Effects of Using Single-Gender Group Exams in a Large, Introductory Geology Class

Rebecca K. R. Ambers

Contact Information:
Dr. Rebecca Ambers
Dept. of Environmental Studies
Sweet Briar College
Sweet Briar, VA 24595
office phone: 434-381-6483
home phone: 434-381-5922
fax: 434-381-6488
email: rambers@sbc.edu

Abstract

Studies have shown that women do better and feel more comfortable in science classes when cooperative learning techniques are used and when groups contain a critical mass of women. Group exams are one way to incorporate cooperative learning into a large-enrollment course. Trials of this method in a large introductory geology class indicate that making groups all-male and all-female helps many women and also some men to feel more comfortable with group work. Single-gender group exams thus provide multiple benefits for students in large classes.

Keywords: group tests, single-gender, large enrollment

Introduction

Large-enrollment courses tend to be difficult for teachers and students alike. Teachers struggle with the management of so many students and often feel that they cannot run the class the way they would in a small-enrollment course. Many students feel lost in the crowd, disconnected from the teacher, and more like audience members in a theater than participants in a classroom. To counteract some of these problems, many instructors have begun incorporating cooperative learning activities into their large classes (e.g., Macdonald and Korinek, 1995; Ebert-May et al., 1997; Wyckoff, 2001). These activities can take many forms, from think-pair-share exercises to group projects. Students were reported by Bykerk-Kauffman (1995) to be more enthusiastic about group exams than any of the other cooperative activities she tried.

Another issue that teachers may struggle with is how to make science a more inviting subject for women, both in terms of improving science literacy and for recruiting majors into fields in which women are traditionally underrepresented (Rosser, 1995). Studies have shown that one of the most effective ways to increase women's comfort level and performance in

science is to use cooperative learning methods in the classroom and laboratory (Rosser, 1992). The make-up of groups is critical in creating a positive learning experience for women, however (Light, 1990). Although it is commonly recommended that teachers choose groups to maximize the gender and racial diversity within each group (e.g., Slavin, 1990), this can be a poor strategy for women and minorities in fields like science that are nontraditional career choices (Rosser, 1997). Women may feel isolated or excluded if they are the only female in a group, at least in part because men have a tendency to interrupt women and dominate classroom discussion (Hall and Sandler, 1982). A "critical mass" of women in each group is important (Etzkowitz et. al, 1994). Taking this idea even farther, "the female-only environment gives women an equal chance," and "cooperative techniques...have proved particularly successful in all-female environments" (Rosser, 1997, p. 56). Most colleges and classrooms are coeducational, however, so special steps have to be taken to create comfortable "microclimates" for women in science courses.

As a teacher handling large, introductory courses, I wondered if giving group exams using single-gender groups would be the optimal way to both encourage active learning and level the playing field for women in the coeducational science classroom. During the Spring 2001 semester at Winona State University in Minnesota, I experimented with this method in the lecture portion of a physical geology course entitled Dynamic Earth. The course had 130 people enrolled, of whom 81 were women and 49 were men. I kept track of the scores of individuals and their groups and administered a detailed student evaluation of the group exam format at the end of the semester. In these data, I found differences between the performance of men and women, as well as individuals and groups. Men and women also showed differences in their degree of comfort with this method, although most students were very positive about it. While

not giving a definitive answer about whether single- or mixed-gender groups are superior, this study provides information that may help other teachers decide whether or not they should start using single-gender groups for cooperative learning exercises in large-enrollment science courses.

Group Exam Format

The group exam format I used is similar to that described by Bykerk-Kauffman (1995) and Mouton and Blake (1975). To eliminate confusion during the test, students sat with pre-assigned, single-gender groups of 4-5 students in pre-assigned seats on exam days. I first administered a 25-question, multiple-choice exam to individuals with no talking allowed. Students were asked to stay seated and quiet and hold onto their score sheets until everyone was done. Once the individual score sheets were taken up, each group was given a new score sheet, and students took the same exam again with their group. Consensus was required on the answer to each question because only one score sheet could be turned in per group. With only 25 questions on each exam (half the number I had previously used when doing only individual exams), I found it easy to finish the whole exam process in one 50-minute class period. I also felt comfortable making the exam questions more challenging and thought-provoking because students would have an opportunity to discuss the questions with others.

