

Group Work in the Science Classroom: How Gender Composition May Affect Individual Performance

Dana Gnesdilow, Amanda Evenstone, Julia Rutledge, Sarah Sullivan, & Sadhana Puntambekar
University of Wisconsin, Madison, WI

Email: gnesdilow@wisc.edu, alevenstone@wisc.edu, jrutledge@wisc.edu, sasullivan2@wisc.edu,
puntambekar@education.wisc.edu

Abstract: Current research on how gender composition within groups influences individual outcomes is both sparse and conflicting. We examined how gender composition within groups affects learning outcomes. Students from sixth, seventh, and eighth grade classes from three US Midwestern public school districts with diverse demographic compositions (N=637, 314 boys and 323 girls) participated in this study as a part of their regular science class during a 12-week design-based physics curriculum, CoMPASS. We conducted two 5 x 2 analyses of covariance to evaluate the effect of group gender ratio and gender on students' physics learning and science practice outcomes. Results indicate that group gender ratio does influence students' science learning and practices as measured by posttest differences. Students in mixed-gender groups performed significantly better than students in same-gender groups. Having at least one group member of the opposite gender increased individual students' posttest performance. Limitations and implications for practice are discussed.

Engaging students in group work during inquiry-based and project-based learning activities has become an increasingly common practice in science classrooms. However, as research suggests, students may not always effectively collaborate in ways that foster learning (Barron, 2003; Rummel & Spada, 2005). Further, collaborative learning may not always result in equivalent learning gains for each individual (Teasley & Fischer, 2008; Gnesdilow, Bopardikar, Sullivan, & Puntambekar, 2010). Several factors such as group size, context, gender, prior knowledge, and individual abilities may affect the collaboration in groups (e.g., Apedoe, Ellefson, & Schunn, 2012; Hawkins & Power, 1999).

In this paper we focus on understanding how the gender composition in groups affects students' learning outcomes in science. The current research on how gender composition in groups influences individual outcomes is both sparse and conflicting. Ding, Bosker, and Harskamp (2011) discussed that while Computer Supported Collaborative Learning (CSCL) has the potential to lessen the gender gap between male and female performance and persistence in physics, the positive findings from CSCL research "are controversial where gender is concerned" (p.325). Leman (2010) pointed out that there is a scarcity of empirical research linking "interactions and collaboration to gender and learning outcomes" (p.218). Research has indicated that there are differences between how boys and girls learn, converse, and interact (Leman, 2010; Kommer, 2006; Rice & Dolgin, 2002), including when within mixed-gender groups (Hawkins & Power, 1999) and also within mixed-gender dyads (Ding, Bosker, & Harskamp, 2011; Harskamp, Ding, & Suhre, 2012). Some studies have found that girls in mixed-gender groups do not perform as well as girls in same-gender groups (e.g. Light, Littleton, Bale, Joyner & Messer, 2000). Similarly, other studies have revealed that high school girls learning physics in mixed-dyads scored significantly lower on posttests than the boys working in the mixed-dyads, as well as the boys and girls who worked in same-sex dyads (Ding et al., 2011; Harskamp et al., 2012). Alternatively, one of the key findings highlighted by Bennett, Hogarth, Lubben, Campbell, and Robinson's (2010) review of studies of small groups in science classrooms was that students in single-sex groups were more purposeful than mixed-gender groups, but ultimately group gender composition did not affect understanding. In another study, girls participated more actively and persistently on collaborative learning activities when in mixed-gender groups, including generating more science and group orchestration talk during computer-based learning activities (Goldstein & Puntambekar, 2004).

Given the contradictions between the findings outlined above, as well as the lack of overall evidence about how gender composition affects students' learning in groups, we believe that understanding these relationships could lead to strategic and easy-to-implement teaching decisions for enhancing collaboration and learning. In this study we examined how gender composition in groups affects students' learning outcomes and attempt to answer the research question: Do differences in gender composition affect middle school science students' learning in groups? We explored this question by examining students' science content knowledge and practices outcomes.

