Paper ID #18633

Partnership Characteristics and Student Performance in an Introductory Computer Science Course

Charles Kowalec, University of Michigan

Charles Kowalec is an undergraduate student at the University of Michigan interested in the science of how students learn.

Dr. Andrew DeOrio, University of Michigan

Andrew DeOrio is a lecturer at the University of Michigan and a consultant for web, machine learning and hardware projects. His research interests are in ensuring the correctness of computer systems, including medical devices, internet of things (IOT) devices, and digital hardware. In addition to teaching software and hardware courses, he teaches Creative Process and works with students on technology-driven creative projects.

Partnership Characteristics and Student Performance in an Introductory Computer Science Course

Charles Kowalec and Andrew DeOrio
crkowale@umich.edu, awdeorio@umich.edu
Department of Electrical Engineering and Computer Science
University of Michigan

Abstract

Group work and programming in partnerships have been shown to have a positive effect on student learning in computer science education. As a result, students in introductory computer science courses often work in teams or partnerships. In this paper, we examine the composition of student partnerships on programming projects in an introductory computer science course and its relationship to student performance.

We analyze data from 1,434 students enrolled in an introductory computer science course at a highly ranked public university, collected over one academic year. These data include student group composition, performance on projects and exams, and demographic information. The course is a second-semester “CS2” programming and introductory data structures course, and is part of the core computer science curriculum for majors and minors. Students had the option to work in partnerships or to work alone on projects.

Our results examine several factors influencing the success of a partnership, including difference in cumulative grade point average (GPA), gender balance, and work habits like starting projects early. After controlling for GPA, we observed an association between starting projects early and increased performance on both exams and projects. The impact was greatest among those in the lowest GPA quartile, where an early start made the difference between an average final letter grade of C+ (lowest early-start quartile) and B- (highest early-start quartile).

1 Introduction and Related Work

An important goal of group work in education is to increase student learning of course material. In computer science courses, group work often takes the form of pair programming. Our goal in this paper is to evaluate how different partnership characteristics affect student learning outcomes.

Past work has demonstrated a relationship between programming in partnerships and increased student performance. In 2002, McDowell *et al.* [1] observed that students who partnered on projects in an introductory computer science course tended to have higher project performance. In 2005 and 2006, Mendes[2 3] found that students who partnered in a second semester computer science course performed better on individual work and had higher exam scores.

Further, Woolley and Chabris *et al.* [4], in 2010, found evidence of a collective intelligence factor which helps to explain a group's performance on a variety of tasks. The intelligence factor they observed was not strongly correlated with the average or maximum individual intelligence of group members but was correlated with other various group dynamics such as the average social sensitivity of group members, the equality in distribution of conversational turn-taking, and the proportion of females in the group.

Not only has the literature measured student performance, but it has also investigated student perception of partnership compatibility. Nagappan *et al.* [5], when investigating the effects of pair programming in a CS1 course, found that students who participated in pair programming desire compatible partnerships. In 2005, Katira *et al.* [6] performed a study to better understand partnership compatibility, and found that mixed gender partnerships were less likely to report compatibility than same gender partnerships.

In addition to compatibility and gender composition, researchers have also analyzed the relationship between personality and partnership performance. In 2009, Salleh *et al.* [7] investigated the effects of differences in personality traits among partners and their effect on academic performance. In this study, personalities were measured using the five factor model, which characterizes personality into five traits, namely: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. They found that differences in personalities according to these traits did not contribute to the academic performance of individuals who pair programmed. In 2010, Hannay *et al.* [8] arrived at similar results and suggested that "more effort should be spent on investigating other performance-related predictors such as expertise, and task complexity, as well as other promising predictors, such as programming skill and learning."

DeOrio and Giugliano[9] found that students who participated in programming partnerships tended to have better project performance than those working alone, especially those in lowest GPA quartile. Additionally, they found that students who pair programmed in a CS2 course also tended to do better on projects in a CS3 course, in which they worked individually.

In this paper, we hope to expand upon previous work by examining the relationship between student performance and additional potential predictors, such as partnership parity, work habits like starting projects early, and partnership gender composition. Our goal is to provide insights about the defining characteristics of a successful partnership.

2 Methods

We use quantitative analysis to examine the performance of student partnerships in an introductory computer science course. Our data set includes information from course grade books. Additionally, we collected partnership information from the computerized grading infrastructure used for programming projects. Lastly, we acquired demographic and GPA data from university records.

