EARLY EXPERIENCE WITH COMPUTER-SUPPORTED COLLABORATIVE EXERCISES FOR A 2nd SEMESTER JAVA CLASS

Jung Hee Kim
Dept. of Computer Science
North Carolina A&T State University
1605 E. Market Street, Greensboro, NC 27411
(336) 285-3695
jungkim@ncat.edu

Taehee Kim
Dept. of Human Development and Services
North Carolina A&T State University
1605 E. Market Street, Greensboro, NC 27411
(336) 285-4388
tkim@ncat.edu

Michael Glass
Dept. of Computing and Information Science
Valparaiso University, Valparaiso, IN 46383
(219) 464-5161
michael.glass@valpo.edu

ABSTRACT

This paper reports on two semesters experience with computer-mediated group discussion exercises in a CS2 computer programming class. The class is a gateway for computer science and computer engineering students, where many students have difficulty succeeding well enough to proceed in their major. The exercises focus on Java concepts. They are designed to require students to rely on each other, but also be individually accountable. Learning gains measured in this trial have been mixed, with the least prepared student (as measured by pretest) in each discussion group showing the highest learning gains, while best prepared student in the discussion group showed score reductions on average. This paper reports on first year results of learning gains and of surveys of student experience with the exercises.

INTRODUCTION

This project reports on the first year of experience in a project to develop computer-mediated group discussion exercises for students in a second-semester computer programming class [9]. These exercise utilize the COMPS computer-mediated problem solving platform [2, 6, 7]. The project's goal is to develop 10 such exercises, along with chat software for hosting the exercises and additional software to aid the instructor administering them. The exercises utilize relational understanding and serve as a complement to the more instrumental skills developed by traditional programming assignments. In this first year four exercises were completed and administered.

An aim of the project is to provide a positive learning experience through group interaction. The positive experience ideally should be expressed in terms of student engagement and self-efficacy. At North Carolina A&T State University the class is a gateway for computer science and computer engineering students, where many students have difficulty succeeding well enough to proceed in their major. The combination of improved understanding and a more positive experience is intended to improve success in the class and persistence in the major.

Guiding and assessing the project thus includes measuring a) learning and understanding of the programming concepts, b) student interest and engagement in the exercises and in the major, and c) student self-efficacy. The results reported here are not from controlled experiments. These measurements are being used to guide development and revision of the labs.

BACKGROUND

The exercises

COMPS Computer-mediated solving exercises are administered during regular programming lab time. Students log into a web site and communicate with typed-chat. Typically they are assigned to 3-person groups. In these exercises the problem prompt is divided into several steps. Students must agree on an answer and have the answer checked by an instructor or TA before proceeding to the next step. The protocol is designed to encourage creative interdependence, where the students must rely on each other to compete the task [3]. The chat part of an exercise typically lasts half an hour to an hour, it is sometimes preceded by a pre-test and followed by a post-test and a survey.

The topics of the four labs were: Lab 1 – classes, objects, and references, Lab 2 – inheritance, Lab 3 – Swing (Fall) or JavaFX (Spring) graphical user interface, Lab 4 – exceptions and exception handling. The two versions of Lab 3 explored the same concepts in the two different GUI APIs.

Student interest, engagement, and self-efficacy

Interest refers to an individual's psychological inclination to participate in particular content over time. There is an intimate relationship between interest, achievement goals, performance and retention [8]. Interest plays a critical role in students' further decisions on engaging and reengaging in the major. Mitchell [11], as an example, reported that using group work activities, computer-based activities, engaging puzzles, and meaningful activities, were correlated with triggering and holding interest in a mathematics classroom.

There can be a strong social component to engagement. Recently Kim and Schallert [10] have investigated how interpersonal interactions affect interest in the college learning environment. It is possible to track student interest in four developmental phases throughout a semester, not just within the time frame of individual activities. Interest is affected not only by the enthusiasm expressed by the teacher and fellow students, but also by factors such as affiliative motivation. The social factors enhancing interest were found through classes in a number of diverse disciplines in both upper and lower level college classes.

Group exercises squarely address many of the components of interest. The exercises are constructed so that students engage with other students, providing the small-group interpersonal contact that transmits enthusiasm. The students know the teacher is watching the conversations and is taking an active interest in the students' progress, sometimes by intervening and sometimes by providing answers and hints as part of the plan of the exercises. It is possible to see students expressing enthusiasm to each other in their dialogue. The COMPS chat interface has additional, extra-linguistic affordances through which students can express affective state. Students sometimes express enthusiasm by switching to a Comic Sans typeface, bigger bolder fonts, or wild colors.