Methods

Participants and Instructional Context

Two hundred sixth grade, 143 seventh grade, and 294 eighth grade students (N=637, 314 boys and 323 girls) from three US Midwestern public school districts with diverse demographic compositions participated in this study as a part of their regular science class. All students took part in the CoMPASS roller coaster unit, a 12-week design-based science curriculum, to learn about forces, motion, work, and energy. They participated in a variety of physical science activities in order to design a fun, safe, and efficient roller coaster for an amusement park whose attendance is waning. Students worked in the same group of three or four throughout the 12-week unit (Group N=178, 54 sixth, 41 seventh, and 83 eighth grade groups), with group composition determined prior to this study by our collaborating teachers. Students took separate pre- and posttests for science content and practices (described below). Students took these tests before starting and after finishing the CoMPASS roller coaster curriculum in their classes.

Data Sources and Analysis

Measures

We used two tests: the Physics Fiesta measured students' content knowledge in physics and the Scientist's IQ tested science practices. The Physics Fiesta consisted of 29 multiple-choice questions and addressed a range of physics concepts and relationships such as mass, work, force, potential and kinetic energy, velocity, acceleration, efficiency, the law of conservation of energy, and Newton's Laws. Each correct item earned a score of one point and incorrect answers were scored as zero, with 29 points being the highest score possible. The Scientist's IQ consisted of 13 multiple-choice and five open-ended questions that assessed students' skills in areas such as interpreting, making inferences, setting up data in graphs and charts, hypothesis writing, variable identification in setting up experiments, using data to back up reasoning and explanations, and identifying measurement and other sources of error in investigational scenarios. Correct multiple-choice responses on the Scientist's IQ earned one point and incorrect responses were scored as zero. The open-ended questions were graded from 0 to 2 or 3 points. A score of 2 (or 3) indicated a more sophisticated, elaborate response or explanation, while a zero indicated that the answer was incorrect, blank, or unintelligible. Answers coded as 1 or 2 points (for 3-point questions) were correct but not explained well, supported, or were partial responses. The maximum score for the Scientist's IQ was 24 points. Interrater reliability for scoring the open-ended responses on the Scientist's IQ pretest was 94.35% and 92.5% for the posttest.

Gender Ratio Group Categories

To examine how the gender composition of groups influenced each student's learning outcome, we used five different Gender Ratio categories. The five categories were: 1) all boys, 2) mostly boys (i.e. 2 or 3 boys in a group of 3 or 4, respectively), 3) even split between boys and girls, 4) mostly girls (i.e. 2 or 3 girls in a group of 3 or 4, respectively), and 5) all girls. Based on the gender composition of the group that a student worked in throughout the CoMPASS curriculum, he or she was labeled as belonging in one of the five categories. For example, if a group consisted of two girls and one boy, each of the three students was labeled as belonging to a mostly girl group. Due to missing data, a total of 574 students (280 boys and 294 girls) completed both pre- and posttests for the Physics Fiesta and 530 students (259 boys and 271 girls) completed both pre- and posttests for the Scientist's IQ. Only the scores of students who completed both pre- and posttests for a given measure were included in our analysis.

Results

We conducted two 5 x 2 analyses of covariance (ANCOVA) to evaluate the effects of group Gender Ratio and Gender on students' scores on the Physics Fiesta and Scientist's IQ tests. The independent variable, Gender Ratio, included five levels: all boys, mostly boys, even split, mostly girls, and all girls. The other independent variable was Gender. The dependent variables were the Physics Fiesta posttest score and the Scientist's IQ posttest score. To meet the assumptions of ANCOVA, we established that the Physics Fiesta pretest score was significantly related to the posttest score, $F(1, 565) = 155.014, p < .001$ and the Scientist's IQ pretest score was significantly related to the posttest score, $F(1, 521) = 895.880, p < .001$. Thus, the relationship between the covariates and their respective dependent variables did not differ as a function of the independent variables.

Table 1 displays the descriptive statistics of each posttest. ANCOVA results showed there was no main effect for Gender, a significant main effect for Gender Ratio, and no interaction. The ANCOVA indicated no significant difference in posttest performance of boys and girls on the Physics Fiesta posttest after controlling for pretest score, $F(1, 565) = .561, p = .454$. Comparably, there was no significant difference between boys and girls in the Scientist's IQ posttest after controlling for pretest score, $F(1, 521) = .154, p = .695$. When comparing differences in mixed-gender and same-gender groups, it was important to find that gender alone was not a significant predictor in posttest performance. Boys did not perform significantly better than girls and

therefore skew group means. Since there was no effect of Gender, we know that the effect of Gender Ratio on posttest scores cannot be attributed to performance based on a specific gender.