Through our analysis, we examine the following research questions:

- What kinds of partnerships form? Are they balanced?
- Do balanced partnerships perform better or worse than unbalanced ones? Does starting projects early affect performance?

2.1 Description of the Course

Our data are collected from a second semester programming course ("CS2") whose target audience is sophomores planning to pursue a major or minor in Computer Science. Although Computer Science majors and minors make up the majority of the students in the course, other majors such as Business Administration, Economics, and Statistics are also represented. CS2 has a single prerequisite, an introductory programming course taught in C++.

CS2 focuses primarily on core computer science topics, but also covers some specific elements of C++. Topics include: Functional abstraction, including specification, recursion, iteration, and functional generalization. Data abstraction, including types, type hierarchies, abstract data types, abstraction and polymorphism. Dynamic resource management, including creation, deletion, and interaction with containers. Specific C++ concepts, including arrays, structs, classes, objects, function and operator overloading, strings, pointers, templates, linked lists, stacks and queues, iterators, and functors

Each semester, CS2 is offered in several lecture sections, each meeting for 3 hours per week (two 1.5 hour lectures), as well as an additional weekly lab section. Lecture sections are synchronized, covering the same material each week. Lab sections are also synchronized, and provide students with short assignments reinforcing the material covered in lecture. Final grades are a weighted average of coding projects (40%), lab exercises (5%), a midterm exam (25%), and a final exam (30%).

The five programming projects in the course require students to write programs ranging from several hundred lines to over 1,000 lines of C++ code. Students had between 2 and 3 weeks to complete each project. The curriculum, course policies, and grading rubric remained consistent through the semesters from which we draw our data.

2.1.1 Partnerships

All students worked alone on the first project, and in optional partnerships for the remaining four projects. During a project, partnerships were fixed, but between projects, students had the option to switch partners or to end a partnership and work alone. Each partnership submitted one project for the pair, and both partners received the same grade. Students working in partnerships were encouraged to pair program, and were strongly discouraged from dividing the work between partners, both in lecture and in the course syllabus.

2.2 Design of the Study

Our data come from two consecutive semesters of CS2. The data set includes each student's cumulative GPA before taking CS2, performance on projects and exams as well as total score in the course. Additionally, we recorded whether each student worked in a partnership for each project and their partner. From grading software records, we obtained project submission timestamps (students could submit several times per day until the deadline). From university records, we obtained records of students withdrawing from the course, final letter grades for each student's 100-level programming course, and demographic information.

To measure balance within a partnership, we used several independent variables. The first computes the difference between two partners' cumulative GPAs, which we refer to as our parity metric. The second independent variable is the gender makeup of the partnership: two men, two women, and mixed gender. Additionally, we measured how early a partnership started the project.

We computed the parity metric for a partnership using following formula:
$parity = \frac{4.0-|(GPA_0-GPA_1)|}{4.0}$, where GPA_0 and GPA_1 refer to the previous semester cumulative GPAs of the two partners. This leads to a parity scale of $[0, 1]$, where 0 is defined as a partnership with opposite GPAs, and 1 is defined as a partnership with identical GPAs.

Our work habits metric is an indicator of how early the partnership started working on a project. Specifically, we measured the number of days between a partnership's first submission to the course's automated grading website and the deadline. Students could submit several times per day until the deadline.

To measure the performance of a partnership on projects, we computed the average project grade for all of the projects completed by that partnership. Partnership performance on exams is also considered, and is represented by the average of the two partners' exam grades.

2.3 Statistical Analysis

Our data collection spanned two semesters of CS2. Each semester had different exams (on the same topics) with different means. We used Z-scores for all grade data, including both exams and projects. After verifying that each category had a similar variance across semesters, we computed Z-scores for each semester's grade data.

Partnership parity, starting early (days between first project submission and the deadline), and gender composition were independent variables. Dependent variables included exam score and project score. We also used overall course letter grade to describe the impact of different exam and project scores. We evaluated the statistical significance of our observations using ANOVA. When performing analysis using our GPA parity metric, we considered the parity and starting early metrics as continuous variables. When performing analysis using our gender balance metric, we partitioned our data into three groups: two men, two women and mixed gender.

2.3.1 Evaluating Partnership Parity

We considered several options for measuring partnership parity: difference in GPA and difference in 100-level programming course final grade. When evaluating 100-level programming course final grade, we encountered several difficulties. First, there are several courses, that are not directly comparable; for example, one is an advanced course. Second, one of the courses had a very small sample size compared to the others, making it difficult to draw accurate conclusions. In addition, we compared our GPA-based metric against the candidate 100-level programming course metric and ultimately found that they were highly correlated. For the remainder of the paper, we use the GPA-based parity metric.

