Repairing Disengagement With Non-Invasive Interventions

Ivon ARROYO, Kimberly FERGUSON, Jeff JOHNS, Toby DRAGON,
Hasmik MEHERANIAN, Don FISHER, Andrew BARTO, Sridhar MAHADEVAN,
Beverly P. WOOLF
Department of Computer Science & Department of Engineering,
University of Massachusetts Amherst

Abstract. We evaluated the impact of a set of interventions to repair students' disengagement while solving geometry problems in a tutoring system. We present a deep analysis of how a tutor can remediate a student's disengagement and motivation with self-monitoring feedback. The analysis consists of a between-subjects analyses on students learning and on students' attitudes towards mathematics and perceptions of the software, and a deeper analysis on students' engagement within the tutor. Results show that the general trend over time is for students to become more disengaged while using a tutor, yet students can be coaxed into reengagement after viewing interventions that promote self-reflection and self-monitoring --a simple open-learner model accompanied by suggestions and encouragement.

Introduction

Students become disengaged overtime when using tutoring software. One particular externalization of disengagement is "gaming" the system, or moving rapidly through problems without reading them, quickly moving through hints, and seeking the final hint that might give the answer away. It has been estimated that students who game the system learn two thirds of what students who do not game the system learn (Baker, Corbett & Koedinger, 2004). This could be a sign of frustration, something especially important to detect for students with special needs in particular. Another possibility is that this is a sign of poor use of meta-cognitive resources –the ability to regulate their own cognitive resources (Hartman, 2001). In either case, by identifying disengagement behavior and repairing it we move closer towards tutors that generate highly individualized, pedagogically sound and accessible material. This research shall lead towards involving more students, including those that might be unmotivated, or those that are anxious about the taught domain or about computers in general, or that simply are not aware of the impact of their unproductive actions on the learning process.

Developing pedagogical approaches to respond online to students who have become disengaged is a challenging task. Past research showed that providing students with examples (extra materials to help in the solving of the problem at hand) can reduce gaming and improve learning (Baker et al., 2006). One possible reason for the success of this feedback is that the origin of gaming was due to frustration for not having enough knowledge to solve the problem. It is still unclear though if students can modify their unproductive behaviours and learn more without the provision of extra content help, just by having them reflect on their actions, and making them aware of productive behaviours. This kind of scaffolding is called 'meta-cognitive' feedback because it tackles self-regulatory processing while learning (Hartman, 2001). De Jong and van Joolingen (1998) described ways in which learning environments can support students' metacognitive skills; however, it has still not been demonstrated whether Software Learning Environments can help students become better learners by providing such meta-cognitive support. Some attempts to remediate unproductive uses of a tutoring software showed that students became even more frustrated when their gaming was blocked with pop-up windows that encouraged proper behavior whenever specific meta-cognitive bugs were observed (Aleven et al., 2004). Not surprisingly, the benefit in actual learning for that intervention was limited (Roll et al, 2006).

One possible approach to making students better learners is to make them more aware of productive and unproductive behaviours, and about the consequences of their actions on their progress, in a way that does not block their unproductive behaviours. With that idea in mind, we designed and evaluated the impact of pedagogical interventions that invite students to reflect on their progress in-between problems, and understand what would be behaviours that will benefit their learning. This paper evaluates the hypothesis that such non-invasive interventions can change a student's engagement state, reduce gaming, enhance learning, while at the same time generate a more positive perception of the system and of the learning experience. More specifically, two hypotheses were tested as part of this research: i) do in-between-problems interventions (performance charts and tips) affect the level of student engagement? and ii) do interventions impact student learning and feelings towards the tutor and towards the learning experience? How about feelings towards their own self-concept and mathematics? The next sections are our attempt to answer these questions.

1. Methodology and Experiment Design

The tutor. Wayang Outpost is a multimedia tutoring system for geometry (Arroyo et al, 2004). It helps students solve challenging geometry standardized tests problems. Wayang is a web-based tutoring software (a database-backed Java servlet with a Macromedia Flash front-end), facilitating the task of logging, pre/post-testing and data collection in general. Students register to use the system, log on and are directed towards the different modules for pre/post-testing and tutoring. Even though Wayang has an adaptive module to tailor the sequencing of problems depending on students' performance at past problems, for this study, the sequencing of problems in Wayang Outpost was fixed (i.e. the same for all students). This decision was made thinking that, if the interventions had an impact, students' engagement in problems would in turn affect problem-solving behavior and interact with the problem sequencing, and the results of this study. Thus, Wayang provided a fixed sequence of problems, chunking problems of similar skills close to each other, and organizing them from easy to hard overall difficulty.