Because this testing procedure is relatively complex compared to a typical exam, at the end of the first week of class, I gave students a 10-question pre-test for no credit. They took the test in exactly the same way as they would the real exams. I think a "practice run" like this is essential to get students comfortable with the group exam format and allow them to identify their assigned seats and group members before the first exam.

I computed students' exam grades, referred to below as their combined scores, as a weighted average of the two scores: 75% individual and 25% group. If students had an excused absence on an exam day, they took the test only as individuals, and this score counted for 100% of their grade. For the final exam, students took only the half of the test that covered new material in the group exam format. They took the comprehensive review portion only as individuals, so 12.5% of the total grade came from the group score. Keeping track of grades became somewhat complex, but I felt it was important that the final exam reflect individual performance as much as possible without sacrificing the educational value of the group exam format.

If a group disagreed with me on the answer to one or more exam questions, it had one week after the graded exams were returned to file a petition for reconsideration. Each petition had to include the group members' names, the question number(s) being petitioned, a persuasive argument why their answer was correct and mine was not (hopefully this included a reference to pages in the text that supported their argument), and the individual and group score sheets. Only groups who submitted a petition that I approved received credit for the question; others had no change in their score. Many students appreciated the opportunity to petition a question because they felt it made the testing procedure fairer (Table 2). I also found the petitioning process valuable, although it made keeping track of grades even more complex. It enabled me to see where students had conceptual difficulty with the material and how they misinterpreted the wording of my multiple-choice questions. This information allowed me to improve my test-writing abilities in an attempt to create questions that were "petition-proof."

In assigning students to groups for the first time, I chose to keep students in the same lab section together so that they would be working with people they saw regularly in a smaller class

setting. Not knowing the students personally, I simply grouped them alphabetically in 32 single-gender groups of 4-5 students. After the second exam, I shuffled students around into a second set of 30 groups. This time, I attempted to place in each group at least one student who averaged a B+ or better on the previous exams; I also tried to distribute the academically weaker students fairly evenly throughout the groups. The composition of the lab sections required me to create a few male groups with people from more than one section. For the final exam, I polled students on whether they preferred to work in the first or second set of groups. The vote was fairly close, but the majority chose to work with the first set again.

An important issue that I had to resolve with this exam format was how to handle students with learning disabilities and/or special exam-taking needs. Such needs may include a low-distraction environment, having extra time, and taking exams orally or with computer assistance. The solution I found was to allow such students to take the individual portion of the exam in the campus testing facility a day or two before the regular exam period. They then came to class, sat quietly through the individual portion of the exam, and participated in the group portion along with everyone else. In exchange for testing accommodation, I found that these students readily agree not to discuss the exam questions with other class members ahead of time.

Testing Outcomes

To analyze student performance on the exams in my class, I calculated the mean and standard deviation of scores for male and female individuals and groups (Figure 1). I also performed a two-sample t-test to test the hypotheses that groups scored better than individuals and that males scored better than females (Table 1). To learn more about group dynamics, I compared the maximum individual score within each group to that group's score (Figure 2).

Individual Versus Group Scores

On all five exams, the groups did significantly better on average than individuals ($p < 0.0005$; Table 1 and Figure 1). The average group score was 12-19 points higher than the average individual score, so the general effect of the group exam on student grades was positive. By weighting the group score at 25%, student's combined scores were only raised an average of 4 points above their individual scores. The maximum increase was 14 points, but this situation only occurred when a student did miserably as an individual yet happened to be in a group that earned a very high grade. Although some teachers may be uncomfortable with this situation, I find it more important that these students actually did get exposed to the correct answers to most of the questions during group discussion. This is more than can be said of many individual exams. The usually small increase in scores seemed to be enough to make students feel good about taking group tests, but I did not find it sufficient to change grade distributions or to keep very weak students from earning poor marks.