2.3.2 Independence of Variables

We first examined the ability of our parity metrics to differentiate between partnerships. Specifically, we wanted to make sure that our independent variables, namely parity, gender composition, and starting early, were not correlated. We confirmed that any relationship among the independent variables was not statistically significant.

2.4 Summary of Data

In total, our dataset contains 1,434 records for students in CS2, from the two semesters of the academic year 2015-2016.

Before performing any analysis, we first filtered our data. We removed students reported for cheating on projects or exams and students who audited the course. We filtered students who withdrew from the course since their number (23 participated in partnerships) was too small to generalize. After filtering, our data set contained 1,343 records. We then grouped records into partnerships, filtering out any records of student who did not partner, leading to a dataset of 510 distinct partnerships. One of these partnerships was removed, because its parity score was an outlier. Dropping this data point did not affect the trends in our results. Thus, our dataset contains 509 partnerships. The number of unique individuals within this dataset is 869. Note that because students were allowed to switch between working alone and in a partnership, and were allowed to change partners between projects, the number of unique individuals can be less than 1,018 and can be odd. Tables 1 and 2 below describe the individual and partnership subsets of our data, respectively. The distribution of project and exam Z-scores can be seen in Figure 1.

Gender	Count	Average GPA	Average Partner GPA Parity	Average Early-start on Projects Z-score
Women	204	3.419	0.890	-0.019
Men	665	3.421	0.889	0.008
All individuals	869	3.420	0.889	0.002

Table 1: Summary of data for individuals who worked in partnerships. Women and men have similar average GPAs, average partnership GPA parity, and average early start on the projects. The GPAs include all university grades earned before taking the introductory programming class. Parity is on a [0, 1] scale with higher parity scores representing a more similar partner GPA. Average early-start is measured as the average number of days before a project's deadline that a partnership first submitted the project to the course autograding system, represented as a Z-score because project durations varied between 2 and 3 weeks. Higher early-start scores represent partnerships who started projects earlier.

Gender Composition	Count	Average GPA	Average Partnership GPA Parity	Average Early-start on Projects Z-score
Two women	62	3.398	0.886	-0.031
Two men	319	3.419	0.890	-0.010
Mixed gender	129	3.416	0.904	0.033
All individuals	510	3.415	0.893	-0.002

Table 2: Summary of data for partnerships. Partnerships of different gender composition had similar average GPA, average parity, and average early-start. Parity is on a [0, 1] scale with higher parity scores representing more similar partner GPA. Average early-start is measured as the average number of days before a project's deadline that a partnership first submitted the project to the course autograding system, represented as a Z-score because duration varied with the assignment. Higher early-start scores represent partnerships who started projects earlier.

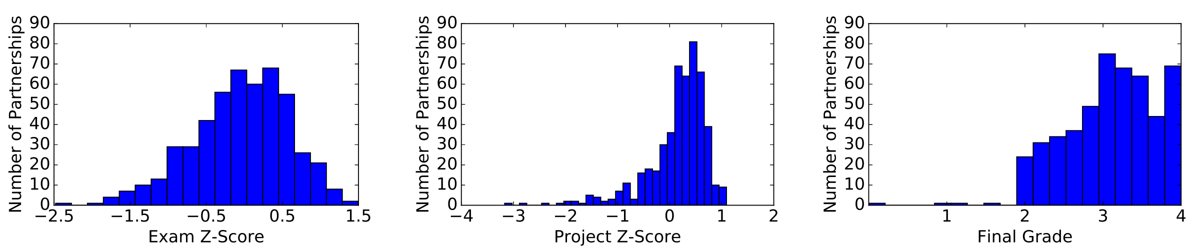


Figure 1: Distribution of exam scores, project scores and final letter grades, which are dependent variables. Exam and project scores are Z-scores, and final grade is on a 4.0 scale.

3 Results

In the following analyses, we examine several independent variables: partnership parity, starting early, and partnership gender composition. To measure student performance, we use several dependent variables: exam scores and project scores. We use course letter grades to describe the magnitude of our observations more intuitively. Finally, we examine differences among partnerships of different gender compositions.

3.1 Multivariate Analysis

In our first analysis, we examined the relationship between two independent variables, partnership parity and starting early, and dependent variables measuring student performance (project and exam scores) while controlling for GPA.

