Leveraging Collaboration to Improve Gender Equity in a Game-based Learning Environment for Middle School Computer Science

Philip Sheridan Buffum, Megan Frankosky, Kristy Elizabeth Boyer, Eric Wiebe, Bradford Mott, James Lester
Center for Educational Informatics
North Carolina State University
Raleigh, NC USA
Email: psbuffum@ncsu.edu

***Abstract*— Game-based learning environments can deliver robust learning gains and also have a unique capacity to engage students. Yet they can unintentionally disadvantage students with less prior gaming experience. This is especially concerning in computer science education, as certain underrepresented groups (such as female students) may on average have less prior experience with games. This paper presents evidence that a collaborative gameplay approach can successfully address this problem at the middle school level. In an iterative, designed-based research study, we first used an experimental pilot study to investigate the nature of collaboration in the ENGAGE game-based learning environment, and then deployed ENGAGE in a full classroom study to measure its effectiveness at serving all students. In earlier phases of the intervention, male students outpaced their female peers in learning gains. However, female students caught up during a multi-week classroom implementation. These findings provide evidence that a collaborative gameplay approach may, over time, compensate for gender differences in experience and lead to equitable learning experiences within game-based learning environments for computer science education.**

Keywords—computational thinking, collaboration, gender, game-based learning, middle school.

I. INTRODUCTION

Computing education has become a focus of attention among both policymakers and researchers. Even as computing skills become increasingly integral to 21st century jobs, computer science is studied by only a fraction of students in the United States, and this fraction is typically not diverse. To address the national need for a computationally skilled workforce, rigorous computer science learning must go hand in hand with increased participation of students from underrepresented groups [1]. Accordingly, the computer science education research community has identified the critical need to create a classroom climate that fosters student learning and retention for these diverse learners [2]. At the K-12 level, many current initiatives in the United States seek to broaden participation in computing, including the development of innovative pre-college curricula such as Exploring Computer Science [3] and the AP Computer Science Principles course [4]. Fundamental to these initiatives is the mission to engage students who are historically underrepresented in computer science, and also to support learning in a measurable way. An increasingly central element that pervades these curricular innovations is *collaborative learning*, in which students work together to solve problems [5], [6].

Evidence suggests that collaborative learning provides many benefits for computer science learners, including improved performance and lower attrition [7], particularly for women [8]–[10], fewer "stuck" moments while problem solving [11], earlier application of critical thinking [12], and the ability to solve problems that may have been just beyond the reach of the students individually [13], [14]. Collaboration has become a central and highly valued skill for the 21st century [15], so much so that efforts are underway to develop collaboration assessment frameworks as part of the CS Principles AP course and within a multi-national K-12 assessment program [16].

Our research team has embarked on a middle school initiative in which we have integrated the benefits of collaborative learning with the engaging nature of game-based learning environments. We hope to leverage the benefits of these two strategies to spark interest in computing and also lead to significant learning gains. Over the past three years, we have developed ENGAGE, an immersive game-based learning environment that adapts learning objectives from the AP CS Principles course [4] for the middle school level. In the game, students take on the role of computer scientists as they develop computational thinking skills [17] while solving a socially relevant mystery. We have iteratively held a series of classroom studies of the game in multiple middle schools with diverse student populations, leading to refinements of both the game and the way we deploy the game. Students now choose partners at the beginning of the intervention and then play the game in pairs over the course of several weeks.

This paper describes the ENGAGE project's strategy for improving gender diversity in computer science activities through collaborative learning in a game-based environment, as well as results establishing the effectiveness of that strategy for supporting learning. First Section II provides some related work. Section III describes a pilot study conducted with two conditions: paired gameplay and single-player gameplay. The results of that pilot study suggested a paired gameplay approach has merit for supporting learning, particularly for female students. We thus proceeded with the paired gameplay approach in a full classroom study of ENGAGE, occurring over eight weeks in the context of a middle school science elective. Section IV describes the learning gains found during that study, with particular attention paid to differences based on gender. The results show the promise of integrating paired gameplay with game-based learning environments to support computer science learning at the middle school level.

II. Related Work

A. Middle School Computer Science

Recently, the computing education research community has placed an increasing emphasis on computer science in middle school. Curriculum frameworks by the Computer Science Teachers' Association and interventions by groups such as Code.org have reached millions of students, contributing to the significant momentum. We are beginning to see longitudinal studies of K-12 students' computing attitudes and self-efficacy [18], as well as emerging design-based research to examine how well a visual programming-based curriculum prepares students for later text-based programming [19]. These research projects join a landscape of increasingly diverse computer science interventions. For example, a recent project has combined jewelry design with 3D printing to teach students about technology and programming [20], while another project has built an intervention that engages children, together with their grandparents, in classroom activities [21].

Because of the widely recognized need to build students' computational thinking skills, a number of computer-based learning environments have emerged. Middle school programs have utilized Scratch programming extensively [22], [23] and emphasized reaching students with disabilities [24], urban youth [25], and underrepresented groups [26]. Alice 3D has been used to integrate computing within the context of a wide variety of subjects such as math, science, and language arts [27], and to help students understand what their future careers in computing might look like [28]. A community of practice for middle school and high school teachers has also emerged around teaching introductory computing with Alice [29]. Emerging work is focusing on building a language and development environment, LaPlaya, tailored for early middle school and upper elementary school [30], as well as integrating computational thinking into middle school science with CTSiM [31].