Instruments. Two mathematics tests of 43 items extracted from SAT and Massachusetts MCAS standardized tests (MCAS) were used for pre and post testing. We examined the standardized test scores from 10th graders who took this exam days after the experiment finished. Post-tutor survey question items were provided, including a student's performance/learning orientation, human-like attributes of tutor and a student's liking of mathematics. Most of these came from a Intervention instrument used by Baker (2006) for studies about gaming the system. Items to measure self-concept in mathematics (Wigfield and Karpathian, 1991) and a 10-question self-efficacy instrument were also part of the survey. Last, we included questions that measured student perceptions of the software and the help that we designed. All items were in a 6-likert-type scale, except for the 2 learning vs. performance orientation items (Mueller&Dweck, 1998).

Experimental design. Eighty eight (88) students from four different classes (10th grade students and some 11th graders) from an urban-area school in Massachusetts used Wayang Outpost for one week. It was 4 time periods for about 2 hours of tutoring (the rest of the time was spent doing pre-testing and post-testing). A second control group (called no-tutor control) consisted of matched classes of students who did not use the tutor at all, but were of the same grade level, equivalent proficiency level, and taught by the same teachers. When students logged on to the software, they were pseudo-randomly assigned to either the experimental (Interventions) or the tutor control group. The latter group used the traditional Wayang –worked on problems with the ability to click on a help button that would provide multimedia hints. The Intervention Group received intervention screens at fixed intervals of 6 problems (i.e., after clicking the 'next problem' button on the sixth problem). Experimental interventions were either i) a performance graph with an accompanying message, similar to Figure 1 (students received a negative graph or a positive graph depending on their recent performance and their past performance) or ii) a tip that suggested a productive learning behavior. The tutor provided two kinds of tips: Tip-read-carefully and Tip-make-guess. Tip-read-carefully encouraged students to slow down, read the problem and hints carefully ("Dear Ivon, We think this will make you improve even more: Read the problem thoroughly. If the problem is just too hard, then ask for a hint. Read the hints CAREFULLY. When a hint introduces something that you didn't know, write it down on paper for the next time you need it"). Tip-make-guess encouraged the student to think about the problem, make a guess and, if the guess was wrong, ask for hints ("Dear Ivon, Think the problem thoroughly and make a guess. If your guess is wrong, no problem, just ask for a hint. If you need more hints, keep clicking on help". Students were addressed by their first name both in the messages accompanying the charts and the tips. Whether a student saw a progress chart or a tip, and which one, was a randomly-made decision.

Procedure. During the first time period, students took an online mathematics pretest. Then they used the tutoring software for part of the first day, second and third days. Posttest was started at the end of the third day and continued during the fourth day. Pre-test and post-tests were completed online (within the software). Pretests and posttest were counterbalanced.

