Student Understanding and Engagement in a Class Employing COMPS Computer Mediated Problem Solving: A First Look

Jung Hee Kim
North Carolina A&T State U.
jungkim@ncat.edu

Michael Glass
Valparaiso U.
michael.glass@valpo.edu

Taehee Kim
North Carolina A&T
tkim@ncat.edu

Kelvin Bryant
North Carolina A&T
ksbryant@ncat.edu

Angelica Willis
North Carolina A&T
awillis@aggies.ncat.edu

Ebonie McNeil
North Carolina A&T
eimcneil@aggies.ncat.edu

Zachery Thomas
North Carolina A&T
zithomas@aggies.ncat.edu

Abstract

COMPS computer-mediated group discussion exercises are being added to a second-semester 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. This paper reports on first results of surveys on student experience with the exercises. It also reports on the affective states observed in the discussions that are candidates for analysis of group functioning. As a step toward computer monitoring of the discussions, an experiment in using dialogue features to identify the gender of the participants is described.

Introduction

The second Java programming class, GEEN 165, at North Carolina A&T State University is a bottleneck for many Computer Science and Computer Engineering students. As an experiment in improving student learning and interest, COMPS computer-mediated discussion exercises (Glass et al., 2014a) have been introduced. This paper reports on first measurements of a) student self-efficacy and interest, b) expressions of affect within the discussions. As a test of our ability to have the computer monitor the conversation, the expressions of affect were applied toward the task of using dialogue features to identify the gender of the participant.

GEEN 165 corresponds to the CS2 (second semester) class in the ACM/IEEE curriculum (ACM/IEEE, 2013). The historical success rate for students attempting GEEN 165 is low. From 2003 to 2012, comprising about 1000 student-semesters, approximately 66% of students succeeded well enough (grade C or better) on the first attempt to continue to the next class. The fact that so many students have difficulty makes it potentially a fertile class for experimenting with educational innovation.

Lab-based computer programming classes traditionally permit unstructured group interaction. Students can talk to each other even as they require the students to write their own software. Therefore problem-solving discussions, where students respond to each other in normal dialogue fashion, are a natural addition to the lab component of a computer programming class.

NC A&T has migrated to an objects-later curriculum, meaning that CS2 contains more object concepts than the first semester CS1 class. The student exercises in this intervention are thus oriented toward object concepts.

Expressions of affect have three potential uses for this project. One is they are indications of emotional states that may effect student enthusiasm, self-efficacy and satisfaction. Another is they will be used in studies of group interaction. Finally, they may detectable by machine, contributing to an instructor's dashboard or other assessment of how well the group discussions are working.

Background

COMPS Dialogue Platform and Exercises

COMPS is a web-delivered computer-mediated chat environment (Kim et al., 2013). It permits the instructor (or a TA) to monitor each conversation. The dialogue data from this study comes from log files. Attesting to the interactivity of the COMPS experience, about half of all typing occurs while several students are typing. Even three students at a time can be typing and responding to each other, all contributing to the same discussion, since they can see each other's keystrokes in real time. In spoken conversation productive dialogue does not happen when three people are talking at once, but we have shown that in the chat domain it indeed occurs (Glass et al., 2015).

The exercises in this project involve students solving multiple-choice questions. When implementing these as group collaborations, we pay attention to three principles that promote successful collaborative learning: a structure or activity script for the students to follow, creative interdependence, and individual accountability (Eberly Center, 2016). The activity is structured as follows. The students are instructed to come to consensus on the answer, then have one student approach the instructor or a TA to verify the answer. That student is responsible for bringing the correct answer (or a hint) back to the group, and they must reach consensus again. Creative interdependence means that students should need each other to complete the exercise, it should not reasonable for one or several students to race ahead and finish it and leave the others behind or let them not participate. During the discussion the obligations of discourse require that students explain themselves in the course of reaching consensus. Having conceptual knowledge as the learning goal promotes explanatory dialogue. We have examples where seemingly the weakest student serves as a metacognitive regulator, challenging or directing every reasoning step and becoming a participant in all dialogue exchanges as the other students seem to teach that weakest one (Glass et al., 2013). Individual accountability typically occurs after the group exercise, where the students have a quiz or an exercise utilizing what they have learned. Individual accountability also occurs within the discussion, as the students find themselves responsible for explaining their positions in order to reach consensus.