In our analysis of variance results, we observed that the relationship between starting early and exam score is significant, even after controlling for GPA (Table 3). We measured early-start as the average number of days before a project deadline that a partnership first submitted the project, represented as a Z-score. A lower Z-score indicates an average submission date closer in time to the project deadline, and a higher Z-score indicates starting the project earlier. Additionally, we measured exam score as the average of the two partners' total exam Z-scores. We observed that starting project earlier is associated with higher average exam scores.

Next, we see that the relationship between starting early and average project score is also significant, even after controlling for other variables (Table 3). Specifically, partnerships that started earlier tended to have higher project grades. Because partnerships may have spanned multiple projects, we measured project grades for a partnership as an average of the project Z-scores over the duration of that partnership, after confirming that projects had similar distributions.

		Average exam score					**Average project score**			
		SS	**df**	**F**	**PR(>F)**		**SS**	**df**	**F**	**PR(>F)**
Parity		0.01	1	0.03	0.871		0.72	1	2.72	0.100
Early-start		2.20	1	8.70	0.003		2.91	1	11.00	0.001
GPA		61.51	1	242.99	0.000		30.61	1	115.84	0.000
Early-start:Parity		0.06	1	0.26	0.613		0.67	1	2.51	0.114
Parity:GPA		0.16	1	0.65	0.422		0.77	1	2.92	0.088
Early-start:GPA		0.04	1	0.15	0.698		0.04	1	0.13	0.715
Parity:Early-start:GPA		0.33	1	1.28	0.268		0.53	1	1.99	0.159
Residual		124.30	491	-	-		129.750	491	-	-

Table 3: ANOVA with independent variables parity, early-start, and average GPA of the partnership. Dependent variables were average exam score and average project score. Starting early had a significant relationship with both exam and project score, after controlling for GPA.

3.2 Partnership Early-start and Course Letter Grade

In Section 3.1, we saw that starting early was associated with both exam and project performance. In this section, we shall quantify these results by examining students' overall course letter grades.

Table 4 shows partnership course letter grades for each quartile of GPA and early-start. In CS2, the mean course letter grade was a B (3.1) and the standard deviation was approximately 2/3 of a letter grade (0.7). From this table, we see that students starting earlier tended to have a higher overall course grade, regardless of their GPA. However, we noticed the most improvement in course letter grade for students whose GPA fell below the median. Specifically, we saw that for these students, the difference in mean letter grade between the lowest and highest early-start quartiles was 0.4. Additionally, these results were found to be statistically significant. Note that early-start is represented as a Z-score due to the variation in project length.

	Early-start Q1	**Early-start Q2**	**Early-start Q3**	**Early-start Q4**
GPA Q1	C+ (2.3)	C+ (2.4)	C+/B- (2.6)	B- (2.7)
GPA Q2	B- (2.8)	B (3.0)	B (3.0)	B+ (3.2)
GPA Q3	B+ (3.2)	B+ (3.3)	B+ (3.3)	B+/A- (3.5)
GPA Q4	B+/A- (3.6)	A- (3.7)	A- (3.7)	A- (3.7)

Table 4: Mean course letter grade, grouped by GPA and early-start quartiles. We notice that students with GPAs below the median were associated with the largest benefit from starting projects early. For example, for students in the lowest GPA quartile, starting early made the difference between a final grade of C+ and B-.

3.3 Gender and Partnership Performance

We now examine the relationship between the gender composition of a partnership and student performance. In our analysis, we found the association between gender composition and exam performance to be statistically significant, while controlling for GPA. The association between gender composition and project scores was not statistically significant, again while controlling for GPA (Table 5). Specifically, we observed that partnerships of two men tended to perform slightly better on exams than other partnerships. On average, partnerships of two men had an average exam Z-score of 0.031, while mixed gender partnerships average -0.138, and partnerships of two women averaged -0.241.