B. Collaborative Learning in CS

Collaborative learning has also been an area of focus in computing education research, most notably in the form of pair programming [32]. Pair programming has been studied with younger learners, and compared to other forms of collaborative learning [33]. Ongoing efforts to develop formal assessments of computational thinking at this level have solidified claims that pair programming has great potential benefits for middle school students [34]. When compared to non-collaborative learning environments, pair programming can have a particularly positive impact on girls' enjoyment and perception of learning [35] and can improve Latina students' perceptions of computer science and aid in developing their identities as computer scientists [36]. However, recent work highlights the potential negative impact of unbalanced collaboration, in which one partner dominates the learning task, leading to inequitable learning experiences [37]. Understanding the nuances of how collaboration affects student learning still stands as a critical open research question.

Another specific area for the study of collaboration in computer science education is game-based learning. The combined benefits of collaborative learning and educational games may lead to increased learning and student engagement in computer science courses [38], [39]. Games that support multiple players may also lead to a more diverse and sustainable learning experience [40]. This paper builds on this research to examine learning gains in a collaborative game-based learning environment that students play over a sustained period of time. The evidence indicates that this fusion of collaboration and game-based learning led to equitable learning gains, regardless of a student's gender or prior experience with similar gaming environments.

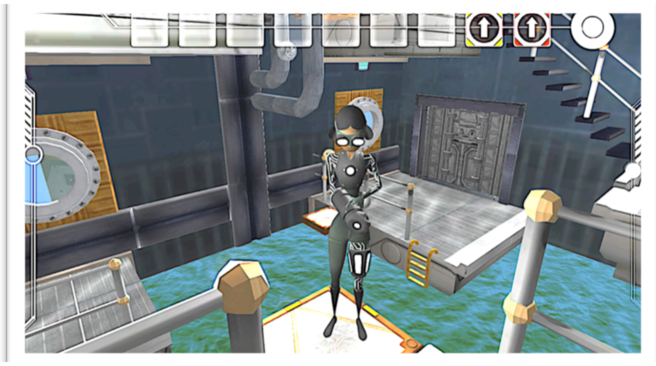
C. Game-based Learning

Game-based learning in general has been widely utilized for computer science education. Moreover, a growing body of evidence is emerging that suggests game-based learning environments hold great promise for middle school students in particular [41], [42]. Recent syntheses of the game-based learning literature have found that games can indeed yield positive learning outcomes across a range of subjects and settings [43]. A recent pair of meta-analyses have independently concluded that, in general, digital game technologies are often found to be more effective than traditional instructional methods in terms of cognitive outcomes, such as learning and retention [44], [45]. The game-based learning community has expanded efforts to conduct empirical game-based learning studies over the past several years. For example, a series of studies with the River City game-based learning environment found that students demonstrate positive learning gains and increased inquiry behaviors [46]. Quest Atlantis, a popular multi-user virtual environment has been the subject of several quasi-experimental studies, which revealed significant student learning gains [47], as well as substantial motivational benefits [48] compared to baseline conditions. Studies such as these have begun critical progress toward establishing an empirical account of the effectiveness and design of game-based learning environments.

III. Pilot Study: Single-Player Vs. Paired Gameplay

As part of our user-centered development, we conducted an exploratory study in which middle school students played the introductory level of ENGAGE [49]. As noted above, ENGAGE is a game-based learning environment for teaching computational thinking to middle school students. Students take on the role of a computer scientist sent to an underwater research station to solve a socially relevant mystery. To accomplish their goals in the game, students write programs for various devices to help their avatars advance through the three-dimensional game environment. Figure 1 shows a screenshot of ENGAGE from the segment of the game that students played during this pilot

Fig. 1. Screenshot of the ENGAGE game-based learning environment



study. The study took place over one week in the spring of 2014, with each student participating in two hour-long sessions on separate days. Students participated during the school day as part of a regularly scheduled class, and were randomly assigned to play the game either individually or collaboratively in pairs.

We designed this pilot study to investigate the impact of collaborative paired gameplay on student outcomes and experiences. As we do throughout this paper, we focus here particularly on cognitive outcomes. While our project team is keenly interested in all aspects of student learning in the game (such as affective outcomes and computer science attitudes), space does not allow for a full discussion of all those results. Instead, this paper reports on students' use of computational thinking, which we assessed through field observations, survey responses, and an early version of our knowledge assessment instrument. We also used the pilot study to refine this instrument, which ultimately became our primary tool for measuring the learning gains of students in full studies of ENGAGE [50]. For the full classroom study described in Section IV, we used the more refined version of the instrument to track students' learning throughout a multi-week intervention. In this current section, we discuss the findings of the pilot study that informed the design of that full classroom study.