One line of evidence showing that cooperative learning was occurring during group work is that in 37-60% of the groups on each exam, the group score was better than the maximum individual score within the group (Figure 2). Students, at least in these groups, were pooling their knowledge and not simply using the answers of the one person they considered the "smartest." These high percentages support the idea that the group exam is an effective cooperative learning method.

On the other hand, 6-25% of the groups earned a lower score than the highest scoring individual within the group (Figure 2). In all of these cases, the difference was only 1-2 more questions wrong on the group exam than the best individual test. The people who earned these lower group scores may have been ineffective at convincing their group members that they knew

the right answers, or perhaps they were in groups with very persuasive but less knowledgeable members. Only four people in the class allowed this to happen to them on more than one exam.

In an interesting illustration of group dynamics, the percentage of groups that earned a lower score than their maximum individual score dropped dramatically the second time students worked in a particular set of groups. For Exams 1 and 3, the value was 22-23%; but for Exams 2 and 4, it was 6-7% (Figure 2). After one exam together, either the group members realized who they needed to listen to, or the high-scoring students learned how to make themselves be heard. Surprisingly, on the final exam when students returned to their original set of groups, this percentage reached a high of 25%. The implication is that if a particular group does not work together for several weeks, they must effectively start over again in their group dynamics. The teacher's goals for the cooperative learning experience are thus relevant in deciding whether it is better to have students work in the same groups throughout the term or to mix them up periodically. Because something is learned in both situations, I believe that the choice is not critical to the success of this cooperative learning method.

Men's Versus Women's Scores

On all the exams in this particular class, men scored higher on average than women: 3-9 points higher individually and 1-4 points higher in groups (Figure 1). Although at least one man and one woman always got a combined score of 97 or 100%, the minimum combined score for women was 2-17 points lower than for men. Performance of two-sample t-tests shows that male scores were significantly better than female scores ($p < 0.05$) only on the individual portion of Exams 2, 3, and 4 and the final, however (Table 1). In an effort to determine why an individual score difference occurred, I put two questions on the end-of-term evaluation that probed

students' early interest in geology-related subjects and the confidence they had in their ability to do well in science (questions 9 and 10, Table 2). The results indicate that more of the men brought strong interest and high confidence levels to the class. Without further information, I cannot say whether this background accounts for the difference in individual performance or if differences in study habits also played an important role.

Although women's individual scores were lower on most exams, their groups appeared to function more effectively than men's groups. In other words, women's groups seemed to pool their knowledge better and work more cooperatively. Average female group scores were 13-21 points higher than the average female individual scores, whereas the average male group scores were only 10-16 points higher than average male individual scores. As a result, women's group scores were not significantly different from men's on any of the exams (Table 1 and Figure 1). (Note that weighting the group score more heavily in the combined score would have helped to equalize men's and women's average test grades even further.) Also, 44-72% of women's groups earned a higher score than their maximum individual score, an indication that cooperative learning was occurring; and this percentage showed marked increase on the last two exams (Figure 2). In contrast, 25-58% of men's groups achieved a higher score, but the percentage showed an overall decline during the term. As would be predicted from previous studies (Rosser, 1997), women appear to have come in with better cooperative skills and strengthened them during the semester. Men, on the other hand, seem to have had more difficulty working cooperatively. Their groups were more likely to earn the same score as the maximum individual score within the group (Figure 2), possibly indicating a group strategy in which the high-scoring individual(s) dictated the discussion.

If enjoyment and skill can be correlated, the results of the end-of-term evaluation support

the idea that women worked together better than men (Table 2). While the vast majority of students liked the group exam format, women in the class enjoyed the group work more than men did. A greater number of women also reported that group discussion made the exams less intimidating than they would otherwise have been. While both sexes experienced some degree of frustration with group members who talked too much or too little, approximately 85% of men and women reported that all their group members participated on some level. Men reported a slightly higher incidence of dominant talkers and silent listeners. Perhaps most notably, only a handful of men and no women felt that the group exam format was so unfair, uncomfortable, or otherwise problematic that they would attempt to avoid it in future classes. I have observed that many students who prefer to work independently find cooperative learning to be distasteful in principle and practice, so I think this result is an important indication of the success of the group exam method.