Addressing Student Learning

Our collaborative inquiry learning exercises are in line with current practices in Computer Supported Collaborative Learning. A key concept is group cognition, where different participants in a conversation contribute different parts of the epistemic knowledge construction task. The Virtual Math Teams project, where students solve math problems through computer-mediated chat, has documented this phenomenon (Stahl, 2009). Learning through group cognition is justified both in terms of learning outcomes and student motivation. There is also research specifically showing that collaborative activity is a desirable pedagogical approach for "relational understanding" or understanding of concepts (Tchounikine et al., 2010). Dialogue that engages in domain reasoning, such as explaining, negotiating, or inferring is observed in these kinds of exercises (Zhou, 2009; Stahl, 2004).

The implication for COMPS technology is that monitoring the health of student conversations could be informed by a) whether students are talking to each other, b) whether they are engaging in reasoning activities.

Addressing Student Interest and Self-Efficacy

Group exercises address many of the components of student interest. Interest refers to an individual's psychological inclination to participate in particular content over time (Hidi & Renninger, 2006). There is a relationship between interest, achievement goals, performance and retention (Harackiewicz et al., 2008). Interest plays a critical role in students' further decisions on engaging and reengaging in the major (Brown, 2012). The four-phase model of interest posits four sequential interest phases: triggered situational interest ("catching"), maintained situational interest ("holding"), emerging individual interest, and well-developed individual interest. Mitchell (1993), 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.

Recently Kim and Schallert (2014) have investigated the mediating effect interpersonal interactions have on student interest. It is possible to track student interest in four developmental phases throughout a semester, not just within the time frame of individual activities. It is affected not only by the enthusiasm expressed by the teacher and fellow students, but also by factors such as affiliative motivations: the desire to belong to the group. The social factors enhancing interest were found within college classes in a number of diverse disciplines (e.g. history, chemistry, religion) in both upper and lower level college classes.

Viewed in this light, group exercises should address student motivation issues through social interaction at the same time as they address learning of concepts through group cognition. The exercises are constructed so that students engage with other students, providing the small-group interpersonal contact that best 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.

The implication for COMPS technology is that monitoring the health of student conversations could be informed by expressions of student affect. Affect, the observable manifestation of emotion, mediates social interaction and is related to student interest.

Self-efficacy, an individual's belief to be capable of performing a particular task (Bandura, 1977), has been widely studied because of its relationship to performance including academic achievement (Choi, 2005; Pajares and Miller, 1995; Wood and Locke, 1987) and even choice of major in college (Hackett, 1985). In accordance with the suggestions of Finney and Schraw (2003), we measured self-efficacy using task-specific survey items rather than generalized questions. This project measures students' self-efficacy both at the level of the skills in individual assignments at the time of the COMPS exercises and overall in the topics of the class at the beginning and end of the semester.

Data and Methods

We have collected data from one semester of the GEEN 165 class. There were 55 students at the start of the semester and 47 at the end. We administered COMPS exercises four times during the semester, with 53 group discussions in total. Most groups had 3 or 4 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. We deviated from this protocol by creating a few all female groups, for comparison with the all-male groups. Altogether there were about 8000 dialogue turns. Students were surveyed near the beginning and end of the semester regarding their enthusiasm for the class, their self-efficacy in programming, and their desire to continue. Every COMPS exercise was also accompanied by a survey of the student experiences.

Transcript Processing and Annotation

Table 1 contains an extract from a COMPS discussion. From COMPS log files we extract dialogue turns in spreadsheet format for processing. The text from one dialogue turn is in one line of the spreadsheet. In addition to the metadata such as problem number, turn number, and time stamp, each dialogue turn is tagged with features. Some are derived by software and some are annotated by hand. These features are available for machine learning experiments and for human analysis and study of dialogues. The machine-derived classifiers are available for feeding software that will monitor the health of the conversation.

