GroupWork: Learning During Collaborative Assessment Activities

William T. Tarimo, Brandeis University, wtarimo@brandeis.edu
Timothy J. Hickey, Brandeis University, tjhickey@brandeis.edu

Abstract: The positive benefits of collaborative and cooperative learning are well documented in educational literature and research. In this paper, we present GroupWork - a group-based assessment tool within TeachBack, and an evaluation of our experience in using it as part of an interactive flipped classroom. GroupWork was used to facilitate real-time group-based formative assessment and other problem-based activities in the classroom. Specifically, we examine the short-term and long-term impact of this implementation on students' learning outcomes and experiences compared to performing formative assessments where students work individually. Results show significant performance improvement for each student and for each assessment question when students are allowed to discuss their solutions with peers. The formative assessment grades after collaborative problem solving are shown to have a strong correlation with final course grades, suggesting that collaborative problem solving may have a lasting benefit.

Introduction

This paper presents a case study and the results of an experiment where students collaborated in small groups during in-class formative assessment activities. The experiment studies whether collaboration during formative assessment activities has better learning outcomes for the participants compared to when they work individually. In assessing the immediate effects on the learning outcomes, we compared students' overall performance when responding individually to formative assessment problems and their corresponding performance when working in collaborative groups. In assessing the impact of collaboration on overall learning outcomes on the course, we looked for evidence of correlations between performance in these assessment activities and final performance in the course. We also surveyed students at the end of the course to get their perspectives on their learning and experiences during group-based formative assessments. All the assessment activities were done using the GroupWork tool within TeachBack (Hickey & Tarimo, 2014; Tarimo, Deeb, & Hickey, 2015; Tarimo, Deeb, & Hickey, 2016; Tarimo & Hickey, 2016; Tarimo, 2016) and the majority of the activities were real-time formative assessments during class time. TeachBack is an in-class web application that offers various interactive tools for active and interactive learning and teaching in the classroom.

Over the last three years of empirical studies using TeachBack, the Questions feature (a lightweight *clicker* functionality) has been used to facilitate assessment activities in the classroom where students respond individually followed by class-wide discussions. For example, a previous study (Tarimo, Deeb, & Hickey, 2016) has shown statistically significant correlation between levels of participation and performance in in-class formative assessments and overall mastery of the material and concepts in a course. In this experiment, we wanted to see whether adding collaboration to these activities would lead to better learning outcomes and experiences for the students.

Conducting this experiment would also inform us on the feasibility and practicality of using web based interactive tools like GroupWork in facilitating collaborative and cooperative learning and assessment activities in classroom settings. The findings would also provide the evidence necessary to propose whether pedagogy design should include collaborative and cooperative activities.

Literature review

The proposed implementation of collaborative formative assessment using GroupWork is based on various instructional activities and learning theories that empirical research and educational literature have shown to support improved learning and teaching outcomes.

First, is the opportunity for students to work collaboratively and cooperatively during learning. Presently, competitive and individualistic (Roger & Johnson, 1988) are the more dominant modes of student-student interactions during learning. However, compared to competition and individualism, cooperation and collaboration have been shown to result in students achieving better learning outcomes, more effectiveness interpersonally and students developing positive attitudes towards learning, each other, instructors and subject areas. This practice is also referred to as Peer Instruction (PI), (Crouch & Mazur, 2001; Crouch, Watkins, Fagen, & Mazur, 2007), where class time includes assessment and learning activities where students work individually and in groups.

Second, having activities that allow a group of students to work together stresses the idea of co-construction of knowledge and mutual engagement of participants. A constructivist classroom focuses on student-centered discourse, over the typical teacher centered classroom, where students drive discussions and the teacher serves as a guide on the side (Palincsar, 2005).

Third, our extended use of formative assessment activities is a form of problem-based-learning based on cognitive apprenticeship (Collins, Brown & Newman, 1987) where students learn in the context of solving complex and meaningful problems. The role of instructor is to create appropriate questions, and then guide students on the learning process by encouraging them to think deeply through discussions in evaluating responses to the problems.

