Casual Collaborations while Learning Mathematics with an Intelligent Tutoring System

Ivon Arroyo[1], Beverly P. Woolf[1] , David Shanabrook[1]

[1] Department of Computer Science, University of Massachusetts Amherst
{ivon, bev, shanabrook}@cs.umass.edu

Abstract. We have started exploring and facilitating natural collaborations between students sitting in contiguous computers. We implemented a mechanism to facilitate this kind of social interactions, instead of struggling against it. By allowing students to "force" specific problems they want, one student can pull up the math activity that another student is going through in their computer, thus capitalizing on this social interaction for a shared problem-solving experience. In a research study, collaboration between friends foster greater development of scientific reasoning when compared with collaboration between mere acquaintances. As part of this research article, we will present pilot data reporting on engagement levels and other "within the system" indicators of learning and progress, as students go through mathematics activities where pairs were collaborating, and compare these within-tutor outcomes to other students that did not have this kind of emergent behavior..

Keywords: informal collaborations, mathematics learning, tutoring systems.

1. Introduction

Tutoring systems have demonstrated effective learning over large amounts of students in classrooms in public schools [1][2], and some studies have shown evidence that the adaptive nature of tutoring systems is responsible for higher learning rates. However, even the most effective tutoring system will fail if the student's behavior is not receptive to the material being presented. For example, disengagement has been shown empirically to correlate with a decrease in learning rate. This is related to students often using tutors ineffectively, e.g., deliberately entering incorrect answers to elicit hints and, eventually, the correct answer from the tutor.

At the same time, we have noted for some time that when students work in the computer lab with tutoring software, natural collaborations emerge with neighboring classmates. This is partly to the fact that, when students arrive at the computer lab, they generally sit next to their friends and people who they feel comfortable to be around. Consequently, and even though the tutoring system in question might have been initially conceived as a "private" one-on-one learning environments, working in some way with a neighboring student is a behavior that naturally emerges in a classroom.

The research presented here starts to explore those collaborations, and promotes them via an interface feature to facilitate them. We describe first the background research on collaboration, the testbed tutoring system. Then, a mechanism for the detection of such informal collaborations, and a way to measure their effectiveness in terms of supporting productive engaged behaviors with the math tutoring activities.

1.1 Background Research

Collaboration has its own set of unique benefits for education. Nearly 700 studies clearly indicate that collaboration results in higher achievement and greater productivity; more caring, supportive, and committed relationships; and greater psychological health, social competence, and self-esteem for [3][4]. In competitive environments (the opposite of collaborative environments), students perceive that they can obtain their goals if other students fail to obtain their goals. Other systems promote neither: they might focus on personal progress, ignoring others.

Collaboration Strategies in classrooms has proven to be successful and powerful for learning. Collaborative student discourse (i.e., reflective discussions among students about content) promotes deep and meaningful learning and enhances students' reasoning skills [3]. It often results in learning that outperforms the ability of the best individuals in the group and emphasizes positive interdependence (of task, identity, role and goals), individual and group accountability (challenging and motivating every student), and authentic interaction (brainstorming, planning, social and team building skills, and solidarity).

The computer supported collaborative learning (CSCL) community has developed several successful computational environments to support collaboration. In some cases, persistence of the system helps support teams, reason about activities, and intervene. The nature of the computational support varies along a continuum from a persistent discussion forum (typed and threaded nodes) to supporting students to process ideas in shared discussion. Environments organize and manage teamwork and collaboration artifacts [10][11]. They help students to explain their reasoning to others and facilitate the receipt of feedback.

1.1.1 Collaborative Learning on the Learning of Mathematics

We are interested in collaboration within our working domain, K-12 mathematics. A case study of online collaboration while solving algebra problems adapted the methodology of conversation analysis to text-based chat room technology [10]. The analysis investigated math problem-solving communication and the circumstances of online chat. A primary goal was to identify technological barriers caused by standard chat technology with an eye to designing a more appropriate and supportive online collaborative learning environment. Cognitive processes were identified as resulting from the dynamic interaction within contexts of group discourse and collaboration.