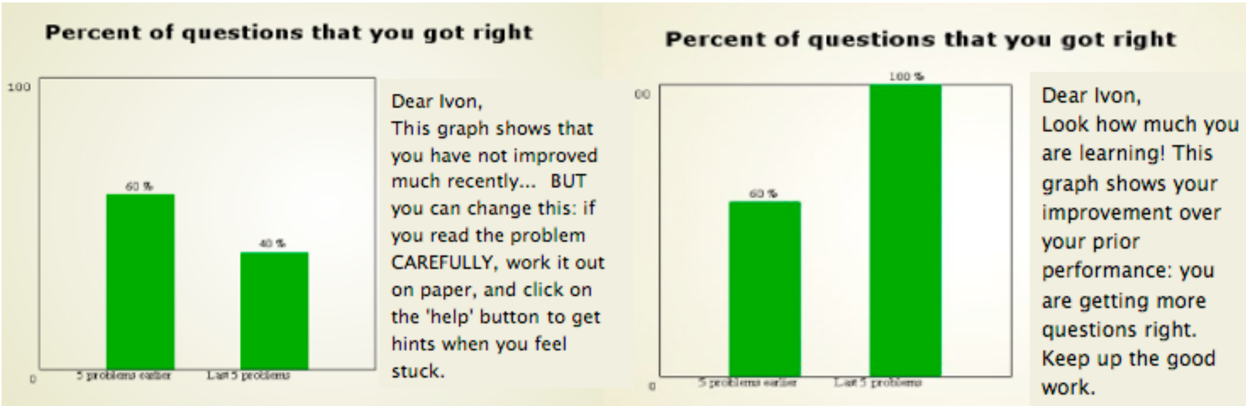

Figure 1. Progress Charts that show students their accuracy of responses from before to recently

2. Results

2.1 Between-subjects analysis of Learning and Attitudes

Table 1 shows the results for pre- and post-test scores for the three groups, i) Intervention Group (used the tutor with interventions every six problems), ii) Tutor Control Group (used the tutor w without the interventions) and iii) No-Tutor Control (matched students who used no software).

Group	Math Pretest	Math Posttest	MCAS Passing Rate
No Tutor Control			76% (N=38)
Tutor Control	40% (20) (N=40)	40% (28)* (N=40)	79% (N=34)
Tutor Intervention	33% (19) (N=36)	42% (22)* (N=36)	92% (N=24)

Table 1. Means and standard deviations in performance measures before and after tutoring

The overall learning gain (Posttest-Pretest) for the experimental group was 7%, while the Tutor Control group showed no improvement (Table 1). These learning gains are smaller than what we had observed previous years (15% in about the same amount of time). We think that the overall posttest scores underestimate students' ability because: 1) it was online, and we observed gaming behavior particularly in the posttest; 2) it was started at the end of a period, when students were already tired. The fixed sequencing might have affected learning gains also (in contrast to the adaptive sequencing of problems). In any case, the low posttest scores do not prevent us from carrying out a between-subjects comparison. An ANCOVA was used to analyze the difference between the learning gains between the two groups (tutor Intervention and tutor-control). The dependent variable was posttest score (percent correct), with group as an independent variable, and pretest as a covariate. The test of between subjects indicated a significant difference in posttest score between the tutor-control and tutor-Intervention groups ($F=4.23$, $p=.04$), suggesting that there is a significant difference in learning gains favoring the interventions-enhanced tutor group.

Because the experiment was carried out days before a statewide-standardized test exam (MCAS), we collected standardized scores as well for students who took the exam, including the matched group of students (same level, same teachers) who did not use the tutor. Note that students in the Interventions group had a higher average learning gain (7% more than the control group) and higher passing rate at the MCAS exam (92% vs. 79%) than their counterparts in the control group, and higher than the no tutor control group (92% vs. 76%). A Chi-square test was carried out on the cross-tabulation between MCAS passing (1 for passed, 0 for did not pass) and version (Tutor Intervention and No Tutor Control), indicating a marginally significant difference ($p=.12$) that suggests that the Tutor Intervention group had significantly higher passing rate.

Table 2 shows results of the surveys that were higher for the Interventions group. Students in the Interventions group agreed more with the statements such as “the Wayang tutor was smart and friendly”. They also had significantly higher learning orientation scores in two items that measure performance vs. learning orientation (Mueller&Dweck, 1998). Marginally significant differences were observed for students thinking they have learned with the Wayang tutor and beliefs about the helpfulness of the help, all favoring the Interventions group. No significant differences were encountered for questions about ‘computers caring about myself’, ‘Wayang is genuinely concerned about my learning’, feeling of control over computers, mathematics liking, ‘the tutor is concerned about my learning’, self-concept about mathematics ability, or self-efficacy. These last ones are deeper feelings about oneself, and the Interventions don’t seem to have impacted at that level, but at a simpler level of perceptions of the system, its helpfulness, and their willingness to learn.