A. Participants

For the pilot, we worked with two seventh grade middle school teachers and their classrooms from an urban middle school. After consent and assent was obtained, 28 seventh grade students were randomly assigned to either the paired (N = 14) or unpaired condition (N = 14), and then played the ENGAGE game across the span of two separate gameplay days. Of the 28 students, 26 completed a post-survey on engagement (two students assigned to the paired condition arrived very late, and did not have time to finish the game or post-survey).

By conducting the pilot study during the normal school day hours and asking all students in the two classes to participate, we expected to achieve a more representative subject pool than a self-selecting, after-school study might provide. This strategy proved successful, as the participants included 14 female students and 14 male students. The demographic composition was 7 African-American students, 7 white students, 6 Latino students, 2 multiracial students and 1 Asian student, (with 5 unknown). Using a survey item that asked, "Have you ever done any activities that involve computer science or computer programming?", we classified 11 students as having prior programming experience and 12 students as not having prior programming experience.[1] Using a survey item that asked, "How often a week do you play computer or video games?", we classified 12 students as frequent video game players (those who responded to the item with "daily" or "almost daily"), and 14 students as less frequent video game players.

B. Task

Once the students had been randomly assigned to their workstations, we had them log onto laptop computers and each individually complete the early version of the knowledge assessment. The version administered for this study consisted of 6 multiple-choice questions on programming concepts we expect students to master in the specific segment of game played in this study and took about 5 minutes to complete. Students in the single player condition were allowed to begin the game immediately upon completion of this assessment. For students in the paired player condition, we waited until both students in a partnership had completed the assessment, and then briefly gave them instructions on how paired gameplay can work before having them start the game.

When played by two people, the game allows each player to select an avatar to represent him or herself. Only one avatar is visible in the game environment at any given time, switching at predefined intervals. We encouraged students to alternate who controlled character movement based on which avatar was visible. In effect, we encouraged a style of gameplay similar to pair programming, in which students alternate between being the driver (at the keyboard) and being the navigator (advising the driver). While we did notice some alternative approaches to pair gameplay among students in other grades, the seventh graders in this study all seemed comfortable adhering to this style.

As mentioned above, the study was designed for two hour-long sessions on separate days. No student in either condition finished the game during the first session, so they all had to stop mid-game and resume two days later. During the second session, students completed the segment of game used in the study. Upon completion, they then took the knowledge assessment again, followed by a survey on their game experience, including items addressing specific game strategies that they may have used while playing.

C. Results

We captured a rich, multifaceted corpus of data, including survey data, field observations, and learning gains derived from the knowledge assessment. This paper focuses on using the data to assess computational thinking. The following subsections include survey data that indicate gender differences were found in the computational thinking strategies used by students during this segment of gameplay, observational data that reveal some of the benefits of paired game-play (along with some caveats), and knowledge assessment data that highlight areas of concern for

[1] Some of the descriptions that students gave of prior programming experience were ambiguous, making it difficult to classify all students.

collaborative gameplay interventions. In total, the results show that female students in general may initially have been disadvantaged due to less prior gaming experience, but that a collaborative gameplay approach has benefits that could help mitigate this inequity if deployed with careful forethought.

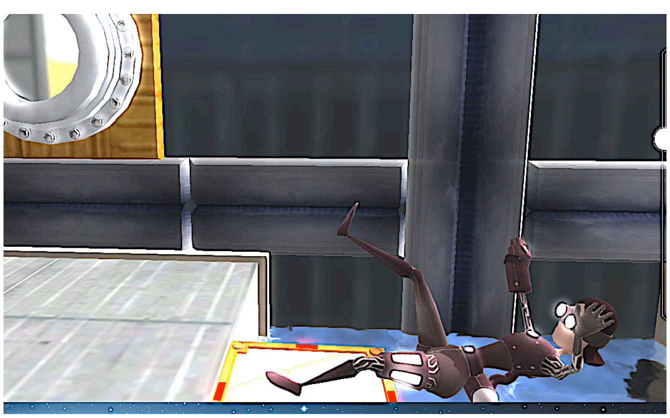
1) Survey Data

After completing the gameplay session, students completed a survey that included several items concerning their reaction to the game, as well as items on their prior relevant computing experiences and their demographic information. In analyzing the data, we looked for differences in game strategies used by students playing the game. In our observations, certain strategies seemed especially beneficial to students as they used their computational thinking skills to progress through the game environment. Collaboration might lead to wider use of these useful tactics, as students share their "best practices". Without collaboration, students with less gaming experience seem to be at a disadvantage, as results show them to be less likely to take advantage of these strategies.

One survey item in particular illustrates this challenge, an item we refer to as *Test Platform*. This item asked student participants to respond to the following question on a 5-point Likert scale: "How often within the game did you test the program for the moving platform without being on the platform?" This item refers to certain locations in the game where students needed to program a moving platform device, which (if programmed correctly) can transport the player's avatar to a desired location. If the player is standing on the moving platform and the platform crashes into an object (i.e. the program was flawed), the player will fall off and be forced to repeat some prior gameplay. Figure 2 shows a screenshot of this happening to a player. This can be frustrating, but it is often avoidable. The beneficial strategy here, Test Platform, is when a player runs the program for the moving platform *before getting on it*. In this case, the player does not risk falling off the platform if it crashes.