Self-efficacy, an individual's belief to be capable of performing a particular task [1], has been widely studied because of its relationship to performance including academic achievement and choice of major in college [5]. The COMPS Java exercises usually address student self-efficacy by pairing concepts with specific tasks and objective answers. For example, an exercise that addresses concepts in Java exception handling will have a component where the students hand-trace some code and predict what it will print.

Measurement of Learning Gains

Learning gains were measured by pre- and post-tests. The tests were revised for the second semester. Learning gains were calculated as:

(post – pre) / (max score – pre)

The pre- and post-tests had a maximum score of 20 for Fall 2015 and maximum score of 15 for Spring 2016. In this calculation, the maximum learning gain is 1.0. When the post-tests are equal to the pre-test the learning gain is 0.0. The learning gain will be less than 0 when the post-tests are lower. Pre- and post-tests were administered online and were not for credit. Administering pre-tests was started in lab 3 of the fall semester, so learning gains were measured for only labs 3 and 4. Furthermore learning gain data is sparse for some labs because a few students did not complete both pre- and post-tests, which weren't graded work.

We also disaggregated the students into three strata, according to the relative preparedness within their discussion group based on their pre-test scores before the discussion. The theory is that different students in a discussion may adopt different roles. The best-prepared student (measured immediately prior to the exercise) will have a different part in the discussion than the worst prepared (also measured immediately prior). The strata are: the highest pre-test scoring person in each discussion, the lowest scoring person in each discussion, and the middle student. A few discussions had only two students with no middle student.

Measurement of Student Interest, Self-Efficacy, and Experience of the Lab

Students were surveyed after each exercise. These questions measure the student's perception of the effectiveness of the group discussion, the student's perception of learning, and the student's interest. Survey questions related to this study are shown in Figure 1.

In addition to the surveys for each lab, students were also surveyed near the beginning and end of the semester regarding their enthusiasm for the class, their self-efficacy in programming certain class topics, and their desire to continue. The enthusiasm questions were derived from [8]. Finney and Schraw [4] emphasized it is important to survey the students regarding their perceived ability in named particular skills, instead of querying more generally. The self-efficacy items on the semester survey were regarding a list of skills that might appear as Java programming learning objectives for the semester class. Not all students completed both surveys.

1. I gained an understanding of the concepts in this exercise.	5. This lab is exciting.
2. I contributed to the understanding of the other students in my group.	6. This lab has things that grab my attention.
3. I would have preferred to figure out this exercise by myself.	7. This collaborative exercise was engaging.
4. The other students in my group contributed to my understanding.	8. This collaborative exercise seemed to drag on forever.

Figure 1. After-Lab Survey Questions

ADMINISTERING EXERCISES

We administered each of the four exercises during both the Fall 2015 and Spring 2016 semesters of the CS2 class at North Carolina A&T State University. Participating in these labs was voluntary. Each exercise was administered shortly after its topics had been taught in class. In Fall there were 54 students enrolled at the start of the semester, of which 29 finished the semester still enrolled with a passing grade of C or higher. There were 53 group discussions in total among the four exercises. Altogether there were about 8000 dialogue turns. In Spring 16 enrollments were higher, total 73 students, of which 49 finished with C or higher grade. There were 75 total group discussions among the 4 labs.

For the lab exercises most groups had 3 participants. The bulk of students were assigned to sessions quasi-randomly as students arrived in lab. Cliques of friends, who tended to arrive together, were split into different random groups. Students in a discussion group were dispersed around the room so they would not communicate except through the chat interface.

From Fall we have complete pre-/post-test pairs for 32 students in Lab 3 and 24 students in Lab 4. In Spring we have 55 completed pairs for Lab 3 and 50 for Lab 4.

FINDINGS

Learning gains

The main finding from the learning gains is that the best prepared student in each group discussion performed quite differently than the other two students.

The class average learning gains were positive. Table 1 shows learning gains in aggregate for the whole class for labs 3 and 4 for both semesters. We also disaggregated the students into the three strata, according to the relative preparedness in their discussions based on their pre-test scores before the discussion.

Disaggregation was telling: the highest stratum best-prepared students exhibited negative or near-zero learning gains. The other strata exhibited positive learning gains.

Within each stratum, we tested whether the pre-to-post score change is significant. Using paired t-tests, the score changes for the three different stratum levels were examined, see Table 2 for Lab 3. For the middle and lower strata there were statistically significant differences between pre- and post- scores, their learning gains were significant at the $p<0.05$ level. However for the high scorers the differences between pre- and post-scores were not significant, with $p=0.08$ (Fall) and $p=0.77$ (Spring). According to Tukey's post-hoc test, there was not a statistically significant difference between the middle and low strata. However both the low and middle groups showed significantly greater score changes than the high group did. Lab 4 results were similar: significant learning gains in the lower two strata, and the near-zero learning gains of the highest stratum were significantly different than the other two.