		Average exam score					**Average project score**			
		SS	**df**	**F**	**PR(>F)**		**SS**	**df**	**F**	**PR(>F)**
Gender Composition		4.21	2	8.51	0.000		0.18	2	0.34	0.713
GPA		72.82	1	294.30	0.000		34.67	1	125.81	0.000
Gender Composition:GPA		0.91	2	1.84	0.160		0.07	2	0.13	0.878
Residual		121.98	493	-	-		135.87	493	-	-

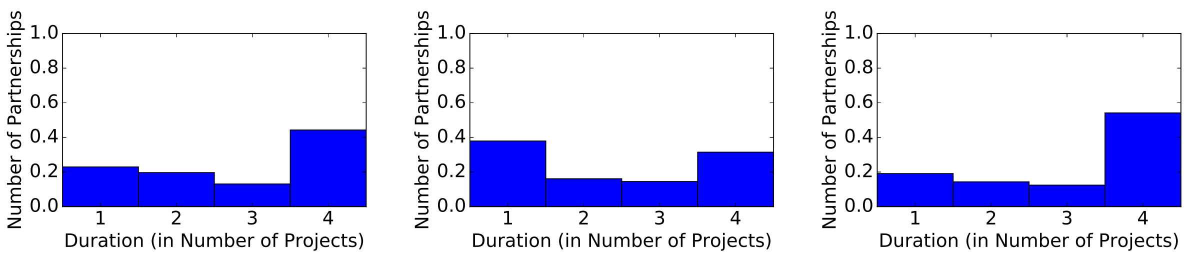
Table 5: ANOVA with independent variables gender composition and average GPA of the partnership. Dependent variables were average exam score and average project score. Gender composition had a statistically significant association with exam score, after controlling for GPA.

To interpret the impact of our project and exam observations, we also examine the association between gender composition of partnerships and overall course letter grade on a 4.0 scale. We found that on average, men who partnered with other men on projects tended to perform slightly better in the course, although the impact was less than a plus/minus.

	Two Men	Mixed Gender	Two Women
GPA Q1	C+/B- (2.6)	C+ (2.4)	C+/B- (2.5)
GPA Q2	B (3.0)	B-/B (2.9)	B-/B (2.9)
GPA Q3	B+ (3.4)	B+ (3.3)	B+ (3.2)
GPA Q4	B+/A- (3.6)	B+/A- (3.5)	B+ (3.4)

Table 6: Mean course letter grade, grouped by gender composition and GPA quartile. We notice that partnerships of two men were associated with slightly higher letter grades. However, we also note that the impact was less than a plus/minus.

Finally, we observed a relationship between the gender composition and partnership duration. In this analysis, partnership duration is measured as the number of projects that a partnership worked on together. We found that the majority of same-gender partnerships spanned all four projects, while the majority of mixed gender partnerships spanned either one or four projects. This relationship is significant with a p-value of 0.000007, and can be seen in Figure 2.



(a) Two women (N = 61) **(b)** Mixed gender (N = 124) **(c)** Two men (N = 314)

Figure 2: Distribution of partnership duration, by partnership gender composition. The x-axis is the duration of the partnership in number of projects. The y-axis is the normalized to the N in each group. We observe that same-gender partnerships tended to last all four projects, while mixed-gender partnerships often lasted either one (the minimum) or four (the maximum) projects.

4 Discussion

Through our analysis, we observed several statistically significant relationships between partnership composition and student performance. Overall, we found a positive association between starting projects early and exam and project scores, even after controlling for GPA. We also found that the gender makeup of a partnership plays a role in how individuals in a partnership score on exams, as well as on the duration of a partnership. Furthermore, we found that partnership parity was not a statistically significant factor.

Next we discuss our interpretation of our statistical analysis, and then present some of the limitations of our study.

4.1 Starting Early

First, we discuss the relationship between starting early (measured in days between the first project submission and the deadline) and partnership performance. Again, note that early-start is represented as a Z-score because of the variation in duration between projects. In Section 3.1, we found that partnerships who started projects early tended to perform better on both exams and projects compared to those who started later. Additionally, we found these results to be statistically significant, even after controlling for GPA (Table 3). We also point out that although this result is significant, the proportion of variance explained by starting early is small compared to the contribution of average GPA of a partnership.

The impact of starting early is observable in final letter grades. This is a logical extension of our observations that starting early is associated with both higher project and higher exam scores. We observed the greatest impact among students with GPAs below the median. For example, for those in the lowest GPA quartile, starting projects early made the difference between a C+ (lowest early-start quartile) and a B- (higher early-start quartile). These results may imply that students who started projects earlier learn the material better. Thus, even when working alone on an exam, they perform better.

We note that starting early could have a positive impact on student performance and learning, regardless of whether or not students worked in a partnership. Additionally, it could be the case that all students, regardless of their choice to partner, benefit from starting their work early.