And finally, GroupWork is closely modeled on the Think-Pair-Share methodology of classroom-based collaborative active learning (Lyman, 1981; Lyman, 1987; McTighe & Lyman, 1988; Kagan, 1989). In a typical Think-Pair-Share activity, students work on a problem posed by the instructor, first individually, then in pairs or small groups, and finally all participate in a class-wide discussion. Some of the benefits of this strategy include promoting engagement, allowing students to express their reasoning, reflect on thinking, and obtain immediate feedback on their understanding and misconceptions (Kothiyal, Majumdar, Murthy & Iyer, 2013). Moreover, the Think-Pair-Share technique is recommended as an instructional activity that engages learners in higher-order thinking, and as a feedback mechanism, both for students and teachers (Cooper & Robinson, 2000).

Experimental design

The class

The experiment was carried out during the fall semester of 2015 in a flipped advanced level Computer Graphics course with enrollment open to undergraduate and graduate students. The course exposes students through a hands-on introduction to the science and practice of rendering three dimensional (3D) images using both resource intensive ray-tracing methods and real-time techniques using the GPU. The course was also intensive in mathematics and programming content and exercises in Java and JavaScript. The course was a typical *flipped* classroom where class meetings started with pre-class reading assignments and reflections, and class times were devoted to interactive teaching and learning activities.

During that semester, 41 students completed the course which consisted of 26 interactive class meetings, with each class lasting for 80 minutes and meeting twice in a week. TeachBack was used in the class to facilitate some of the instructional activities and interactions in the classroom, and this required all students to bring laptops or tablets to class. This included formative assessments, students giving feedback to the instructor, the back-channel assistance forum for asking and answering questions as well as discussions on class materials, and taking notes by students.

Collaborative assessment using GroupWork

The majority of the exercises were given one-by-one in real-time during classes as a way to assess students' understanding and to inform the instructor on whether to reiterate the just covered concepts or move on to new material. The exercises also counted as an active activity during classes in getting students to actively engage with the material, gauge their own learning, reveal misconceptions, and more importantly engage in discussions and share the learning with others in the class.

Answering questions in GroupWork follows three steps. During the first step, each student takes the time to think and attempt the problem individually and then submit a personal answer which cannot be changed. During the second step, group members share their personal answers and have a collaborative discussion where the team attempts to improve on their initial answers and perhaps reach an agreement on better responses. Groups are made of three students, and they are created using a random assignment of members for each class meeting based on attendance for the day. During the group discussions, the team members can each create one or more new answers based on the level of agreement in the group. Afterwards, each group member is required to submit a final answer by voting on any of the answers from the team. Students are not forced to agree on a single group answer, and that's why a voting system is used, requiring each team member to vote/indicate a final answer. This strategy is intended to encourage a sense of individual accountability and autonomy as shown in cooperative learning theories.

Figure 1 shows a students' view of a question in GroupWork with access to answers panel and the group mini-forum. The interface is designed to include a group discussion area for each question so that group members can still have a discussion in the cases where members are not physically seating next to each other or some members are attending class remotely. In a related study reported in (Tarimo & Hickey, 2016), we studied the effects of offering optional remote attendance to class meetings where students are required to follow along