Survey question item	Tutor Interventions	Tutor Control
“The Wayang Tutor is friendly” ANOVA: F=6.5, p=.01**	4.8 (1.0) N=21	3.9 (1.4) N=35
“The Wayang tutor is smart” ANOVA: F=6.5, p=.01**	5.1 (1.0) N=21	4.3 (1.3) N=35
Learning Orientation (average over 2 items) ANOVA: F=4.2, p=.045*	0.60 (.8) N=21	0.39 (.6) N=37
Did you learn how to tackle math problems by using the Wayang system? ANOVA: F=2.9, p=.09 (marginal)	3.5 (.74) N=22	3.1 (.82) N=37
Helpfulness of the help (average over 3 items) ANOVA: F=2.5, p=.1 (marginal)	4.2 (.73) N=21	3.8 (.9) N=36

Table 2. Means and standard deviations for responses to post-tutor surveys

2.2 Within-subjects analysis of changes in engagement

The second analysis has to do with students’ variation of engagement within the tutor. If the interventions were effective, it would be expected that there is a difference in indicators of engagement, from the problem before the intervention was given to the problem after the intervention was given. Some other authors have suggested time as an indicator of engagement (Beck, 2004). If the student is not engaged, then the student will either skip fast through the problem, or abuse help quickly to get to the correct answer, and all gaming behaviours are related to timing issues (Baker, 2006). Spending long time in a problem is an insufficient condition to learning (the student might be off-task), but it is also a necessary condition (if the student doesn’t invest time on the problem, they will surely not learn). Students might be off task when spending much time in a problem, but if outlier cases of time spent per problem are discarded, or median/quartile statistics are taken into account instead of means/standard-deviations, we can look at the time spent per problem variable with some confidence that what we are looking at is engagement and focus on the problem.

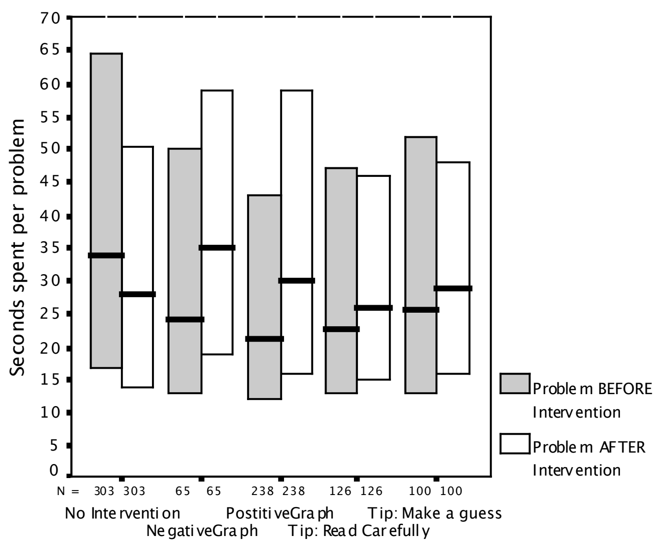

Figure 2. Medians and Quartiles for time spent in the problem immediately before and after the intervention

2.2.1 Time spent per problem. Difference in subsequent problems.

The main question is how a student’s behavior changes during the time spent per problem from before to after the intervention. How do different interventions affect the time spent in a problem? What is the difference between the times spent during the two problems? The first two boxes in the Box Plot of Figure 2 show the median and quartile seconds spent per problem for 303 random pairs of subsequent problems, *for students in the tutor-control* group. These two boxes suggest that students tend to get more disengaged as the session progresses (median and quartiles are lower in the second problem), by a median 6 seconds less in the next problem. The eight boxes to the right correspond to the seconds spent in 469 problems immediately before and immediately after each intervention, for students in the Interventions group. Clearly, there is a reverse effect of students decreasing time in subsequent problems: students increase the median time spent in the problem after seeing any intervention.

There are two reasons why carrying out regular statistical analyses is difficult with this kind of data. First, a standard analysis of variance or means comparison relies on the fact that probability distributions should be normal. This is not the case of this data set, as distributions for time-based variables have minimums at zero and long tails for outlier times. Second, each case corresponding to the time spent in a pair of problems is not independent from each other, because there are several pairs corresponding to each student. Thus, ANOVA statistical analyses are merely exploratory, but we believe worth as an indication of the strength of the results. Taking this into account, and after removing outlier time values, repeated measures ANOVA indicated a significant difference in time change within (F=8.79, p=.003) and between the two groups (F=7.3, p=0.007). However, note that the shift in time is more pronounced for the graph interventions than for the tip interventions. A paired-samples t-test gave a significant difference for time spent from the problem before to the problem after a Graph Intervention (t=-2.9, p=0.004), but not for before and after the tips (t=.69, p=.49).