The ANCOVA further indicated that there was no interaction between Gender and Gender Ratio on the Physics Fiesta posttest score when accounting for pretest score, $F(2, 565) = .304, p = .738$. Controlling for pretest score, there was also no interaction between Gender and Gender Ratio on the Scientist's IQ posttest score, $F(2, 521) = .454, p = .635$. Results showing no interaction for either test suggest that Gender did not change posttest score based on group composition and indicates that posttest score effects are based solely on Gender Ratio. Because there was no effect of Gender or interaction between Gender and Gender Ratio over three different grades and three different school districts, our results support the idea that group composition is a more important determinant of posttest score and that this result does not vary based on the student's gender.

Gender Ratio

There was a significant effect of Gender Ratio on the Physics Fiesta posttest after controlling for the effect of Physics Fiesta pretest score, $F(4, 565) = 3.024, p = .017$. The students in the mostly girl groups had the largest adjusted mean ($M = 17.431$), and students in the even split groups had the next largest adjusted mean ($M = 17.238$), followed by the third largest adjusted mean in mostly boy groups of students ($M = 16.960$). The two lowest adjusted means were observed in students in all boy ($M = 16.220$) and all girl groups ($M = 15.854$).

There was also a significant effect of Gender Ratio on the Scientist's IQ posttest after controlling for the pretest, $F(4, 521) = 2.680, p = .031$. Similar to the Physics Fiesta test, the mostly girl groups of students had the largest adjusted mean score ($M = 15.140$), followed by students in mostly boy groups ($M = 14.942$) and then students in even split groups ($M = 14.920$). The groups of students with all girls ($M = 14.592$) and groups of students with all boys ($M = 13.436$) had the lowest adjusted means on the posttest. The Gender Ratio main effect indicated that the mixed-gender groups of mostly girls, even split of boys and girls, and mostly boy groups tended to have higher posttest scores than the same-gender groups that had all girl or all boy students.

Table 1: Descriptive Statistics for posttest score by Gender Ratio

Posttest	Gender Ratio	Posttest score			
		Observed Mean	Adjusted Mean	SE	n
Physics Fiesta	All Boys	16.03	16.220	.452	73
	Mostly Boys	16.87	16.960	.414	104
	Even Split	17.61	17.238	.268	211
	Mostly Girls	17.63	17.431	.412	106
	All Girls	15.03	15.854	.437	80
Scientist's IQ	All Boys	12.14	13.436	.426	70
	Mostly Boys	14.03	14.942	.400	91
	Even Split	15.89	14.920	.258	192
	Mostly Girls	15.37	15.140	.383	105
	All Girls	13.97	14.592	.418	72

Post Hoc Analysis

In order to explore the posttest differences based on Gender Ratio, we conducted contrasts to observe any potential differences between mixed-gender and same-gender group composition. To do this, we grouped the means of all of the heterogeneous gender groups (mostly girl, even split, and mostly boy groups) into a mixed-gender category. Then, we combined the means of the all girl and all boy groups together into a same-gender category. Contrasts revealed that Physics Fiesta posttest scores for students in mixed-gender groups were significantly higher than for students in same-gender groups, $p < .01$. Likewise, mixed-gender groups of students had higher Scientist's IQ posttest scores than same-gender groups of students, $p < .01$.

Discussion

In attempting to answer our research question, our analysis supports the idea that group gender composition does influence students' science learning and practices as measured by the Physics Fiesta and Scientist's IQ, respectively. We found that students in mixed-gender groups outperformed students in same-gender groups on both the content and practices posttests, when controlling for pretest score. No difference at the gender level may indicate that students' individual success may be better explained by gender composition within a group. Our results contrast with Light et al.'s (2000) finding that females perform better in same-gender versus mixed-gender groups. And, while group size may be a confounding factor in making clear-cut comparisons between studies, our results also seem to conflict with studies that show that girls in mixed-gender dyads scored lower than their boy counterparts on physics posttests (Ding et al., 2011; Harskamp et al., 2012). In addition, we found

that students in mixed-gender groups with at least two girls performed slightly better than students in other groups, and that students in same-gender groups were less successful. From these results, one might think that collaborative work within this science context may favor the skillset of female students. However, if girls are the key, then students in all girl groups should outperform students in other groups, but this was not the case. Presence of at least one member of the opposite gender increased individual students' posttest performance in both science content and practices. These are nuanced dynamics; therefore, we will need to qualitatively examine the discourse of students in groups of different gender compositions to better understand and elaborate on how these outcomes may have occurred.