One large study found that collaborative learning produced the largest effect size for learning mathematics [9]. Eight studies, four of which were randomized experiments or randomized quasi-experiments, found strong impacts (median ES=+0.30) of cooperative learning programs, all of these studies not involving technology. Research on the achievement outcomes of mathematics programs for middle and high schools was reviewed and effect sizes were generally very small (median ES=+0.07 in 38 studies) for mathematics curricula, larger (median ES=+0.16 in 39 studies) for computer-assisted instruction and larger for instructional process programs (median ES=+0.21 in 23 studies). This study concluded that programs that affect daily teaching practices and student interactions have larger impacts on achievement measures than do those that emphasize textbooks or technology alone. These suggest that focus is needed on how the classroom is organized to maximize student engagement and motivation, and programs that fundamentally change what students do every day in their math classes.

1.1.2 The testbed: Wayang Outpost

Wayang Outpost is an Intelligent Electronic Math Tutoring Software that incorporates multimedia, animated characters and animated adventures to assist middle and high-school students practice solving math word problems, with problems in the format of standardized math tests such as the SAT and standardized tests such as MCAS and CA-Star. Currently funded by research programs at the National Science Foundation and Department of Education, Wayang Outpost is state-of-the-art multimedia technology designed to increase standardized test scores and help teachers in their assessment of students' strengths/weaknesses.

As the student progresses though the curriculum, Wayang Outpost learns about how much the student knows, and how much the whole class knows, regarding each math skill. Wayang Outpost has shown to improve standardized test scores. Rigorous controlled research studies showed that student scores improved an average of 10% in tests after only 3 hours of instruction and 20% after 4-5 hours using our tutoring system to practice on problems from the MA Standardized Test. Wayang was originally designed for 10th grade mathematics, but it has been extended to cover the lower grades (6-8th grades), plus extended upwards to prepare students for college placement tests (SAT-Math and Accuplacer)

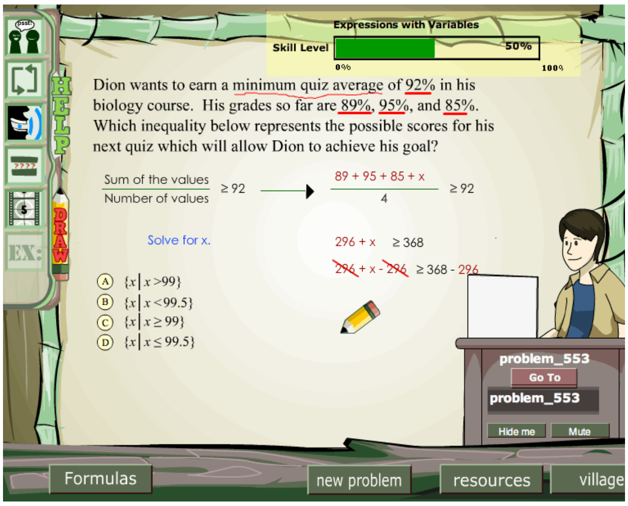

Figure 1. The tutoring module in Wayang. "Jake" is an animated study partner, who emphasizes the importance of effort and perseverance. Hints are available from the "Help" bar on the right. Toolbar on the left allows to make notes and scribble via pen mice. Formulas are available on demand.

and further on, developmental math classes in colleges. In addition, Wayang has the ability to go down to grade 4 math, whenever necessary, as it includes basic math problems at that level, providing a seamless continuum of mathematics coverage that should not be restricted by grade, but by the possibilities of each student. This is possible thanks to its Adaptive Problem Selection algorithm described in [2] which models individual problem difficulty as well as students' effort, mastery and emotions.

Research projects have addressed the question of how to best customize teaching to each student, and how to learn and improve the systems' teaching from vast amounts of information from previous students [2], while addressing emotional factors such as frustration and anxiety while problem-solving. One of those had to do with experimenting various ways to address specific emotions as they arise, such as variations in the dialog that characters have with students when feelings emerge that may help to cope with frustration and anxiety. Characters can empathize with the student, get excited when they do well, especially if they have demonstrated effort to work through problems.