with a live-streaming of the class and participate in all class activities (including the group activities reported in this paper) using TeachBack. This hybrid style was practical and feasible; however, we did not implement an entirely online mode.
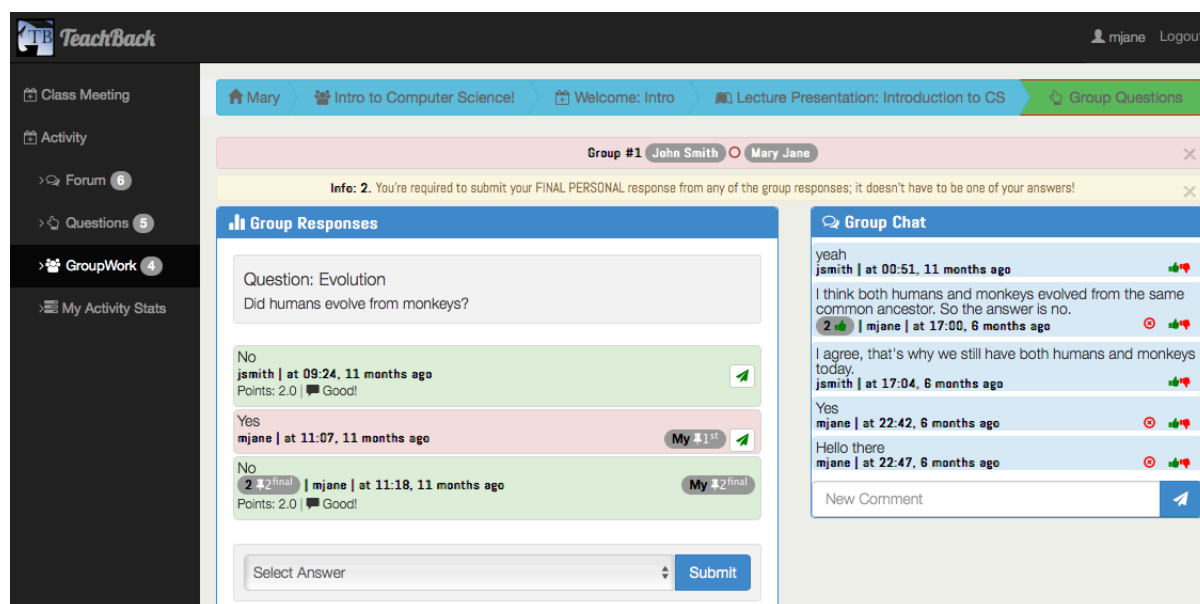

Figure 1. TeachBack: Students' question view in GroupWork.

An important design feature in GroupWork is making sure that group members attempt to answer questions individually before engaging with other group members, or see each other's attempts. In order to achieve this, the interface is designed such that a group member cannot see answers or messages from the rest of the group until that student has submitted his/her individual answer. This design closely models the Think-Pair-Share methodology.
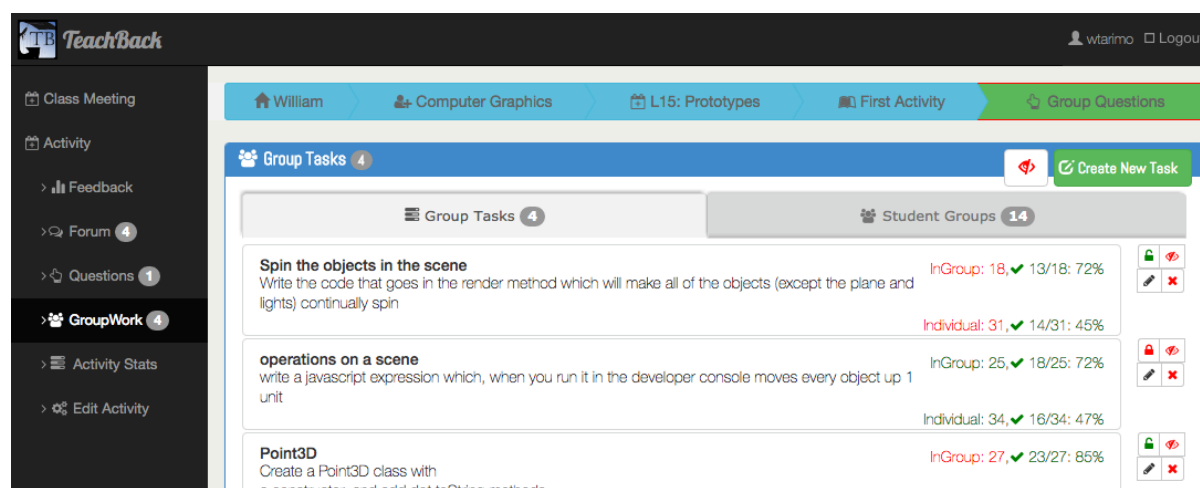

Figure 2. The main GroupWork page listing questions with participation and grading summaries.

In the main GroupWork page, shown in Figure 2, each listed question displays the response counts and grading summary for individual and in-group answers. During class, the counts update in real-time as students respond to questions, this allows instructors to more easily gauge and manage the progress of the assessment. During the course reported in this experiment, an instructor would review the answers and a teaching assistant (TA) would do the actual grading at the same time.