2.2.1 Do disengaged students game less after seeing an intervention?

If time spent in a problem reflects level of engagement, students apparently become more disengaged as time progresses in the tutoring session if the tutor does nothing except presenting additional problems. The main question is whether there is a way to show that students reduce specific disengagement actions. In Wayang, we may answer this question by referring to a model developed by Johns and Woolf (2005) from past data of Wayang users, which allows to classify students in either an 'engaged' state or otherwise three types of disengagement: i) quick-guess, ii)hint-abuse, and iii) skipped-problem. A hidden markov model infers the probability that a student is in any of these states for each problem, depending on their behavior on the current problem and the motivation state in the previous time step. The student is classified as engaged/disengaged based on the engagement state of highest likelihood. Disengagement states are similar to those identified by Baker (2006) regarding gaming the system –all measures of disengagement that correlate to learning.

An important question to answer regards the potential of these interventions to re-engage students compared to the control group –what are the odds that a student will turn from one disengagement state back to the engaged state after seeing an intervention, compared to those students who did not see any interventions at all. 347 episodes of 6 subsequent engagement states were extracted for the control group (corresponding to six subsequent problems), and 433 sequences were extracted for the experimental group. The latter sequences had the peculiarity that they corresponded to three problems before an intervention was shown, and three problems after an intervention was shown. Figure 3 shows how many students were classified as engaged or disengaged at each time step, for each group. Note that the probability that a student will continue to stay in the same engagement state or switch to the other state (labelled arrows in the figure) were computed from students' data, for example: 79 students from the interventions group were considered disengaged in time t_4. 49% of those 79 students switched to being engaged in time t_5. Thus, the probability of re-engaging at t_5, or $P(E_{t5}|\sim E_{t4})$, is 0.49. The other transition probabilities are the probability of disengaging, (e.g. $P(\sim E_{t5}|E_{t4})$), staying engaged (e.g. $P(E_{t5}|E_{t4})$), staying disengaged (e.g. $P(\sim E_{t5}|\sim E_{t4})$) and were computed in a similar way. All probabilities were rounded to the second decimal.

	t_4	T_5	t_6		t_7	t_8	t_9
Interventions	0.18	0.18	0.20		0.18	0.21	0.20
Control	0.17	0.19	0.22		0.22	0.19	0.18
Difference	*0.01*	*-0.01*	*-0.02*		*-0.04*	*0.02*	*0.02*

Table 3. Probability of a student being disengaged at $t_4, \ldots, t_9$

Arrows going from bottom states to top states in Figure 3 indicate the odds that a student will become re-engaged in the task, or $P(E_{ti}|\sim E_{ti-1})$. Note that these re-engagement probabilities are generally larger for the interventions group, except for the transition between t_5 ad t_6, where they are equal. Noticeably, the largest difference between the two groups happens immediately after the intervention, between t_6 and t_7 (there is a 16% advantage for the interventions group in t_6-t_7, 5% advantage in t_7-t_8, 11% in t_8-t_9). However, note that in general students in the interventions group are also more likely to disengage by a small margin. This suggests that once students are engaged, that engagement status might be hard to maintain. The interventions group is 1% more likely to disengage in t_6-t_7 than the control group, 3% in t_7-t_8, 4% in t_8-t_9. It is possible that the interventions may have only a short-term positive impact on engagement, though it is also possible that exposing students who are already engaged to an intervention might be unproductive. These two reasons help explain why there is not much of a difference between the groups in total percentage of engaged students at each time step (Table 3), nor an 'overall' increase in total students considered engaged for the interventions group from time t_4 to time t_9. Students in the