4.2 Gender

Next, we discuss the association between gender balance and partnership performance. As seen in Section 3.3, partnerships of men tended to score slightly better on exams compared to others. As a result, these partnerships tended to receive slightly higher overall course letter grades, although the magnitude was less than a plus/minus. According to the multivariate ANOVA found in Table 5, the contribution of gender to exam score is significant even after considering the contributions of the average GPA of partnerships. We also point out that although this result is significant, the proportion of variance in exam scores explained by gender composition is small compared to the contribution of average GPA of a partnership.

In addition, we observed that same-gender partnerships were more likely to span all four partner projects, while mixed-gender partnerships tended to span either one or four projects. While we do not have any data on romantic relationships within this course, it seems possible that the high number of mixed-gender, short duration partnerships could be the result of the mortality of romantic relationships within the course. Another possibility is a negative group dynamic in mixed gender partnerships that causes them to end earlier.

4.3 Limitations

Because our study was observational in design, rather than experimental, there are a number of factors out of our control that could affect the validity of our results.

First, our parity metric could be affected by class standing. Students of higher class standing have more datapoints included in their GPA calculation than students of lower class standing, and therefore there would be less variance in a senior's GPA than a freshman's. Our parity metric is calculated directly from partnership GPAs. We believe that this limitation is less of a concern as we also evaluated the difference in previous programming course grades as a parity metric, and found similar results.

Second, students had the choice whether to partner or work alone, and further, students could choose their partner. Because of this, we are not able to control the composition of students who chose to partner or any biases students may have had when selecting partners. We also do not have data on the dynamics of student interaction within a partnership.

Finally, our data set is compiled from multiple semester offerings of the introductory computer science course. Although the content presented in lecture remains largely the same between course offerings, there could be some differences in presentation. Additionally, while the content tested on exams remained largely the same between semester offerings, the exact questions were different.

5 Conclusions

With this study, we aim to provide insights about the defining characteristics of a successful partnership. We examined data from 509 partnerships of students enrolled in an introductory computer science course at a highly ranked public university.

In our results, we observed a relationship between starting early and course performance. On the other hand, parity of partner GPAs was not a significant factor. We also point out that the average GPA of a partnership plays a significant role. After controlling for GPA, We found that partnerships who began working on a project early were associated with both higher project scores and higher exam scores. Students with below-median GPAs were associated with the greatest gains from starting projects early. For example, students in the lowest GPA quartile and the lowest early-start quartile averaged a C+ in the course, while those from the same GPA quartile, but the highest early-start quartile averaged a B-.

We also found that the duration of a partnership was associated with the gender composition of the partnership. Same gender partnerships often spanned the entire semester, while mixed gender partnerships tended to span either only one project or the entire semester. In future work, we hope examine the dynamics of partnerships with different gender compositions.

References

[1] Charlie McDowell, Linda Werner, Heather Bullock, and Julian Fernald. The effects of pair-programming on performance in an introductory programming course. In *Proc. SIGCSE*, pages 38–42, 2002.

[2] Emilia Mendes, Lubna Basil Al-Fakhri, and Andrew Luxton-Reilly. Investigating pair-programming in a 2nd-year software development and design computer science course. In *Proc. SIGCSE Conference on Innovation and Technology in Computer Science Education*, 2005.

[3] Emilia Mendes, Lubna Al-Fakhri, and Andrew Luxton-Reilly. A replicated experiment of pair-programming in a 2nd-year software development and design computer science course. In *Proc. SIGCSE Conference on Innovation and Technology in Computer Science Education*, 2006.

[4] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686–688, 2010.

[5] Nachiappan Nagappan, Laurie Williams, Miriam Ferzli, Eric Wiebe, Kai Yang, Carol Miller, and Suzanne Balik. Improving the CS1 experience with pair programming. In *Proc. SIGCSE*, 2003.

[6] Neha Katira, Laurie Williams, and Jason Osborne. Towards increasing the compatibility of student pair programmers. In *Proc. ICSE*, 2005.

[7] Norsaremah Salleh, Emilia Mendes, John Grundy, and Giles St. J. Burch. An empirical study of the effects of personality in pair programming using the five-factor model. In *Proc. SIGCSE International Symposium on Empirical Software Engineering and Measurement*, 2009.

[8] J. E. Hannay, E. Arisholm, H. Engvik, and D. I. K. Sjoberg. Effects of personality on pair programming. *IEEE Transactions on Software Engineering*, 36(1):61–80, Jan 2010.

[9] Andrew DeOrio and Andrew Giugliano. Long-term effects of partner programming in an introductory computer science sequence. In *Proc. ASEE*, 2016.