The results show that female students used the Test Platform strategy less often than male students. Whereas male students responded to the 5-point Likert scale item on this strategy with a 3.54 (SD = 1.450), female students responded with an average 2.21 (SD = 1.251). A one-way ANOVA found this to be statistically significant ($F(1, 26) = 6.482$, $p < .05$). Although we can view the Test Platform strategy as an example of computational thinking, it is interesting to note that no significant differences were found between students who reported having had previous programming experience and those students who reported none.[2]

Fig. 2. When an error in a program causes the moving platform to crash, the player's avatar will fall off (if it is currently riding the platform)



However, while prior programming experience may not have had an influence on how often students used the Test Platform strategy, prior *gaming* experience did. Students who reported playing video games "daily" or "almost daily" responded to the item on this strategy with a 3.58 (SD = 1.621), while those with less frequent video game experience responded with an average 2.27 (SD = 1.100). A statistically significant difference was also found between these two groups ($F(1, 26) = 6.302$, $p < .05$). It must be noted that female students reported a lower amount of weekly video game experience, responding with an average 3.21 on a 5-point Likert scale, compared to an average for male students of 4.0. A one-way ANOVA found this to be statistically significant ($F(1,26) = 5.667$), $p < .05$).

Taken together, these preliminary results show how students' prior experiences inform the actions that they take within a game-based learning environment. If playing alone, a student might eventually discover these strategies over time, but she will be much more likely to do so if collaborating with other students with diverse prior experiences. We can thus take inspiration from the software engineering practice of pair programming, which leads to the "diffusion of innovation" among software developers [51].

2) Observational Data

The field observations support this claim that students can gain significant benefits when playing collaboratively with a partner. Overall, observations of students' interactions within the game indicated that students from both conditions had an enjoyable experience. Students' postgame comments echoed this sentiment. From a paired team member: "I think that this game was awesome and that I had fun playing with my partner", and from a single player: "I had lots of fun playing it by myself". The classroom teachers, who were in attendance for the entire duration and played the game themselves, commented that the students seemed particularly focused compared to a typical day.

Throughout the overall gaming experience, observations revealed advantages for paired gameplay versus single gameplay. For example, a given student might use her teammate as a "sounding board", or her teammate might provide suggestions for what to try next or reasoning about what was happening within the game. One student put it succinctly: "I enjoyed working with my partner because he helped me when I was trying to figure the game out". Additionally, in cases where one member of a pair was having great difficulty with character navigation (likely due in part to lack of gaming experience), her partner could take over during

[2] A similar proportion of male and female students reported having previous programming experience.

those times when character navigation was particularly tricky. Finally, students playing in pairs had the potential to receive superior affective support from their partners. If one student started feeling frustrated or discouraged, her partner could revive her spirits through the social collaboration.

We also paid close attention to potential disadvantages, however. While the single player students did not have the advantage of a peer who could provide encouragement and support, they also did not have anyone criticizing their actions. While most instances were in jest, there was some element of cross-partner frustration. Because the pairing of students was random, it is unsurprising to see some partnerships led to more successful interactions than other partnerships. In later classroom implementations of ENGAGE (such as the one described in Section IV), we allowed students to choose partners rather than use random pairing.

3) *Knowledge Assessment Data*

Considering the benefits mentioned above of collaboration, we hoped to see superior learning gains from those students who play the game in pairs. At the time, we were still developing our knowledge assessment [50], which students took both before and after playing the game. Because the instrument had not been fully validated at this point, we must interpret the score results (which were similar for both conditions) conservatively. An examination of the results from individual questions, however, can provide specific insight into how well the students mastered certain concepts we hoped them to learn. In particular, we were interested in how well the students mastered the concept of broadcasting, since field observations revealed that many students found the introduction of this concept in the game to be particularly challenging.

The knowledge assessment addressed this concept in Question 6. In the single player condition, every student answered Question 6 correctly. In the paired player condition, however, only 8 out of 14 students answered it correctly. This illustrates one of the potential pitfalls of the paired gameplay approach from a pedagogical point of view, an issue of equity that has been seen in other K-12 computer science studies that emphasize collaborative learning [37]. While having students play in pairs may better engage them with the cognitive challenges of the game, it also sets up the possibility that a stronger partner can advance the pair through a challenge without the weaker partner understanding how the challenge got solved. Of the seven partnerships, this problem seemed to manifest in four, as evidenced by one of the partners answering Question 6 correctly and the other answering it incorrectly. Addressing this drawback is a major open question for designing game-based learning environments that support paired gameplay. As the next section describes, however, this negative consequence may decline over time in longer-term collaborative gameplay interventions.

IV. FULL STUDY: LEARNING GAINS AMONG STUDENTS PLAYING IN PAIRS

Having revised the game-based learning environment and the knowledge assessment instrument following the results of that pilot study, we then conducted a full study of ENGAGE within two urban middle schools in Raleigh, NC in the fall of 2014. Contrary to the pilot study, which we had conducted in a controlled environment, we deployed this full study in the context of a quarterly science elective. At each school, a cohort of students attended the elective five days a week during their regular school day. One of the school's full-time teachers taught the elective, with members of our research team attending the gameplay session to provide support and record field observations. Over the course of the quarter (approximately 2 months), several class sessions a week were given for students to interact in the game-based learning environment.