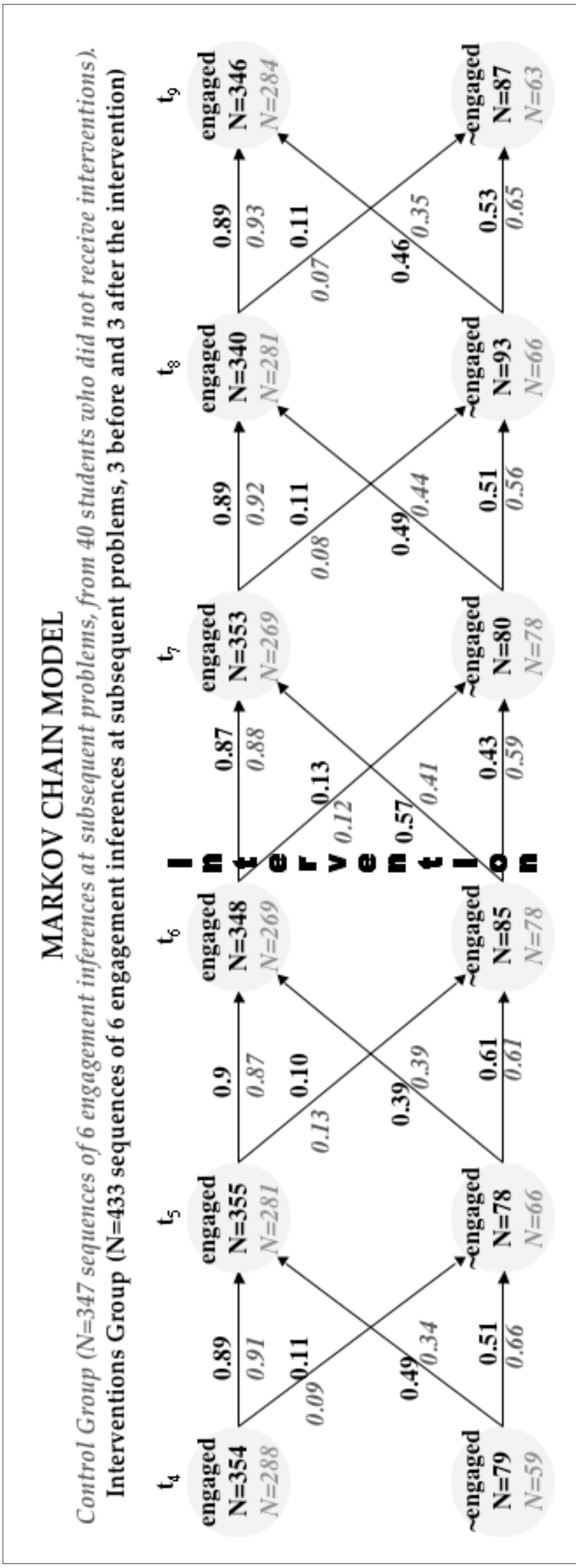

Figure 3. probabilities of disengagement, re-engagement, or staying in the same state (computed from data)

experimental group clearly oscillated more between engaged and disengaged states, as both probabilities of disengagement and re-engagement are generally higher than the control group. There is much more movement between engaged and disengaged states for students in the experimental group.

Another interesting question to explore is whether students modified specific kinds of disengagement behaviors. With that thought in mind, Table 4 shows transition probabilities between the before-problem and the problem after the intervention --between t_6 and t_7. These transition probabilities were computed discriminating by type of disengagement. For instance, the first column in table 4 indicates the probability of transitioning from any state into the engaged state after seeing an intervention. At the same time, the first column in Table 5 is our control, the probability of transitioning into an engaged state, after not seeing any interventions).

Table 6 is the difference between the previous matrices: positive values indicate that the probability is higher for the motivational group, negative values indicate the opposite. Therefore, the first column in Table 6 shows that students re-engage more (transition more from disengaged states into the engaged state) after seeing an intervention. Students who quick-guessed in the problem before the intervention had a 12% higher chance to re-engage than those in the control group (same problem, but no intervention). Students who abused help in the problem before the intervention had 10% higher chance of re-engaging in the following time step than the control group. Students who skipped the problem before the intervention have 51% higher chance of being considered engaged in the following time step than the control (though it is a low number of students skipping problems, 5 out of 7 skipping-students re-engaged after the intervention vs. 1 out of 5 in the control group). Also, as the second column in Table 6 shows, students have less chance of transitioning into the quick-guess state after an intervention than in the control group. Table 6 also shows that students in the Interventions group are less prone to stay in the same disengagement state (bold diagonal). This may mean transitioning from one form of disengagement state into another form of disengagement. For instance, students have 10% higher chance to transition from hint-abuse to skipping the problem after seeing an intervention. However, this might just be due to the low number of cases, as students did not abuse help much. Fortunately, most students who skip the problem before the intervention (71%) re-engage in the following problem (51% higher than the control group).