The final stage in a GroupWork exercise is a class-wide discussion that is led by the instructor to go over responses to questions. The interface is designed such that instructors can view and grade both types of answers, see Figure 3. After a question is graded, the percentage of correct answers is also displayed for the

individual and group answers. As seen in Figure 2, this detail provides a convenient way to see and compare the students' performance when responding individually and in groups.

Data collection

The first analysis involved looking for the evidence to support whether students perform better in formative assessment activities when they work collaboratively in small groups compared to when they work individually. For this, we used the grading data from GroupWork for all graded problems that were given throughout the semester. Each individual and in-group problem response was graded using the same rubric. The correctness of each answer was assessed on the scale of either being fully correct (2.0 points), partially correct (1.0 points) or fully incorrect (0 points), this can be seen in Figure 3.
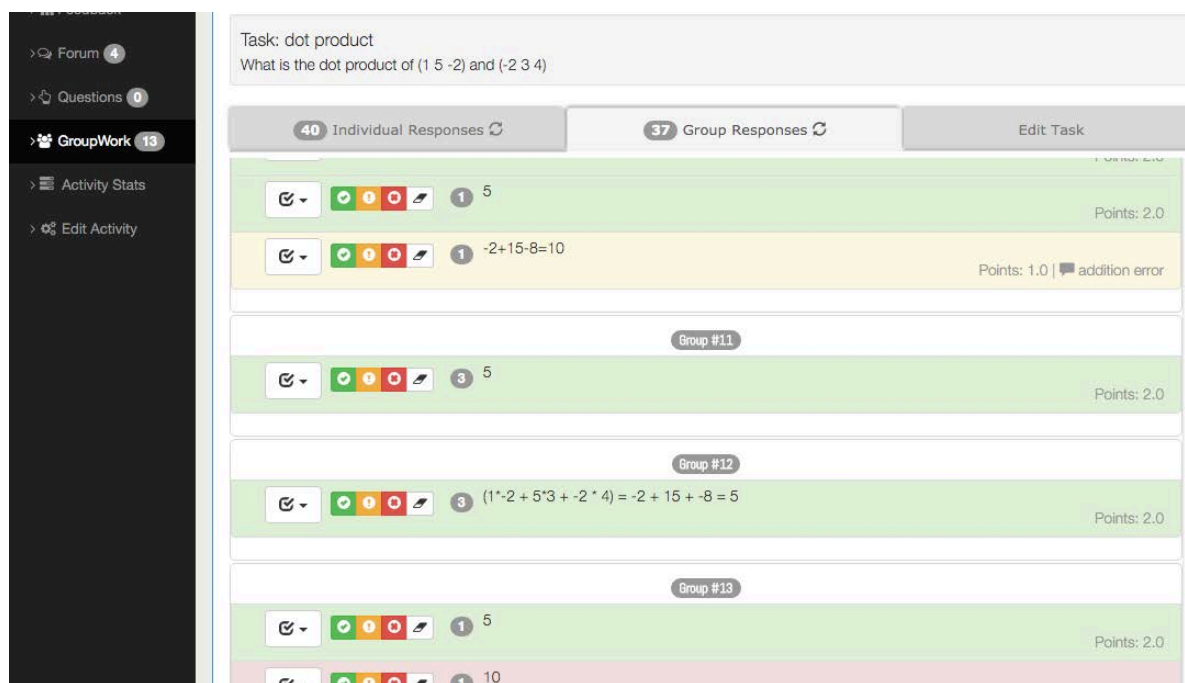

Figure 3. A GroupWork page showing group responses to a question.

The second analysis looked for evidence of correlation between performance in collaborative formative assessments and overall mastery of concepts covered in the course. Overall learning was measured as the course grade earned at the end of the semester which included cumulative grades from weekly or bi-weekly small programming assignments, three major programming projects, and a final written exam. The final exam was a cumulative exam which was administered at the end of the course and covered most of the materials taught in the course. 20% of the official course grade accounted for class participation (based on participation in TeachBack activities and pre-class readings), however, we omitted this participation component in the course grade that was used in our analysis as it didn't exactly reflect mastery of concepts.