The system provides a variety of help aids to students who are practicing how to solve standardized math problems. These consist not only of help aids, but of videos and animated worked-out examples. It has been used with thousands of students from middle and high schools and after-school programs in urban and rural Massachusetts areas. Like a human tutor, Wayang identifies student skills and modifies its presentation to provide further work on weak skills, see http://wayangoutpost.com. The software has an adaptive mechanism that tailors the sequencing of problems, using cutting-edge research in educational data mining, multimedia learning, embedded assessment, feedback, and scaffolded learning.

It identifies the most critical cognitive skills, alters the choice of the next problem, provides individualized responses and predicts the likelihood of success on future problems [2]. Mathematics content is organized in a continuum across grades, so that students who are not ready for the mathematics corresponding to their grade level can fall back into earlier mathematics topics; in a similar way, students who are beyond their grade level and ready for high school-level math can progress ahead.

One important feature recently added was the "Go To" button, shown at the bottom right corner of Figure 1. This button was created to ease the natural interaction and collaboration of students sitting in contiguous computers. Some students were found to synchronize their answers and solve together, so that the behavior of the system tended to be the same for both, as much as they could, or the system would allow. The "Go To" button was created so that students might type in a specific problem number, so that a specofic problem would show up. It was told students that this way they could "play" the same problems that their neighboring classmate had. This was a choice for students and was never forced. The purpose of the Go To button was to start to explore the extent in which students collaborated, and the possibility of developing more sophisticated tools to allow contiguous students to work together.

2. Hypotheses and Methodology

While the synchronization of problems was totally optional, we can already start to preliminary analyze how students work together. The main hypothesis for this preliminary work was to analyze if students working together had more positive and productive behaviors that have before been found to correlate to learning and others [12]. Productive behaviors on a problem are indication of high levels of engagement, instead of disengagement and gaming.

In order to analyze these collaborations, we needed to carry out several steps. **First**, understand the state of engagement of a student in a problem interaction; **second**, understand which students were actually collaborating in simultaneous problems in these casual interactions; and **third**, analyze the effectiveness of such collaborations in terms of productive behaviors that are a sign of engagement with the software.

2.1 Engagement States

During a **Data Pre-processing Phase**, the raw tutor data is classified in a set of of “student-activity” interactions, or “student-problem engagement states”, see Table 1. A variety of typical behavioral states are determined (e.g., student is *not reading the problem* --NOTR, or *solving a problem after hints are seen* --SHINT). These states are mutually exclusive, so that each student-problem interaction is classified as a single state. For example, an original problem metric, timeToFirstAttempt was used as a continuous predictor. But it produced a highly skewed distribution with a large number of values less than five seconds; a length of time insufficient to even read the problem [12]. While we don’t have a clear understanding the implications of timeToFirstAttempt in a normal range, i.e., the significance of a student taking 20 seconds to the first action versus 5 seconds --student attempted to answer before reading the problem. We categorize this as not reading (NOTR) problem state without regard for the other metrics.

SOF categorizes all problems that are solved on first attempt without invoking help; this indicates content level is too low. BOTT implies gaming and the intervention could prevent this by not providing bottom out solution. GIVEUP indicates user and problem difficulty would be decreased. ATT indicates a student working on-task at an appropriate level; support provided to ensure continued success. SHINT similar to ATT but no support needed. GUESS indicates the student needs either help or lower problem difficulty. These states are mutually exclusive

State	Description
NOTR	Not enough time to read the problem, before an action is taken.
SOF	Solved on first attempt without help.
BOTT	Getting answer from hints or help. Possibly seeing problem as example.
GIVEUP	Moved on before answering.
ATT	Made 1-2 incorrect attempts and self-corrected, without help.
SHINT	Answered with help, without guessing possible solved on first attempt
GUESS	Guess answer after several attempts

Table 1. Student states that will be analyzed, with possible tutor interventions

checked in this order. All of the students' problem states were encoded for each student problem interaction.

We define NOTR, GIVEUP and GUESS as disengagement student-problem interactions. We also define SOF, ATT and SHINT as productive behaviors. We don't bin BOTT as productive or unproductive, as it could be a sign of searching for the right answer in act of disengagement, but it could also be that the student is playing the question in "example mode", asking the computer to solve it in order to learn.