		Problem after the Intervention was seen			
		Engaged	Quick-Guess	Hint Abuse	Skipped
Problem Before Intervention	Engaged (348)	**.88 (305)**	.06 (22)	.04 (13)	.02 (8)
	Quick-guess (58)	.53 (31)	**.45 (26)**	0 (0)	.02 (1)
	Hint Abuse (20)	.60 (12)	.10 (2)	**.20 (4)**	.10 (2)
	Skipped (7)	.71 (5)	0 (0)	0 (0)	**.29 (2)**
	Totals (433)	*.82 (353)*	*.12 (50)*	*.10 (17)*	*.03 (13)*

Table 4. Transition Probabilities between t_6 and t_7 (before and after Intervention), Interventions group.

		Problem after the Intervention was seen			
		Engaged	Quick-Guess	Hint Abuse	Skipped
Problem Before Intrvention	Engaged (269)	**.88 (237)**	.09 (23)	.03 (8)	.00 (1)
	Quick-guess (61)	.41 (25)	**.57 (35)**	0 (0)	.02 (1)
	Hint Abuse (12)	.50 (6)	.25 (3)	**.25 (3)**	0 (0)
	Skipped (5)	.20 (1)	.20 (1)	0	**.60 (3)**
	Totals (347)	*.78 (269)*	*.18 (62)*	*.03 (11)*	*.01(5)*

Table 5. Transition Probabilities between t_6 and t_7 in the Control Group

		Problem after the Intervention was seen			
		Engaged	Quick-Guess	Hint Abuse	Skipped
Problem Before Intervent ion	Engaged	**0.00**	-0.02	0.01	0.02
	Quick-guess	0.12	**-0.13**	0.00	0.00
	Hint Abuse	0.10	-0.15	**-0.05**	0.10
	Skipped	0.51	-0.20	0.00	**-0.31**
	Totals	*0.04*	*-0.06***	*0.01*	*0.02*

Table 6. Matrix Difference: Intervention – Control transition probabilities (t-test p<.01 **)

Because these transition probabilities can also be considered means computed over sample data, an independent samples t-test can help us to explore if these differences are significant. Again, there is an assumption that the cases are independent when they really are not (several cases correspond to the same student) which is why this is an exploratory analysis. Only about 10% of students were classified as disengaged at any time step, so significant differences are rare due to low number of cases. Still, a t-test revealed that students quick-guessed significantly less in the problem after the intervention, 6% less than in the matched problem in the control group. Table 7 breaks down the number of quick-guess cases in the problem after (how many quick-guesses happened after a tip intervention, and how many after a graph intervention?). There is a significant difference in mean percent of quick-guesses after a graph intervention (Graphs – Control= -.08, p=0.005), but not after the tip-interventions (Tips – Control = -.04, p=0.18). It is the problems after a student received a graph performance-monitoring intervention, not the problems after receiving a tip intervention, in which students have lowest amount of quick guesses (compared to the engagement state in matched problems after no intervention).

Kind of Intervention	N	Mean % of quick-guess	Std. Deviation	ANOVA
No Intervention-Control	347	.18	.38	F=6.3 p=.02*
Tip Intervention	191	.14	.31	
Graphs Intervention	242	.10	.30	

Table 7. Means, std. deviations and univariate ANOVA for mean quick guesses in t_7
(* significant difference at p<0.05)

3. Discussion and Conclusion

Despite the fact that students in the Interventions group started off with slightly lower pre-test scores, they achieved higher posttest scores than the control group, higher passing rates in standardized tests, higher learning-orientation in a post-tutor survey, had the feeling the tutor was more helpful and that they learned more, and attributed more human-like characteristics to the tutoring software. We attribute these differences to the non-invasive Interventions provided to students in between problems. Students have a higher chance to become re-engaged immediately after seeing an intervention, regardless of whether they quick-guessed, abused help or skipped the problem before the intervention. It is in particular the progress-monitoring interventions that helped students have less quick-guesses, and higher changes in time spent per problem.