At the end of the course a survey was given to students in order to gather their perspectives and evaluation of their learning experiences from the use of collaborative formative assessment and how it was implemented using the GroupWork feature of TeachBack.
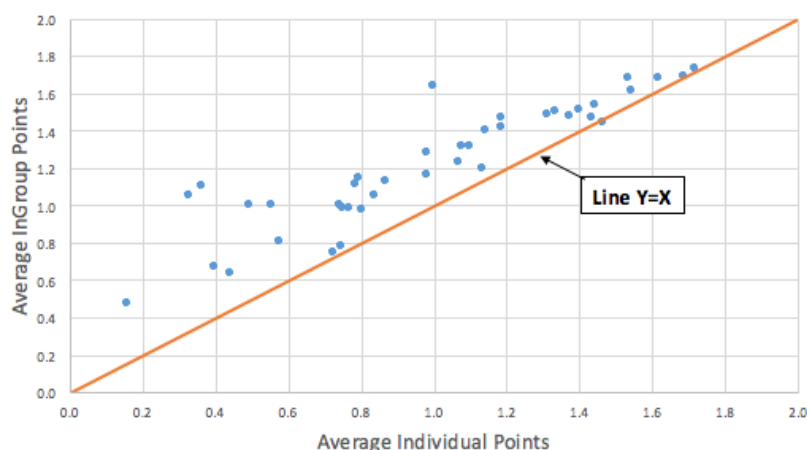
Results

### Learning during collaborative formative assessments

For each student, we looked at the average performance when the student responded individually and when the student responded after the group collaboration stage. Average performance was calculated as the average number of points earned; that is, the sum of points earned divided by the number of questions attempted.

Figure 4 shows a scatter plot of average points earned by each student when responding individually versus when working in groups. The plot demonstrates that, on average, most students tend to earn more points as a result of group collaboration, and this is true even for students who score highly during the individual part.

We used a two-tailed *t test* to compare the mean of the average number of points from the individual and in-group populations. The class average in points when working individually is 0.995 points per question, and the average while working in collaboration is 1.223. The improvement in the mean of 0.228 is statistically significant with a *P value* of less than 0.0001, with a 95% confidence interval of 0.170 to 0.286. This improvement represents an effect size of 0.618, which is calculated as a Cohen's d Effect Size (Cohen, 1988). According to Cohen, an effect size of 0.618 is just above the 'medium' range. And according to a survey of academic research on instructional interventions, an effect size of 0.5 or higher would represent significant gains as it is higher than the results found in most instructional interventions (Albanese, 2000; Dubin & Taveggia, 1968).

Figure 4. A scatter plot of average individual points vs average in-group points for all students.
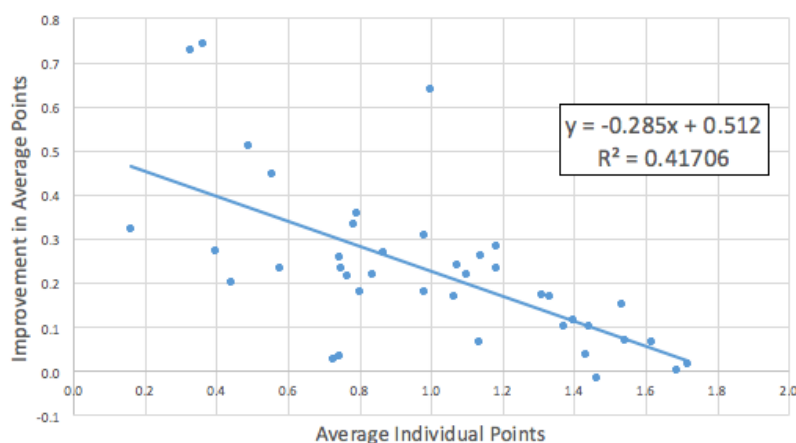
Figure 5. Average improvement in points from individual to in-group responses.