Some of the existing machine-derived features (Glass et al., 2014b) that have been relevant to transcript studies and machine monitoring of the health of the conversation are:

- The presence of discourse marker words, e.g. “now” or “therefore” near the beginning of a dialogue turn. These are linguistically associated with reasoning, and are therefore possibly indicative of productive discussion.
- The presence of pronouns that include another participant in the dialogue: “you,” “we,” “us.” These are possibly indicative of transactive discussion.
- The presence of question marks.
- The presence of emoticons. It is possible that emoticons are associated with students attending to each others affect.
- The length of a turn in words.
- Whether typing this turn overlapped with other people typing.

Affective States Evinced in Dialogue

Of particular interest are six affective states that we have chosen as initial targets. These are annotated by hand. They were chosen because they may be salient for monitoring both the learning aspects (whether the students are reasoning together) and the social health of the conversation. We show here some of the definitions that the coders have applied for consistency in recognizing and coding.

- Excited.
- Apologetic. Refers to a user expressing regret for previous action. This type of message is usually aimed towards another user or towards the group as a whole.
- Humor.
- Frustrated.
- Confused. User explicitly expressing confusion, or exhibiting confusion e.g. through questions.
- Sad. A negative emotion determined by keywords and sad emoticons that are usually directed at self.

Some of these affective states have been tagged and illustrated in the Table 1 dialogue. Table 2 indicates some of the textual indications for the various states. These are being used by the coders at present, but will become machine-derived features for the purpose of machine-annotating the affective states.

Surveys

The survey administered to all students at the beginning and end of the semester has an interest part and a self-efficacy part. The end-of-semester survey also inquires about student plans for continuing in the major and registering for the next programming class. All items use a 6-point scale. The interest survey items are derived from a survey from Harackiewicz et al. (2008). One of the authors of this paper has utilized these items to assess how much a student's interest in a class is affected by the enthusiasm of fellow students (Kim and Schallert, 2014). Some representative items are “What we are learning in GEEN165 this year can be applied to real life” and “To be honest, I don’t find what we do in the GEEN165 class interesting.” The self-efficacy items inquire about student confidence in completing 13 tasks corresponding to class topics. This list was obtained from the instructor. A typical item is “Design inner classes that implement event handling interfaces.”

The after-COMPS-lab survey had items covering student perceptions in three areas: student interest, whether the student learned from the lab, and how well the group exercise functioned. An example item is “I contributed to the understanding of other students in my group.”

Results

Survey Results

Table 3 shows the students' perceptions of interest and efficacy at the beginning and end of the semester. All interest items were combined into one mean and the same for all efficacy items. In total 28 students participated in both pre- and post-surveys.

- Regarding students’ interest toward the course, their interest did not change. The averages of stu-

dents' interest toward the course in the beginning of the semester and end of the semester were 4.33 and 4.32 respectively.
- Self-efficacy with respect to the course content indicated significant improvement between the beginning and the end of the semester, rising from 2.83 to 3.81.

The increase in self efficacy was significant, $p < 0.01$.

Table 4 shows students' perception of the COMPS labs, surveyed immediately after each lab. There seemed to be a clear improvement between the first part of the semester (Labs 1 and 2) and the later part (Labs 3 and 4). Students perceived:

- more effective group work in the second part (means rose from about 3.1 to about 3.4)
- better understanding of concepts in the second half (means rose from about 3.4 to about 3.9).

Multiple one-way ANOVA supports the hypothesis that mean scores are indeed different, $p = 0.03$ 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.

However, students' interest in each exercise in the lab sessions seemed to fluctuate throughout the semester. Lab 3 had the highest interest, which corresponded with the highest effectiveness and understanding. But interest in Lab 4 was the approximately the same as Labs 1 and 2.

Affective States by Gender

We annotated the 14 group discussions of one COMPS exercise, comprising 2147 dialogue turns, for the six affective features. In total 199 turns showed evidence of one or more feature, or 9.3%.