Showing students their progress with a simple open-learner model can be considered a type of meta-cognitive feedback. It encourages self-monitoring by highlighting that both the software and themselves can keep track of whether they are answering correctly or incorrectly, and whether they are making progress. Our simple progress charts talk to the student in terms of correctly answered questions so that students can clearly link their actions to the chart, and the chart to their learning progress. We think it is showing students' performance in particular that re-engages students because being hinted of ones' progress together with tips provided for lower quick-guesses and higher improvements in time spent per problem than the tips by themselves. Time spent per problem especially improved if the graph provided was negative --indicated that no improvement had been made. One possible explanation is that negative feedback may have encouraged students to be more cautious, to reduce the chances of getting negative feedback again. However, it is important to note that the data analyzed indicates no clear benefit in sustaining engagement. A measure of success for future interventions will be not only how much it can re-engage a student in the following problem, but for how long the student continues to be in such state. One other important aspect is that the timing of interventions may be relevant. They may be more beneficial to show immediately after the problem where the student is disengaged, though this has not been established yet. The decision of whether to present an intervention after each problem or not, depending on the assessed disengagement state and other factors may be relevant to enhance engagement, and learning and motivation in the long run.

In the near term we plan to design of new kinds of interventions, as well as analyzing which ones are beneficial for which students, and at which moments. Firstly, something to revisit is the accuracy of the graph interventions in assessing students' knowledge. Our current graph interventions are very simple, and possibly inaccurate some times at showing the student how much they have learned --comparing percentage correct responses in the last problems is not an accurate enough measure to assess how much students know, in the face of large variation of problem difficulties and skills involved. On the other hand, we believe that merely reporting number of correct responses offers a very tangible measure for students to understand and relate to their actions and their progress, so we want to find a representation to externalize students' progress that is thorough but also easily understandable. Secondly, new kinds of meta-cognitive feedback (beyond self-monitoring) could be beneficial for students, such as specific feedback on planning how to solve a specific kind of problem, or motivational feedback that acknowledges students' feelings of frustration, or helps establish productive attributions for failure and success (Mueller and Dweck, 1998).

We conclude that non-invasive interventions that help students reflect on their progress are beneficial to students learning, attitudes towards learning, and towards the tutoring software. These research results provide valuable insight in support of in-between problem meta-cognitive feedback, and in support of open learner models for the design of effective software interactive learning environments.

References

Arroyo, I., Beal, C. R., Murray, T., Walles, R., Woolf, B. P. (2004). Web-Based Intelligent Multimedia Tutoring for High Stakes Achievement Tests. In Proceedings of the 7th International Conference on Intelligent Tutoring Systems, pages 468-477.

Baker, R.; Corbett, A.; Koedinger, K. (2004) Detecting student misuse of intelligent tutoring systems. In Proceedings of the 7th International Conference on Intelligent Tutoring Systems, pages 43--76.

Baker, D.J., Beck, J. (2006) Adapting to When Students Game an Intelligent Tutoring System. Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 392-401

Beck, J. (2004) Using response times to model student disengagement. Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments, August, 2004

de Jong, & van Joolingen, W. R. (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains . Review of Educational Research, 68, 179-201.

Johns, J.; Woolf, B.P. (2006). A Dynamic Mixture Model to Detect Student Motivation and Proficiency. Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI-06), Boston, MA.

Hartman, H. J. (2001). Metacognition in Learning and Instruction, Hartman, HJ (ed). Springer.

Mueller, C.M., Dweck, C.S. (1998). Praise for intelligence can undermine children's and performance. Journal of Personality and Social Psychology , 75 (1), 33-52.

Roll, I., Aleven, V., McLaren, B. M., Ryu, E., Baker, R., and Koedinger, K. R. (2006). The Help Tutor: Does Metacognitive Feedback Improve Students' Help-Seeking Actions, Skills and Learning?

Wigfield, A. & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in academic situations. Educational Psychologist, 26, 233--262.