Looking at the improvement in performance resulting from collaboration, Figure 5 shows that students who tend to do poorly when working individually are the ones who take the most advantage of group collaboration. For example, a student with an average individual performance of 0.85 (out of 2.0) points improved performance by 0.28 points, whereas a student with individual performance of 1.7 points only improved by 0.02 points. This is contrary to, for instance, the hypothesis that collaboration would affect all students in the same way.

Secondly, for each GroupWork problem, we compared the average class performance from individual responses and in-group responses. Figure 6 shows a scatter plot of average points earned in each GroupWork question when students responded individually versus when they responded within groups. In only one of the 60 questions asked we see no average improvement in students' performance when they worked in groups.

Therefore, regardless of the question type or difficulty, students' performance for each question improved as a result of collaboration compared to working individually. Again, a two-tailed *t test* was used to compare the means of the average number of points for the two populations representing individual and in-group performances. The mean of average number of points earned increased by 0.233 from 1.003 points per

question to 1.236, and the *t-test test* showed this difference to be statistically significant with a *P value* of less than 0.0001. The 95% confidence interval is 0.190 to 0.276. The corresponding *effect size* of the improvement is calculated as 0.609, which is again considered an impressive gain from an instructional intervention.
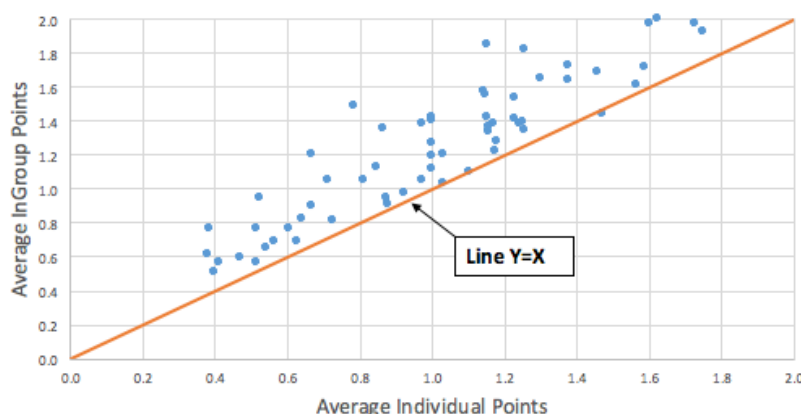

Figure 6. A scatter plot of average individual points vs average in-group points for all GroupWork questions.

Correlation with overall learning in the course

In this analysis, we looked for evidence of correlation between performance in collaborative formative assessments and overall mastery of concepts covered in the course. Mastery of the material was measured as the course grade earned at the end of the semester which included cumulative grades from weekly or bi-weekly small programming assignments, three major design and programming projects, and the final written exam. Linear regression was used to test the hypothesis that performance in collaborative formative assessment could predict the learning outcomes in the course. Separate regression analyses were used to study the correlation between the final course grade and each of the individual and in-group average points in GroupWork problems. The results of the analyses are shown in Table 1. Both regression models indicate statistically significant, (P < 0.0001), positive correlation between course grade and each of the individual and collaborative performances in formative assessment. Indeed, the correlation coefficients are 0.647 and 0.568 between course grade & average individual points, and course grade & average in-group points, respectively.

Table 1: Linear Regression Analysis: Course Grade vs. Average Individual Points and Course Grade vs. Average In-Group Points

Average Points	R^2	P-Value	Coefficient
Individually	0.4188	< 0.0001	13.074
In Group	0.3232	0.0001	14.429

Figure 7 shows a visual representation of the regression analysis using scatter plots. On the left is a scatter plot of course grade versus average individual points, and on the right, is a scatter plot of the course grade versus average in-group points. And even though both models are statistically significant, performance while working individually is a slightly stronger predictor of course performance compared to performance when working in groups. This observation is still consistent with an earlier study which found significant positive correlation between performance in a course and performance in formative assessment activities where students worked individually. For this class, final course grade is comprised of summative assessments in which students worked individually, this explains why there is a stronger correlation between course grade and individual performance in the GroupWork exercises compared to in-group performance. Moreover, despite the fact that each group member is required to submit a personal answer after group discussions, the group discussions and collaboration has a significant influence on subsequent group answers which in turn averages down the differences in mastery levels among group members.