As a first test of the utility of these features along with the machine-generated ones, we tried to use them to predict the gender of the participant. Among 49 students we had 16 women and 33 men. First we aggregated all the turns from each student, and looked at statistical differences between the two populations. Two-tailed t-tests revealed that none of the features were significantly different between the genders at the $p < 0.05$ level. However expressions of apology were different at the $p = 0.06$ level. The most common affective feature was confusion, with 62 instances of utterances expressing confusion. Women expressed confusion in 4.6% of turns, and men in 2.2%. It suggests the two genders behave differently, but the $p < 0.22$ level does not show significance. The two genders also showed differences in the amount of participation. Men each uttered an average of 46 turns per dialogue and women 36 turns.

We then trained a J48 decision tree classifier and a multiple-regression linear classifier using the Weka data mining tool (Witten and Frank, 2005). The task is to classify each dialogue turn with the gender of the speaker, to mimic the task of monitoring a conversation in real time turn-by-turn. These classifiers have not been successful. The same features that are statistically correlated with gender are discovered by the decision trees, but accuracy has been quite low.

Discussion and Future work

Survey Results

The students experienced improvement in their experience of the COMPS exercises during the semester. They reported that the groups worked better in the last two exercises and that they learned more. It is not clear why student interest was lower in the last lab. Anecdotally there are two reasons that have been suggested by the instructor and lab TAs who supervised this session. One is that the last lab was optional, presented during Thanksgiving week. That fewer students attended could indicate that the general level of engagement was lower than usual. The other is that perhaps the novelty was wearing off. Some students expressed as much during the session. We will need to find some way to survey the reasons for student interest.

The pre- and post-semester survey is hard to interpret because of low participation rate and dropouts. In the next semester we are enforcing better participation. The increase in self-efficacy was striking, but we do not have yet any comparison with other classes. Future work includes comparing interest and self-efficacy with learning gains on the pre- and post-tests. Future work also includes comparing pre- and post-semester survey results with individual lab surveys, to see whether there are correlations between overall student interest and the situational interest in individual COMPS exercises.

Another analysis in the future will be between the participants of the same group: do they agree about learning and group functioning, do they have similar learning gains.

Affective States in Dialogue

Hand-annotating the remaining 6000 turns of dialogue may result in more reliable statistical correlations. We are also at work toward machine-annotation of these features.

Annotation of the affective states so far has relied entirely on the text of the dialogues. Future work will include extra-linguistic features. In COMPS group exercises in other classes evidence of student engagement sometimes presents through Comic Sans typeface, big or bold fonts, and wild colors. We are also exploring using timing features from the overlapped typing. Students can all type simultaneously while seeing each other's developing chat text (Glass et al., 2015). We think that typing speed, degree of simultaneous typing, and pauses as they look at each other's turns, may provide indications of affective states such as being excited or indications of when they are attending to each other's utterances.

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

ACM/IEEE. Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. New York: ACM, 2013.

Bandura, A. (1997). Self-efficacy: the exercise of control. New York: Freeman.

Brown, J. (2012). Developing a freshman orientation survey to improve student retention within a college. *College Student Journal*, 46, pp. 83−851.

Choi, N. (2005). Self-efficacy and self-concept as predictors of college students' academic performance. *Psychology in the Schools*, 42, pp. 197−204.

Eberly Center (2016). 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/group-projects/design.html. Retrieved March, 2016.

Finney, S. and G. Schraw (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, 28, pp. 161−186.

Glass, Michael, Melissa Desjarlais, Jung Hee Kim, and Kelvin Bryant. (2013). COMPS Computer-Mediated Problem Solving Dialogues (poster abstract). International Conference on Computer Supported Collaborative Learning, Madison, WI.

Glass, Michael, Jung Hee Kim, Kelvin Bryant, Melissa Desjarlais, Micayla Goodrum, and Thomas Martin. (2014a). Toward Measurement of Conversational Interactivity in COMPS Computer Mediated Problem Solving. Proceedings of the 25th Modern Artificial Intelligence and Cognitive Science Conference (MAICS-14), Spokane, WA.

Glass, Michael, Jung Hee Kim, Kelvin Bryant, and Melissa Desjarlais. (2014b). Indicators of Conversational Interactivity in COMPS Problem-Solving Dialogues, Third Workshop on Intelligent Support for Learning in Groups (ISLG) at Twelfth International Conference on Intelligent Tutoring Systems, Honolulu, Hawaii, June.