Each student chose a partner on the first day and then collaboratively played the game with that same partner throughout the quarter. This paired gameplay model was motivated by the results from the pilot study, and also by logistical concerns. Because of limited technology in the computer labs of the two schools (a common issue in under-resourced schools), it would have been infeasible to have every student play the game individually on separate computers. The paired gameplay model thus allows deployment of the game-based learning environment with half as many working computers as there are students in a given class.

A. *Participants*

This section reports on the 48 students who played the game in pairs during Quarter 1 of the elective (and gave consent for their data to be used). Of these students, 26 were male and 22 were female. The demographic composition was 21 White students, 13 Asian students, 8 African-American students, 2 Latino students, 1 Middle Eastern student, and 3 other. On the survey item asking about prior computer programming experience, we classified 11 students as having prior programming experience and 29 students as not having prior programming experience. On the survey item asking how often they play computer or video games, 21 students responded "every day" or "almost every day", while the remaining 27 students responded "occasionally" or "almost never".

B. *Task*

During the first week of Quarter 1, before the initial introduction of the game-based learning environment, students completed the refined version of the knowledge assessment instrument. We used this as a pretest to measure their incoming computational thinking skills. The full test consisted of 23 items that covered the entire gameplay, which is distributed over three distinct game levels. However, since only the first two game levels were deployed for this study (Level Three was still in development), we will report only on the 15 items aligned with the first two game levels.

Students then played the game in pairs during the class sessions scheduled for gameplay (roughly every other day was reserved for gameplay, with the interceding days full of complementary science activities). When a pair finished Level One of Engage (which occurred after three to five gameplay sessions, on average), both partners individually completed an interim posttest. This interim posttest included the items on the knowledge assessment that we expected students to learn while playing Level One.

Upon completion of that test, the pair would then resume gameplay in Level Two. This level, which is both longer and demands more complex computational thinking, took students approximately 7 to 10 additional gameplay sessions. After a pair finished this level, they completed the full knowledge assessment instrument as a posttest. For the purpose of this paper, we break down the knowledge assessment into "Level One content" and "Level Two content", depending on where in the game we expected students to learn the concept targeted by an individual assessment item. There are four items that assess concepts introduced in Level One, and 11 items assessing concepts from Level Two.

C. Results

1) Investigation of Learning Gains for Underrepresented Students.

The computational thinking knowledge assessment addresses our need to evaluate how well the game-based learning environment serves all students. By administering it as a pretest, we were able to assess the extent to which students already had these targeted computational thinking skills. We expected that some students would enter with more knowledge than others. Indeed, even at the middle school level, students may have widely varying exposure to computer science. Moreover, we hypothesized that students traditionally underrepresented in computer science would score lower on the pretest than their peers. This paper is focused on underrepresentation based on gender, as well as whether there are differences based on students' prior programming or gaming experiences. Table 3 illustrates the differences we found along these three metrics. Overall, students scored an average of .458 (SD = .219) on the pretest (a perfect score would be 1.0), but significant differences were found between female students and male students.

TABLE II. COMPARISON OF SCORES ON PRETEST

	Performance on Pretest	**Sig.**
Gender	*Female*: .383 (SD = .188) *Male*: .517 (SD = .227)	p<.05
Prior Programming Experience?	*No*: .42 (SD = .152) *Yes*: .63 (SD = .243)	p<.01
Frequent Video Game Experience?	*No*: .385 (SD = .187) *Yes*: .558 (SD = .225)	p<.01
Overall	.458 (SD = .219)	

To better understand the difference in pretest scores between male and female students, we ran a One-Way ANOVA in SPSS and the results showed a statistically significant difference ($F(1,43) = 4.486$, $p < .05$). The disparity was even greater between students based on their prior experience with programming ($F(1,39) = 10.456$, $p < .01$) and video games ($F(1,43) = 7.952$, $p < .01$). Thus, just as we had found during the pilot study that frequent video game experience correlated with more frequent use of the beneficial Test Platform strategy, our assessment instrument revealed a similar disparity in pre-knowledge of computational thinking concepts.

Having established that underrepresented students did indeed enter with less knowledge than their peers, we next compared pretest scores with posttest scores to examine learning progressions of students. As mentioned above, we also administered an interim posttest after students completed Level One of the game to measure the extent to which student had mastered those concepts early on. Table II shows the average scores for female and male students on each of these tests, broken down by level.

TABLE I. COMPARISON OF COMPUTATIONAL THINKING KNOWLEDGE AT MULTIPLE STAGES

	Female (n = 19)	**Male (n = 21)**
	Level 1	
Pretest (Level One content)	.51 (SD = .282)	.58 (SD = .266)
Interim Posttest	.434 (SD = .261)	.691 (SD = .315)
Posttest (Level One content)	.645 (SD = .268)	.726 (SD = .315)
	Level 2	
Pretest (Level Two content)	.36 (SD = .183)	.53 (SD = .261)
Posttest (Level Two content)	.63 (SD = .246)	.70 (SD = .212)

The results indicate that, while female students demonstrated less knowledge early on, they made great gains as they progressed in the game. A one-way ANOVA found the difference between genders on the Interim Posttest for Level One to be statistically significant ($F(1,39) = 7.735$, $p < .01$). The difference between genders on the pretest of Level Two content was also statistically significant ($F(1,39) = 5.193$, $p < .05$). Thus, at this early stage of the gameplay, we do not see the gender gap closing. Indeed, the normalized learning gain here is disheartening. We calculated learning gain as (Post – Pre)/(1 – Pre). When using the Interim posttest, this calculation showed the male students as having a higher learning gain (.387 for males, compared to .110 for females). A one-way ANOVA found this to be statistically significant ($F(1, 39) = 5.684$, $p < .05$).