Students' opinions

Students were given a survey at the end of the course in order to get their learning experiences and evaluation of GroupWork and its use in facilitating formative assessment exercises. 37 of the 41 students participated in the survey. Using a range of 1 (not useful) to 5 (very useful), students were asked to assess the extent to which the use of GroupWork based collaborative assessment activities was useful towards their learning. 80.5% of the students responded with a rating of 3 or above, with 55.5% believing that it was useful with a rating of 4 - 5. Moreover, using a scale of 1 (not satisfied) to 5 (very satisfied), 75% of the respondents reported being satisfied with how GroupWork tool was used in the course with a rating of 3 - 5. 44.4% of the respondents reported a satisfaction rating of 4 - 5.
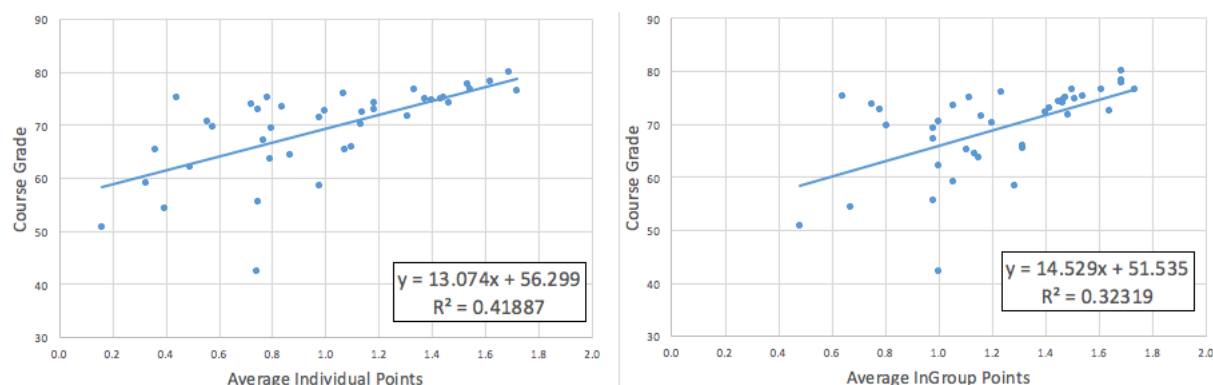

Figure 7. A scatter plots of course grade vs. average individual points, and course grade vs. average in-group points for all students.

Among the aspects of the implementation that students liked were how it forces them to start by attempting problems individually before engaging with a group in discussion and collaboration. The requirement to attempt problems individually was commented to be especially useful in tackling complex concepts as it revealed various personal misconceptions and confusions that would then be corrected during group and class discussions. Moreover, enforcing individual attempts results in many alternative answers to problems which would then be refined to fewer in-group answers. Students pointed out that having access to various alternative answers to a problem opens a new opportunity to learn from seeing how others think and approach answering questions.

Students also appreciated the introduction of group-based formative assessment using GroupWork as it "provided the opportunity for collaborative activities in class". GroupWork activities were part of most of the class meetings during the semester, and as a result, student collaboration and formal discussions were part of the class routine. This introduced the component of collaboration to the pedagogy that would otherwise not be there as in traditional classrooms. Working with ever changing groups of classmates was pointed out by students in the survey as enabling them to "get to know other(s)" in the class.

Conclusions

The findings of this experiment demonstrate the benefits of cooperative, collaborative and problem-based learning in group-based formative assessment activities. The data analysis compared students' performance in formative assessment activities when they worked entirely individually and when they worked in small groups using the Think-Pair-Share methodology. Results showed that students reached a significantly higher level of achievement when formative assessment activities included collaboration. In general, all students performed better when they worked in groups than when they worked individually alone. Similarly, students' performance was higher for each assessment problem when students worked in collaboration.