Glass, Michael, Jung Hee Kim, Kelvin Bryant, and Melissa Desjarlais (2015). Come Let Us Chat Together: Simultaneous Typed-Chat in Computer-Supported Collaborative Dialogue. *Journal of Computing Sciences in Colleges*, 31 no. 2, Dec.

Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: a path analysis. *Journal of Counseling Psychology*, 32, pp. 47–56.

Harackiewicz, J. M., A.M. Durik, K.E. Barron, L. Linnenbrink-Garcia, and J.M. Tauer. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology*, 100, pp. 105−122.

Hidi, S., and K. Renninger. (2006). The four-phase model of interest development. *Educational Psychologist*, 41, pp. 111−127.

Kim, Jung Hee, Melissa Desjarlais, Kelvin Bryant, and Michael Glass. (2013). Observations of Collaborative Behivor in COMPS Computer-Mediated Problem Solving, Proceedings of the 24th Midwest AI and Cognitive Science Symposium, New Albany, IN.

Kim, Taehee, and Diane L. Schallert. (2014). Mediating effects of teacher enthusiasm and peer enthusiasm on students' interest in the college classroom. *Contemporary Educational Psychology,* 39, no. 2 pp. 134−144.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology*, 85, pp. 424−436.

Pajares, F., and M. D. Miller. (1995). Mathematics self-efficacy and mathematics outcomes: The need for specificity of assessment. J*ournal of Counseling Psychology*, 42, pp. 190–198.

Stahl, Gerry (2004). Building collaborative knowing: Elements of a social theory of CSCL. In J. W. Strijbos, P. Kirschner and R. Martens, eds., What we know about CSCL and implementing it in higher education, Boston, MA: Kluwer Academic Publishers, pp. 53−86.

Stahl, Gerry, ed. (2009). Studying Virtual Math Teams, Springer, pp. 57−73.

Tchounikine Pierre, Nikol Rummel, and Bruce M. McLaren. (2010). Computer Supported Collaborative Learning and Intelligent Tutoring Systems. In R. Nkambo, J.Bourdeau, & R. Mizoguchi (Eds.) Advances in Intelligent Tutoring Systems. Chapter 22, pp. 447−463, Springer.

Witten, Ian H. and Eibe Frank. (2005). Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufmann.

Wood, R.E. and E.A. Locke. (1987). The relation of self-efficacy and grade goals to academic performance. *Educational and Psychological Measuremen*t, 47, pp. 1013–1024.

Zhou, Nan (2009). Question Co-Construction in VMT Chats. In Stahl, 2009, pp. 141−159.

Table 1: Example of Dialogue Transcript with Affective Features

Turn	**Student**	**Time**	**Dialogue turn**	**Affective State**
1	A	06:44.2	f and foo are the refernece variables	
2	A	07:05.2	so those together make 16? for the refrence types	
3	B	07:11.9	yup yup	
4	A	07:27.9	16 bytes	
5	C	07:30.2	2a = 20	
6	C	07:36.0	**:D**	**Excited**
7	B	07:39.7	**there ya go lol**	**Humor**
8	D	07:54.9	**Wait where did you get 16?**	**Confused**
9	D	08:05.8	wouldnt it be 48 at least for the main method	
10	D	08:18.3	because the array creates 5 object	
11	A	08:26.1	**oh yeah i looked over that** was just counting m f and foo	**Apologetic**
12	C	08:28.7	those are on the heap not the stack	
13	D	08:48.0	So the objects created by an array are on the heap	
14	A	09:13.8	yeah run time stack = 48	

Table 2: Example of Feature words

Excited	**Apologetic**	**Confused**	**Frustrated**	**Sad**
:D	sorry	i'm confused	D:<	:(
yay	my bad	how	):<	):
yes!	nvm	why	This is hard	I feel stupid
!!!	whoops	what is		
cool!	i messed up	I don't under-stand		

Table 3: Beginning and end of semester surveys

Time	**Interest**	**Efficacy**
beginning of sem.	4.33 / 5	2.83 / 5
ending of sem.	4.32 / 5	3.81 / 5

Table 4: After lab surveys

	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