Gender is an important factor in collaboration (Leman, 2010), and Kommer (2006) stressed that it is important for teachers to understand how to organize groups to optimize students' strengths. The gender composition of groups may be a key factor and certain ratios may prove to be more beneficial than others to foster students' learning and collaboration. While we had a large number of students, one limitation of this study is that it does not include process data to help explain why these findings may have occurred. There are other factors that could result in differences in students' performance, such as the influence of different classroom contexts, teacher variables, grades, or districts. It will be important to explore these factors in future studies, as well as use group process data, to understand what is qualitatively different in the interactions of successful mixed-gender groups so that these findings may be applied to all groups in a classroom, no matter the gender composition. In combination with our future work on analyzing groups' discourse, our findings may inform teaching decisions about how to structure CSCL group composition in practical and simple ways.

Acknowledgements

The research described in this paper is supported by the Bill & Melinda Gates Foundation Next Generation Learning Challenges Wave 2 grant program and EDUCAUSE.

References

- Apedoe, X.S., Ellefson, M.R., & Schunn, C.D. (2012). Learning together while designing: Does group size make a difference? *Journal of Science Education and Technology*, 21(1), 83-94.
- Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, 12(3), 307-359.
- Bennett, J., Hogarth, S., Lubben, F., Campbell, B., & Robinson, A. (2010). Talking science: The research evidence on the use of small group discussions in science teaching. *International Journal of Science Education*, 32(1), 69-95.
- Ding, N., Bosker, R.J., & Harskamp, E.G. (2011). Exploring gender and gender pairing in the knowledge elaboration processes of students using computer-supported collaborative learning. *Computers & Education*, 56, 325-336.
- Gnesdilow, D., Bopardikar, A., Sullivan, S.A., & Puntambekar, S. (2010). Exploring convergence of science ideas through collaborative concept mapping. In Gomez, K., Lyons, L., & Radinsky, J. (Eds.) *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers*, 698-705. International Society of the Learning Sciences: Chicago, IL.
- Goldstein, J., & Puntambekar, S. (2004). The brink of change: Gender in technology-rich collaborative learning environments. *Journal of Science Education and Technology*, 13(4), 505-522.
- Hawkins, K., & Power, C.B. (1999). Gender differences in questions asked during small decision-making group discussions. *Small Group Discussion*, 30(2), 235-256.
- Harskamp, E., Ding, N., & Suhre, C. (2008). Group composition and its effect on female and male problem-solving in science education. *Educational Research*, 50(4), 307-318.
- Kommer, D. (2006). Boys and girls together: A case for creating gender-friendly middle school classrooms. *The Clearing House*, 79(6), 247-251.
- Leman P.J. (2010). Gender, collaboration and children's learning. In K. Littleton & C. Howe (Eds.), *Educational Dialogues: Understanding and Promoting Productive Interaction* (pp. 216-239). New York, NY: Routledge.
- Light, P., Littleton, K., Bale, S., Joiner, R., & Messer, D. (2000). Gender and social comparison effects in computer-based problem solving. *Learning and Instruction*, 10, 483-496.
- Rice, F.P., & Dolgin, K.G. (2002). *The Adolescent: Development, Relationships, and Culture* (10th ed.). Boston: Allyn and Bacon.
- Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem-solving in computer-mediated settings. *Journal of the Learning Sciences*, 14(2), 201-241.
- Teasley, S.D., & Fischer, F. (2008) Cognitive Convergence in Collaborative Learning. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F. Prins, (Eds.), *International perspectives of the learning sciences: Creating a learning world. Proceedings of the Eighth International Conference of the Learning Sciences (ICLS 2008), Vol 3* (pp. 354-359). International Society of the Learning Sciences, Inc.