Yet the longer students interacted in the game-based learning environment, the less these differences manifested themselves. On the posttest, no statistical differences were found between male and female students, as female students' greater learning gains leveled the playing field. Indeed, female students mastered the Level One concepts as they saw them and applied them more often during Level Two, and this did not take away from their learning of the Level Two concepts. Here we see the importance of persistence. We propose that the collaborative nature of the gameplay better enabled this persistence.

2) *Examination of Pretest Knowledge Differences between Partners*

A major concern for this paired gameplay approach arises when considering pairs of students in which Student A has more prior knowledge than Student B. The potential exists for Student A to dominate the interaction, solving problems rapidly and leaving little opportunity for Student B to explore and learn. With this concern in mind, we sought to examine if and how differences in pretest knowledge between two partners correlated to learning. Unsurprisingly, we found a range of differences between partners. In only three pairs did both partners score exactly the same on the pretest, while the greatest difference between the pretest scores of two partners was .467. To aid this investigation, we defined a new variable, *Difference_Pretest* as (*students' pretest score – partner's pretest score*). Students who scored lower on the pretest than their partners thus have a negative *Difference_Pretest*, while those who scored higher than their partners have a positive *Difference_Pretest*.

A test for correlation between *Difference_Pretest* and learning gains found no statistically significant differences. Students therefore did not seem to be disadvantaged if their partners had more prior knowledge. In fact, they may have benefitted from having a stronger partner. Classifying each student as either *Negative Difference_Pretest* or *Nonnegative Difference_Pretest*, we found that *Negative Difference_Pretest* students achieved a superior learning gain of .247 (SD = .191) compared to their peers' .113 (SD = .113). A One-Way ANOVA found this result to be statistically significant ($F(1,39) = 4.714$, $p < .05$). It should be noted that a ceiling effect may limit the learning gains of some students in the *Nonnegative Difference_Pretest* category. Regardless, the positive learning gains of the students with less prior knowledge further support the paired gameplay approach, as it contradicts the fear that such students will be left behind as their partners dominate the learning experience.

V. Conclusion

The studies reported here indicate that paired gameplay has significant potential for improving the gender equity of game-based learning environments. From observing both those students who played the educational game in pairs and those who played it individually, we noted several beneficial aspects of pair gameplay, as well as one or two caveats. When students play in pairs, they can provide each other various types of support, although the quality of this support depends somewhat on the two individuals. While the learning benefits of collaboration (as measured by a validated knowledge assessment) might not manifest in the initial session of gameplay, we saw collaborative gameplay lead to equitable learning gains as students continued playing the game over time. Indeed, whereas female students (and students with less gaming experience) used certain key computational thinking strategies less often than their peers during the introductory level of the game, we observed that collaboration led to a sharing of best practices as time went on. Through this "diffusion of innovation," students achieved significant learning gains regardless of their gender or their prior gaming experiences.

Future work should investigate how the combination of educational games and collaboration affects students of other underrepresented groups. Although we had a diverse pool of student participants, this paper has not examined differences based on race or ethnicity, for example. To do so presents significant challenges, but it is crucial for our understanding of how to create game-based learning environments that are equitable for all learners. Additionally, future work should also look at collaboration in games at a finer granularity by looking at game-trace data and multimodal data. This will provide better insight into the nature of collaboration and what collaborative strategies lead to equitable learning gains for both partners in a paired gameplay scenario. Finally, future work should explore how pedagogical agents can be integrated into human-human collaboration within virtual learning environments in order to even more fully support a diverse ranger of learners with different needs.

Acknowledgment

This work is supported in part by the National Science Foundation through Grants CNS-113897 and CNS-1042468. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