2.2 Inference of Paired Collaborations

In this section we describe how we start to understand collaborations, even if they are informal, or achieved via untraditional methods. Let *SP* be a matrix of "simultaneous problem activity" of size NxN where N=total number of students in the class. Each element SP_{ij} corresponds to the total number of problem instances that student s_i and student s_j solved simultaneously. "Simultaneously" means that one of the two students started a practice problem activity p when the other student was already on that same activity, so that there was some fraction of time when both students were interacting with the same activity.

This situation could happen mainly used in three possible scenarios: 1) a student s_i who is sitting next to a student s_j deliberately uses the "Go To" problem box to pull up the same activity their classmate had on the screen to either help him/her or work together; 2) a student s_i in the class coincided with student s_j in a math problem activity just by chance, but they are not aware of it; 3) a student s_i sitting close to a student s_j coincide in a math problem activity, notice this, and then start working together and typing in the same answers, thus having a long sequence of same math problems as a result of the adaptive problem selection algorithm.

Each student has a mean *number* of simultaneous problems with other students, which we call the mean overlap with other students . This might differ from student to student depending on the total amount of problems seen, and a variety of other factors. We define O_i as overlap for a student s_i , $i=1...N,$ as the mean number of simultaneous problems with other students, minus the diagonal which corresponds to a student interacting with himself. Equation 1 shows this computation.

$$\hat{O}_i = \frac{\sum_{\substack{j=1 \\ i \neq j}}^{N} SP_{ij}}{N} \qquad \text{Eq. 1}$$

The idea is to use the mean overlap O_i as a baseline to classify pairs that appear to be heavy collaborators from students who may have coincided on a problem by chance. Thus, we define a threshold of simultaneous problems for each student, T_i , and in this case we set that theshold to be $T_i = 2O_i$ to imply that a student-pair (s_i, s_j) has high potential to be collaborating and working a synchronized way when SP_{ij} *is* *greater than twice the average overlap of problems with any other student in the class*. This is obviously a heuristic and can be tightened or losened to determine a tighther threshold, or a loser threshold. We can formalize that by defining a new matrix called *COLL*, where each element encodes which students are high collaborators and which not, in other words, as specified in Equation 2:

$$\begin{aligned} COLL_{ij} &= 1 \quad when \quad SP_{ij} > T_i \\ COLL_{ij} &= -1 \quad when \quad SP_{ij} \leq T_i \end{aligned} \qquad \text{Eq. 2}$$

From now on, we will refer to student pairs with $COLL_{ij}=1$ as "collaborators" and $COLL_{ij}=-1$ as "non-collaborators", where collaborators are pairs of students that have a large amount of same simultaneous problems, as specified by our threshold T_i.

2.3 Analyzing the effectiveness of Collaborations

The last step is to analyze the effectiveness of such paired of students at working in synchronization over a large amount of problems, compared to other students who were not apparently collaborating. As mentioned before, we analyze the effectiveness in terms of productive and unproductive (disengagement behaviors) for high collaborator pairs.

We define ***PB***, were each PB_{ij} corresponds to the total of Productive Behaviors in the interaction of a student with a math practice problem, for those specific problems where the pair (s_i, s_j) overlapped and potentially collaborated, and PB_{ij} was *undefined* if the pair was not a collaborator. We also define ***RPB*** as the matrix of ratios of productive behaviors PB_{ij}/SP_{ij} where each element then represents the likelihood (percent if multiplied by a hundred) of a student in the pair having a productive behavior in relation to that problem (SOF, SHINT, ATT).

We last define ***DB***, were each $\boldsymbol{DB_{ij}}$ corresponds to the total of Disengagement Behaviors in the interaction of a student s_i or s_j with specific math practice problems where the pair (s_i, s_j) overlapped and potentially collaborated, and DB_{ij} was *undefined* if the pair was not a collaborator. We also define ***RDB*** as the matrix of ratios of DB_{ij}/SP_{ij} where each element represents the likelihood of a student in the pair having a disengagement behavior (NOTR, GUESS, GIVEUP).

3. Results

We describe and summarize the results of *RPB* and *DPB* for two math classes, a 7th and an 8th grade public school in Massachusetts. The first class had N=20 students and the other one had N=23. Students were seeing different topics, aligned to their curriculum and had used the tutoring system for at least two 45 minute time periods during recent months.