Table 1. Average Learning Gains.

	Class average		Highest Stratum	Middle	Lowest
Fall 2015 Lab3 (n=32)	0.23		-0.32	0.45	0.34
Fall 2015 Lab4 (n=24)	0.32		0.02	0.34	0.47
Spring 2016 Lab3 (n=55)	0.26		-0.05	0.35	0.40
Spring 2016 Lab4 (n=50)	0.54		-0.13	0.57	0.73

Table 2. Lab 3 Pre-test vs. Post-test Scores by Stratum

		Pre-Test Mean / SD		Post- Test Mean / SD		t	p
Fall 15 Lab 3	High	12.25	3.41	9.75	6.06	t (11) = 1.94	p = 0.078
	Mid	5.92	5.05	12.25	4.41	t (11) = -3.74	*p = 0.003
	Low	3.75	3.10	9.25	4.56	t (7) = -4.72	*p = 0.002
Spring 16 Lab 3	High	10.41	2.08	10.18	3.08	t (21) = 0.303	p = 0.765
	Mid	8.13	2.23	10.50	2.79	t (16) = -2.57	*p = 0.021
	Low	5.29	2.79	9.15	4.00	t (16) = -3.34	*p = 0.004

Student attitudes

Table 3 shows the results of after-lab surveys of the student experience in Fall 2015. Survey items were on scale from 1 to 5. Labs 3 and 4 were generally better perceived than the first two labs. Students perceived:

a) More effective group work in Labs 3 and 4 (mean rose from about 3.1 to about 3.4)

b) Better understanding of concepts in Labs 3 and 4 (mean rose from about 3.4 to about 3.9).

c) Interest was higher in Labs 3 and 4 than the other two labs, but fluctuated.

Multiple one-way ANOVA supports the hypothesis that mean scores are indeed different among the four labs, $p = 0.03$ for both effectiveness and understanding. Post hoc analyses using the Tukey test for significance indicated that the mean scores of Lab 3 were significantly higher than Lab 2 for both effectiveness and understanding.

Table 3. After Lab Surveys in Fall 2015 [9]

	Effectiveness of group work Mean / SD		Understanding of concept Mean / SD		Interest in lab Mean /SD	
Lab1	3.17	0.68	3.45	0.96	3.19	0.94
Lab2	3.08	0.93	3.42	1.05	3.08	0.93
Lab3	3.47	0.71	4.03	1.06	3.65	0.76
Lab4	3.40	0.61	3.78	0.85	3.17	0.89

Changes from beginning to end of semester

At the beginning and the end of the semester, students' interest toward the course and self-efficacy were assessed in order to examine any changes throughout the semester. Survey items were on the same 5-point scale as the lab questions. For example, one of the interest questions is "This year, I really enjoyed what we did in GEEN 165." A self-efficacy question is: I am able to "explain constructor chaining between a series of classes and subclasses." Regarding the students' interest toward the course, their level of interest stayed very similar as shown in Table 4. Meanwhile, students' self-efficacy toward the course content indicated significant improvement between beginning and end of the semester.

Table 4. Student Interest and Efficacy Changes from Beginning to End of Semester

	N	Beginning Mean	Ending Mean	t	p
Fall 15 Interest	28	4.33	4.32	$t(27) = 0.09$	$p = 0.299$
Spring 16 Interest	39	4.29	4.23	$t(38) = 0.47$	$p = 0.664$
Fall 15 Efficacy	28	2.83	3.81	$t(27) = -5.70$	$^{*}p = 0.000$
Spring 16 Efficacy	39	3.23	3.69	$t(38) = -0.264$	$^{*}p = 0.012$

DISCUSSION

The presence of learning gains for most of the class justifies using the COMPS discussion exercises, but the negative learning gains (on average) for the best prepared student in each discussion was puzzling. These students had not come near to maximum scores on their pre-tests. Had they done so simple randomness could produce an average negative learning gain. A possible explanation comes from the average post-test scores, as shown in Table 5. In both labs the post-test scores were not significantly different among the three strata. One possibility is that within a discussion group the students came closer to a common understanding of how to approach the problem, leading to a commonality of post-test scores. The class average learning gains of 0.2 to 0.5 indicate that there may be room to get more effective learning from the exercises.

Between Fall and Spring semesters the pre- and post- tests changed. The primary change was to recast the pre-test to match the first few parts of the exercise. The motivation was partly to have the pre-test more accurately assess the state of knowledge on the particular items of the exercise, and partly to increase student engagement with the exercise. The principle is that immediately after encountering the individual problem on the pre-test, the students would be primed to want to know the correct answers and to discuss with each other.