References

[1] "BPC Portal," 2014. .
[2] L. J. Barker, M. O'Neill, and N. Kazim, "Framing Classroom Climate for Student Learning and Retention in Computer Science," in *Proceedings of the 45th ACM Technical Symposium on Computer Science Education (SIGCSE '14)*, 2014, pp. 319–324.
[3] J. Goode and J. Margolis, "Exploring Computer Science," *ACM Trans. Comput. Educ.*, vol. 11, no. 2, pp. 1–16, Jul. 2011.
[4] O. Astrachan, T. Barnes, D. D. Garcia, J. Paul, B. Simon, and L. Snyder, "CS Principles: Piloting a New Course at National Scale," in *Special Interest Group on Computer Science Education (SIGCSE) Conference*, 2011, pp. 397–398.
[5] B. Coleman and M. Lang, "Collaboration Across the Curriculum: A Disciplined Approach to Developing Team Skills," in *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education (SIGCSE '12)*, 2012, pp. 277–282.
[6] K. Falkner, N. J. G. Falkner, and R. Vivian, "Collaborative Learning and Anxiety: A phenomenographic study of collaborative learning activities," in *Proceedings of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)*, 2013, pp. 227–232.
[7] P. Lasserre, "Adaptation of team-based learning on a first term programming class," in *ITiCSE'09*, 2009, vol. 41, no. 3, pp. 186–190.
[8] L. L. Werner, B. Hanks, and C. McDowell, "Pair-programming helps female computer science students," *J. Educ. Resour. Comput.*, vol. 4, no. 1, p. 4, 2004.
[9] C. M. Lewis, N. Titterton, and M. Clancy, "Using collaboration to overcome disparities in Java experience," *Proc. Ninth Annu. Int. Conf. Int. Comput. Educ. Res. - ICER '12*, pp. 79–86, 2012.
[10] J. Liebenberg, E. Mentz, and B. Breed, "Pair Programming and Secondary School Girls' Enjoyment of Programming and the Subject Information Technology (IT)," *Comput. Sci. Educ.*, vol. 22, no. 3, pp. 219–236, 2012.
[11] B. Hanks, "Problems Encountered by Novice Pair Programmers," in *ICER'07*, 2007, pp. 159–164.
[12] Z. Li and E. Kraemer, "Social effects of pair programming mitigate impact of bounded rationality," in *Proceedings of the 45th ACM technical symposium on Computer science education - SIGCSE '14*, 2014, pp. 385–390.
[13] L. S. Vygotsky, "Mind and society: The development of higher mental processes." Cambridge, MA: Harvard University Press, 1978.

[14] L. Williams and R. Kessler, *Pair Programming Illuminated*. Boston: Addison-Wesley, 2003.

[15] J. Strijbos, "Assessment of (Computer-Supported) Collaborative Learning," *IEEE Trans. Learn. Technol.*, vol. 4, no. 1, pp. 59–73, Jan. 2011.

[16] "PISA 2015 Draft Collaborative Problem Solving Framework," 2013.

[17] J. M. Wing, "Computational Thinking," *Commun. ACM*, vol. 49, no. 3, p. 33, Mar. 2006.

[18] C. Aritajati, M. B. Rosson, J. Pena, D. Cinque, and A. Segura, "A Socio-Cognitive Analysis of Summer Camp Outcomes and Experiences," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15*, 2015, pp. 581–586.

[19] S. Grover, R. Pea, and S. Cooper, "Expansive Framing and Preparation for Future Learning in Middle-School Computer Science," in *International Conference of the Learning Sciences (ICLS) Conference*, 2014, pp. 992–996.

[20] C. Starrett, M. Doman, C. Garrison, and M. Sleigh, "Computational Bead Design: A Pilot Summer Camp in Computer Aided Design and 3D Printing for Middle School Girls," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15*, 2015, pp. 587–590.

[21] A. Sauppé, D. Szafir, C.-M. Huang, and B. Mutlu, "From 9 to 90: Engaging Learners of All Ages," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15*, 2015, pp. 575–580.

[22] E. Schofield, M. Erlinger, and Z. Dodds, "MyCS: a CS curriculum for middle-years students," *J. Comput. Sci. Coll.*, vol. 29, no. 4, pp. 145–155, Apr. 2014.

[23] Q. Burke and Y. B. Kafai, "The Writers ' Workshop for Youth Programmers Digital Storytelling with Scratch in Middle School Classrooms," *Proc. 43rd ACM Tech. Symp. Comput. Sci. Educ. - SIGCSE '12*, pp. 433–438, 2012.

[24] J. C. Adams, "Scratching middle schoolers' creative itch," in *Proceedings of the 41st ACM Technical Symposium on Computer Science Education (SIGCSE '10)*, 2010, pp. 356–360.

[25] J. Maloney, K. Peppler, Y. B. Kafai, M. Resnick, and N. Rusk, "Programming by Choice : Urban Youth Learning Programming with Scratch," in *Proceedings of the 39th ACM Technical Symposium on Computer Science Education (SIGCSE '08)*, 2008, pp. 367–371.

[26] D. Franklin, P. Conrad, G. Aldana, and S. Hough, "Animal Tlatoque: attracting middle school students to computing through culturally-relevant themes," in *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education (SIGCSE '11)*, 2011, pp. 453–458.

[27] S. H. Rodger, H. Qin, and J. Hayes, "Engaging Middle School Teachers and Students with Alice in a Diverse Set of Subjects," in *Proceedings of the 40th ACM Technical Symposium on Computer Science Education (SIGCSE '09)*, 2009, pp. 271–275.

[28] H. C. Webb and M. B. Rosson, "Exploring Careers While Learning Alice 3D: A Summer Camp for Middle School Girls," in *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education (SIGCSE '11)*, 2011, pp. 377–382.

[29] S. Cooper, S. H. Rodger, M. Schep, R. H. Stalvey, and W. Dann, "Growing a K-12 Community of Practice," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15*, 2015, pp. 290–295.