We computed a Vector $\hat{PBC}_i$ for each student s_i, corresponding to $MEAN_j\,(\boldsymbol{RPB_{ij}})$ or the mean of likelihoods of productive behaviors with a student that s_i collaborated with (it turned out to be that the number of students each s_i collaborated with, as specified in $COLL_{ij}$, was low and about 2-3 students; this was along the lines of what we had seen in the classroom, so it was a good validation of our T_i). Similarly, we computed the mean disengagement likelihood $\hat{DBC}_i$ for each student s_i, corresponding to $MEAN_j\,(\boldsymbol{RDB_{ij}})$ or the mean of likelihoods of productive behaviors for collaborator students. These two vectors indicate the general likelihood and general trend of collaborators to be disengaged, or have productive behaviors, when they interacted with their collaborators on simultaneous problems.

StudId	Mean Difference in Likelihood of Productive Behavior, for Collaborators minus Non-collaborators $\hat{PBC}_i - \hat{PBNC}_i$	Mean Difference in Likelihood of Disengagement Behaviors, for Collaborators minus Non-collaborators $\hat{DBC}_i - \hat{DBNC}_i$
9647	0.07	-0.07
9648		-0.02
9652	0.05	-0.05
9653		-0.02
9654	0.13	-0.12
9655	0.26	-0.26
9656		-0.10
9658	0.01	-0.03
9660	0.02	-0.01
9661		-0.02
9662	0.05	-0.05
9663		0.05
9664	-0.03	-0.03
9665	0.30	-0.12
MEAN	**0.10**	**-0.06**

Table 1. Mean difference in likelihood to have productive learning behaviors or get disengaged, when students interacted on simultaneous problems in a "collaborator" student-pair, vs. non-collaborator pairs who might have coincided on a problem by chance. There is a 10% advantage in productive behaviors for collaborator pairs, and a 6% less likelihood of gaming behaviors for collaborator pairs. Missing cases in the table and missing rows correspond to students who did not have a value associated in the *PBC* or *PBNC* Vectors. This can happen for a variety of reasons, such as lack of simultaneous problems with others

However, in order to make any assessment of how good these general collaboration trends of engagement and disengagement are for pairs of students, we need to find a good comparison. Thus, we created a counterpart "control" vector $\hat{PBNC}_i$ (Productive Behaviors for Non-Collaborator pairs) and another counterpart vector $\hat{DBNC}_i$ (Disengagement Behaviors for Non-Collaborator pairs). Each entry for a student s_i represents the general likelihood of a productive (or disengagement) behavior respectively, but for student-pairs (s_i, s_j) that were NOT considered collaborators, but coincided –most likely by chance-- on a few problems.

The results we present in Table 1 correspond to the differences $\hat{PBC}_i - \hat{PBNC}_i$ and $\hat{DBC}_i - \hat{DBNC}_i$ for students who overlapped on problems with collaborators and with non-collaborators. Thus, the results show the difference of likelihood of good behaviors between collaborators and non-collaborators, both for engagement and disengagement behaviors.

3. Conclusions and Future Work

It has been hard to understand what to do in order to address the fact that students game the system. One possibility is that fostering and promoting collaborations between students, which add social value, can help students to slow down and have more productive and fun experiences while solving math problems. Some collaborative behaviors that may be promoted, instead of normally forbidden due to the one-on-one nature of a tutoring system, are the natural behaviors of working with neighboring students who naturally cluster together. These collaborations are different in that: 1) they are unstructured, and do not use chat facilities to communicate with students in another part of the room, or a different site; 2) they are highly social, possibly involving friends, and may include off-topic conversation.

This article has introduced ways to assess, understand the value of, and consider the possibility of promoting collaborations and synchronizations of neighboring students as they move along with a traditional tutoring software. While it might be easier in the future to ask contextual factors such as whom is sitting next to whom, it is important that just by looking at the logs we can start to understand students who might be working heavily together, and start adjusting practice work and other activities, and even the interface, to the collaborating pair.