Table 5. Post-test Scores for Each Scoring Level

	Fall 2015 Lab 3	Fall 2015 Lab 4	Spring 2016 Lab 3	Spring 2016 Lab 4
High	9.75	12.9	10.18	12.18
Middle	12.25	11.0	10.50	11.72
Low	9.25	11.5	9.15	12.50

The student attitudes show that they found the experience significantly better during the second two exercises than the first two, both in terms of perceived effectiveness of the exercises and self-efficacy. Anecdotally, from reports of lab teaching assistants, it seems that the students became better accustomed to doing the discussions as they became routine. However, these mean scores in the 3.5 range have room for improvement.

FUTURE WORK

Digging into the different experience of the higher stratum is a priority. The student experience surveys will be disaggregated by their preparedness strata. It is possible that the best-prepared student in each discussion was, on average, having a worse experience than the other two. This would possibly correlate with their negative learning gains. The professor reports, again anecdotally, that a number of the higher achieving students in the class became dissatisfied with spending effort teaching the others students during this exercise.

Another priority is to try to understand whether it is possible to predict group functioning based on prior data from the students. Do the student perceptions or pre-/post-test of earlier labs inform performance on the next lab? When three students who have had low interest or low perceived value of the former group exercise are put together in the same group, does it affect either learning gains or student experience?

This project is developing more exercises for CS2, with the goal of ultimately making ten such exercises available to the CS education community. An exercise on one- and two-dimensional arrays will be added. Existing topics will be split into two, shorter exercise sets. The aim is to achieve higher learning gains. Partly this will be attempted through breaking some of the harder problems into shorter steps and introducing some repetition. Another change will be to add a related hands-on Java experience immediately following each exercise within the 110-minute lab period, necessitating shorter discussions. These exercises may be tested in a design experiment against a condition where the same work is completed by the students, alone without the group discussion.

As a way to stimulate engagement, there is also discussion of adding an inter-group competitive element to the exercises.

ACKNOWLEDGEMENTS

Partial support for this work was provided by the National Science Foundation's Improving Undergraduate STEM Education (IUSE) program under Award No. 1504917. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Bandura, A. *Self-efficacy: the Exercise of Control*. New York: Freeman. 1997.

[2] Desjarlais, M, Kim, J. H., Glass, M. COMPS Computer Mediated Problem Solving: A First Look, *Proceedings of the Midwest AI and Cognitive Science Conference* (MAICS 2012), 50–57. 2012.

[3] Eberly Center. Carnegie Mellon, Eberly Center for Teaching Excellence and Innovation. What are best practices for designing group projects? http://www.cmu.edu/teaching/designteach/design/instructionalstrategies/groupprojects/design.html. Retrieved March, 2016.

[4] Finney, S. J., & Schraw, G. Self-efficacy Beliefs in College Statistics Courses. *Contemporary Educational Psychology*, 28 (2), 161-186. 2003.

[5] Hackett, G. Role of Mathematics Self-efficacy in the Choice of Math-related Majors of College Women and Men: A Path Analysis, J*ournal of Counseling Psychology*, 32, 47–56. 1985.

[6] Glass, M., Kim, J. H., Bryant, K. S., Desjarlais, M. Indicators of Conversational Interactivity in COMPS Problem-Solving Dialogues, *Intelligent Support for Learning in Groups (ISLG-3)*. 2014.

[7] Glass, M., Kim, J. H., Bryant, K. S., Desjarlais, M. Come Let Us Chat Together: Simultaneous Typed-Chat in Computer-Supported Collaborative Dialogue, *Journal of Computing Sciences in Colleges*, 31 (2), 96–105. 2015.

[8] Harackiewicz, J. M., Durik, A.M., Barron, K.E., Linnenbrink-Garcia, L, Tauer J.M. The Role of Achievement Goals in the Development of Interest: Reciprocal Relations Between Achievement Goals, Interest, and Performance, *Journal of Educational Psychology*, 100, 105–122. 2008.

[9] Kim, J.H., Glass, M., Kim, T.H., Bryant, K., Willis, A., McNeill, E., Thomas, Z. Student Understanding and Engagement in a Class Employing COMPS Computer Mediated Problem Solving: A First Look, *Proceedings of the Modern AI and Cognitive Science Conference* (MAICS 2016), Dayton, OH. 2016.

[10] Kim, T.H. and Schallert, D. Mediating Effects of Teacher Enthusiasm and Peer Enthusiasm on Students' Interest in the College Classroom, *Contemporary Educational Psychology,* 39 (2) 134–144. 2014.

[11] Mitchell, M. Situational Interest: Its Multifaceted Structure in the Secondary School Mathematics Classroom, *Journal of Educational Psychology*, 85, 424-436. 1993.