[30] C. Hill, H. A. Dwyer, T. Martinez, D. Harlow, and D. Franklin, "Floors and Flexibility: Designing a Programming Environment for 4th-6th Grade Classrooms," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15*, 2015, pp. 546–551.

[31] S. Basu, A. Dukeman, J. S. Kinnebrew, G. Biswas, and P. Sengupta, "Investigating Student Generated Computational Models of Science," in *International Conference of the Learning Sciences*, 2014, pp. 1097–1101.

[32] B. Hanks, S. Fitzgerald, R. McCauley, L. Murphy, and C. Zander, "Pair programming in education: a literature review," *Comput. Sci. Educ.*, vol. 21, no. 2, pp. 135–173, Jun. 2011.

[33] C. M. Lewis, "Is pair programming more effective than other forms of collaboration for young students?," *Comput. Sci. Educ.*, vol. 21, no. 2, pp. 105–134, Jun. 2011.

[34] L. Werner, J. Denner, and S. Campe, "The Fairy Performance Assessment: Measuring Computational Thinking in Middle School," in *Proceedings of the 43rd ACM technical symposium on Computer science education (SIGCSE '12)*, ACM, Ed. New York, NY: ACM, 2012.

[35] J. Liebenberg, E. Mentz, and B. Breed, "Pair programming and secondary school girls' enjoyment of programming and the subject Information Technology (IT)," *Comput. Sci. Educ.*, vol. 22, no. 3, pp. 219–236, Sep. 2012.

[36] H. Thiry and S. Hug, "'We Should All Help Each Other': Latina Undergraduates' Practices And Identities In The Figured World Of Computing," in *International Conference of the Learning Sciences*, 2014, pp. 214–221.

[37] N. Shah, C. Lewis, and R. Caires, "Analyzing Equity in Collaborative Learning Situations: A Comparative Case Study in Elementary Computer Science," in *International Conference of the Learning Sciences (ICLS) Conference*, 2014, pp. 495–502.

[38] A. Nickel and T. Barnes, "Games for CS Education: Computer-Supported Collaborative Learning and Multiplayer Games," *Foundations of Digital Games (FDG) Conference*. pp. 274–276, 2010.

[39] N. Anderson and T. Gegg-Harrison, "Learning Computer Science in the 'Comfort Zone of Proximal Development,'" in *Proceeding of the 44th ACM Technical Symposium on Computer Science Education - SIGCSE '13*, 2013, pp. 495–500.

[40] S. R. Foster, S. Esper, and W. G. Griswold, "From competition to metacognition: designing diverse, sustainable educational games," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 2013, pp. 99–108.

[41] S. Esper, S. R. Wood, S. R. Foster, S. Lerner, and W. G. Griswold, "Codespells: how to design quests to teach java concepts," *J. Comput. Sci. Coll.*, vol. 29, no. 4, pp. 114–122, Apr. 2014.

[42] D. C. Webb, A. Repenning, and K. H. Koh, "Toward an emergent theory of broadening participation in computer science education," in *Proceedings of the 43rd ACM technical symposium on Computer science education (SIGCSE '12)*, 2012, pp. 173–178.

[43] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Comput. Educ.*, vol. 59, no. 2, pp. 661–686, Sep. 2012.

[44] S. K. M. Douglas B. Clark, Emily E. Tanner-smith, "Digital Games for Learning: A Systematic Review and Meta-Analysis," 2013.

[45] P. Wouters, C. van Nimwegen, H. van Oostendorp, and E. D. van der Spek, "A meta-analysis of the cognitive and motivational effects of serious games.," *J. Educ. Psychol. Psychol.*, vol. 105, no. 2, pp. 249–265, 2013.

[46] D. J. Ketelhut, B. C. Nelson, J. Clarke, and C. Dede, "A multi-user virtual environment for building and assessing higher order inquiry skills in science," *Br. J. Educ. Technol.*, vol. 41, no. 1, pp. 56–68, Jan. 2010.

[47] D. T. Hickey, A. A. Ingram-Goble, and E. M. Jameson, "Designing Assessments and Assessing Designs in Virtual Educational Environments," *J. Sci. Educ. Technol.*, vol. 18, no. 2, pp. 187–208, Feb. 2009.

[48] S. J. Warren, M. J. Dondlinger, and S. A. Barab, "A MUVE Towards PBL Writing," *J. Res. Technol. Educ.*, vol. 41, no. 1, pp. 113–140, Sep. 2008.

[49] M. H. Frankosky, E. N. Wiebe, P. S. Buffum, and K. E. Boyer, "Spatial Ability and Other Predictors of Gameplay Time: Understanding Barriers to Learning in Game-based Virtual Environments," in *AERA Applied Research in Immersive Environments for Learning SIG*, 2015.

[50] P. S. Buffum, E. V. Lobene, M. H. Frankosky, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "A Practical Guide to Developing and Validating Computer Science Knowledge Assessments with Application to Middle School," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education - SIGCSE '15*, 2015, pp. 622–627.

[51] E. Murphy-Hill and G. C. Murphy, "Peer interaction effectively, yet infrequently, enables programmers to discover new tools," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11*, 2011, p. 405.