Interesting results over one class of 23 students suggested that these casual, informal and natural collaborations that happen in the computer lab work well to promote learning, in the sense that such pairs have higher incidence of engagement behaviors and lower incidence of disengagement behaviors. In past work, we have shown that such "productive behaviors" that include high incidence of time spent on problems and hints are significantly help to predict learning [5][6]. Within the context of collaborative learning, it is important to understand the positive and negative behavioral implications of a pair that collaborates, because it is quite possible that a pair might either benefit learning or help to distract, and it is not clear which one is more prone to happen. The fact that collaborations helped to change student behaviors from gaming ones to productive ones, suggests that inviting students to collaborate would be an extremely valuable way to repair student gaming. So far, there has not been concrete evidence that shows that collaborations can make a difference at a deep level of detailed student behaviors (e.g. some tutoring systems that encouraged collaboration did not fully assess the benefits of collaboration at the level we have proposed).

Future work will regard the exploitation of these paired collaborations to intervene and help students synchronize their work. We will also extending this analysis to other measures of learning, such as measures of Mastery for collaborating pairs. It will also regard the validation of the algorithm that determines who are collaborator pairs with other (external) measures, such as asking the students who they collaborated with at posttest time.

References

[1] Arroyo, I., Beal, C.; Murray, T.; Walles, R.; and Woolf, B. (2004). Web-based Intelligent Multimedia Tutoring for High Stakes Achievement Tests. *In Proceedings of the Seventh ITS conference.* 468–477.

[2] Arroyo, I., Mehranian, H., Woolf, B. (2010b) Effort-based Tutoring: An Empirical Approach to Intelligent Tutoring. Proceedings of the 3rd International Conference on Educational Data Mining. Pittsburgh, PA

[3] Johnson, D. W., & Johnson, R. (1999). Learning together and alone: Cooperative, competitive, and individualistic learning. Boston: Allyn & Bacon. First edition, 1975.

[4] Johnson, D., R. Johnson, J. Johnson, and D. Anderson. (1976) "Effects of Cooperative Versus Individualized Instruction on Student Prosocial Behavior, Attitudes Toward Learning and Achievement." JOURNAL OF EDUCATIONAL PSYCHOLOGY 68: 446-452.

[5] Arroyo, I.; Woolf, B. (2005). Inferring learning and attitudes from a Bayesian Network of log file data Proceedings of the 12th International Conference on Artificial Intelligence in Education. Amsterdam, 2005.

[6] Arroyo, I., Murray, T., Woolf, B. P., Beal, C. R. (2004). Inferring Unobservable Learning Variables from Students Help Seeking Behavior. James C. Lester, Rosa Maria Vicari, Fábio Paraguaçu (Eds.): Intelligent Tutoring Systems, 7th International Conference, ITS 2004, Maceiò, Alagoas, Brazil, Proceedings. Lecture Notes in Computer Science 3220 Springer 2004.

[7] Harrer, Andreas; McLaren, Bruce M.; Walker, Erin; Bollen, Lars; and Sewall, Jonathan, "Collaboration and Cognitive Tutoring: Integration, Empirical Results, and Future Directions" (2005). Human-Computer Interaction Institute. Paper 151. http://repository.cmu.edu/hcii/151

[8] Kumar, R., Gweon, G., Joshi, M., Cui, Y., Rosé, C.P. (2007) Supporting students working together on Math with Social Dialogue, Workshop on Speech and Language Technology in Education, Farmington, Pennsylvania, 2007.

[9] Slavin, R., Lake, C., & Groff, C. (2007). Effective programs in middle and high school mathematics: A best-evidence synthesis. United States Department of Education.

[10] Stahl, G. (2006). Sustaining group cognition in a math chat environment. Research and Practice in Technology Enhanced Learning (RPTEL), 1(2), 85-113.

[11] Soller, A. (2004). Computational Modeling and Analysis of Knowledge Sharing in Collaborative Distance Learning (preprint). User Modeling and User-Adapted Interaction: The Journal of Personalization Research, 14 (4), 351-381

[12] Arroyo, I., and Woolf, B. (2005). Inferring Learning and Attitudes from a Bayesian Network of Log File Data. *In Proceedings of the Twelfth International Conference on Artificial Intelligence in Education*.