Ed.), *Three worlds of CSCL: Can we support CSCL?* 61-91.
9. Kollar, I., Fischer, F., & Slotta, J. D. (2005). Internal and external collaboration scripts in web-based science learning at schools. In T. Koschmann, D. Suthers, & T.-W. Chan (Eds.), *The next 10 years! Proc. CSCL 2005* (pp. 331-340). Mahwah, NJ: Lawrence Erlbaum Associates.
10. Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting Collaborative Learning and Problem Solving in a Constraint-based CSCL Environment for UML Class Diagrams. *International Journal of Computer-Supported Collaborative Learning*, 2 (2-3), 159-190.
11. Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. In *Proc. AIED 2007* (pp. 383-390). IOS Press.
12. Walker, E., Rummel, N., & Koedinger, K. R. (2011). Adaptive support for CSCL: Is it feedback relevance or increased accountability that matters? In *Proc. CSCL 2011* (pp. 334-342).
13. Suebnukarn S and Haddawy P. (2006). Modeling Individual and Collaborative Problem-Solving in Medical Problem-Based Learning. *User Modeling and User-Adapted Interaction*, 16(3-4):211-248.
14. Constantino-González, M. A., Suthers, D., & Escamilla de los Santos, J. (2003). Coaching web-based collaborative learning based on problem solution differences and participation. *IJAIED*, 13, 263-299.
15. Fantuzzo, J. W., Riggio, R. E., Connelly, S., & Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology*, 81(2), 173-177.
16. Ploetzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. In P. Dillenbourg (Ed.), *Collaborative Learning: Cognitive and Computational Approaches* (pp. 103 – 121). Elsevier Science Publishers.
17. Roscoe, R. D. & Chi, M. (2007) Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*. 77(4), 534-574.