Furthermore, the results show that students benefit in various amounts from collaboration based on individual student's mastery levels. Even though all students gather benefits from collaboration, it is the students who would usually do poorly when working individually who benefit the most. This sounds like a trivial outcome but in this case the benefit comes from having heterogeneous groups where groups are randomly generated for each class meeting. When it comes to choosing students assignment into groups based on individual capabilities, heterogeneous groups tend to be more powerful than homogeneous groups (Roger & Johnson, 1988). This is because in collaboration learning comes from discussion (presentation and arguing of

different perspectives and alternatives), explanations (to other group members, especially those in need of help), justification of answers, and shared resolutions.

The results demonstrated positive correlation between overall performance in the course and students' average performance in formative assessments when working individually and in groups. Performance working individually has a slightly stronger correlation compared to in-group performance. And since we were interested in studying the impact of collaboration on overall learning, a better study approach would have been comparing course performances between experimental and control populations of students, one with collaborative formative assessment and the other with only the traditional individualistic formative assessment.

In this work, we have shown an implementation of collaborative formative assessment that is practical and feasible. Students thought the implementation of was beneficial and useful towards their learning. Students were generally satisfied with the technical implementation of GroupWork even though they had recommendations for improvement. The demonstrated benefits to the teaching and learning outcomes imply that this instructional approach is worthy of further adoption, investigation and testing in various subject areas and curriculum designs.

References

Albanese, M. (2000). Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills. *Medical education*, 34(9), 729-738.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. *Hillsdale, NJ*, 20-26.

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children*, 8(1), 2-10.

Cooper, J. L., & Robinson, P. (2000). Getting started: Informal small-group strategies in large classes. *New directions for teaching and learning, 2000*(81), 17-24.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9), 970-977.

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-based reform of university physics*, 1(1), 40-95.

Dubin, R., & Taveggia, T. C. (1969). *The teaching-learning paradox: A comparative analysis of college teaching methods*. Eugene; Center for the Advanced Study of Educational Administration, University of Oregon, 1968. 78 pp.

Hickey, T. J., & Tarimo, W. T. (2014). The affective tutor. *Journal of Computing Sciences in Colleges*, 29(6), 50-56.

Kagan, S. (1989). The structural approach to cooperative learning. *Educational leadership*, 47(4), 12-15.

Kothiyal, A., Majumdar, R., Murthy, S., & Iyer, S. (2013, August). Effect of think-pair-share in a large CS1 class: 83% sustained engagement. In *Proceedings of the ninth annual international ACM conference on International computing education research* (pp. 137-144). ACM.

Lyman, F. (1981). The responsive classroom discussion: The inclusion of all students. *Mainstreaming digest*, 109-113.

Lyman, F. (1987). Think-pair-share: An expanding teaching technique. *MAA-CIE Cooperative News*, 1(1), 1-2.

McTighe, J., & Lyman, F. T. (1988). Cueing thinking in the classroom: The promise of theory-embedded tools. *Educational Leadership*, 45(7), 18-24.

Palincsar, A. S. (2005). 12 Social constructivist perspectives on teaching and learning. *An introduction to Vygotsky*, 285.

Roger, T., & Johnson, D. W. (1988). Cooperative learning: Two heads learn better than one. *Transforming Education: Awakening the Full Human Potential... in Everyone, Context Institute, Langley, WA*, 34-36.

Tarimo, W. T., Deeb, F. A., & Hickey, T. J. (2015, May). A flipped classroom with and without computers. In *International Conference on Computer Supported Education* (pp. 333-347). Springer International Publishing.

Tarimo, W. T., Deeb, F. A., & Hickey, T. J. (2015). Computers in the CS1 Classroom. In *Proceedings of the 7th International Conference on Computer Supported Education (CSEDU) (2)*, 67–74.

Tarimo, W. T., Deeb, F. A., & Hickey, T. J. (2016). Early detection of at-risk students in CS1 using TeachBack/Spinoza. *Journal of Computing Sciences in Colleges*, 31(6), 105-111.

Tarimo, W. T., & Hickey, T. J. (2016, October). Fully integrating remote students into a traditional classroom using live-streaming and TeachBack. In *Frontiers in Education Conference (FIE), 2016 IEEE* (pp. 1-8). IEEE.

Tarimo, W. T. (2016). *Computer-Supported Agile Teaching* (Doctoral dissertation, Brandeis University).