Also encouraging was the response that 95% of students felt that they studied as hard or harder for the group exams than they would have for a purely individual test (Table 2). Apparently this format did not encourage students to rely on others to pull their grade up. Although roughly half the class only studied for exams by themselves, women studied in groups somewhat more frequently than men did. Despite the fact that students knew their group assignments well in advance and had lab with group members, they did not seem particularly motivated to study with their testing groups.

Without being able to compare results for a series of exams with the same students in mixed-gender groups, it is impossible to say for certain if using single-gender groups significantly improved the group exam experience for students. On the end-of-term evaluation, most students (66% of women and 73% of men) reported that they felt equally comfortable in

single-gender groups as they would have in mixed-gender groups. I think it is very significant, however, that 34% of women and 20% of men reported that they felt more comfortable in single-gender groups (Table 2). In contrast, none of the women and only 7% of men (3 students) said they felt less comfortable in single-gender groups. Using single-gender groups thus appears to do no harm and, in fact, to help many men and women feel more comfortable interacting in groups.

Conclusions

Given the more effective cooperation of women's groups and the greater degree of comfort many students feel in an all-female or all-male group, I think that there is good reason to use single-gender groups for at least some cooperative learning exercises in large-enrollment science classes. It takes no more time for a teacher to assign single-gender groups (actually less time than for highly diversified groups), and students do not seem to find them objectionable. With the weight of evidence that has been amassed for the different needs of women and minorities in science classes (Rosser, 1997), it is worth reconsidering the common recommendation to assign groups in a way that mirrors classroom diversity. In fields like science where many students are phobic of the subject matter, it can be important to provide some degree of comfort in the group work environment so that other problems do not get in the way of learning. The results of this study show that single-gender groups appear to provide that extra comfort for many men as well as women.

From these experiences with the group exam, I recommend it as an excellent way to incorporate cooperative learning into a large-enrollment course. Students seem to like the format because they recognize the positive effect it usually has on their grade. I like it because it

reinforces the concepts covered on the test, motivates a high level of student participation in group work, and often leads to improvement over time in a group's ability to work together as a team. As a result, group exams are a real learning experience for students and not simply an assessment tool for the teacher. By using single-gender groups on these exams, even greater benefits can be realized, particularly for women.

References

Bykerk-Kauffman, A. 1995. Using cooperative learning in college geology classes. *Journal of Geoscience Education* 43:306-316.

Ebert-May, D., C. A. Brewer, and S. Allred. 1997. Innovation in large lectures--Teaching for active learning. *Bioscience* 47:601-607.

Etzkowitz, H., Kemelgor, C., Neuschattz, M., Uzzi, B., and Alonzo, J. 1994. The paradox of critical mass of women in science. *Science* 266:51-54.

Hall, R. and Sandler, B. 1982. The classroom climate: A chilly one for women. Washington, DC, Project on the Status and Education for Women, Association of American Colleges.

Light, R. 1990. Explorations with students and faculty about teaching, learning, and students life. Cambridge, MA, Harvard University Press.

Macdonald, R. H., and L. Korinek. 1995. Cooperative-learning activities in large entry-level

geology courses. *Journal of Geoscience Education* 43:341-345.

Mouton, J. and Blake, R. 1975. Instrumented team learning: A behavioral approach to student-centered learning. Austin, TX, Scientific Methods, Inc., 79 p.

Rosser, S. V. 1992. Are there feminist methodologies appropriate for the natural sciences and do they make a difference? *Women's Studies International Forum* 15(5-6):535-550.

Rosser, S. V. 1995. Reaching the majority: Retaining women in the pipeline. *in* Teaching the Majority, S. V. Rosser, ed., New York, Teachers College Press, p. 1-21.

Rosser, S. V. 1997. Re-engineering female-friendly science. New York, Teachers College Press.

Slavin, R. E. 1990. Cooperative learning, theory, research, and practice. Englewood Cliffs, NJ, Prentice-Hall.

Wyckoff, S. 2001. Changing the culture of undergraduate science teaching. *Journal of College Science Teaching* 30:306-312.

Tables

Table 1. Results of two-sample t-tests on exam scores						
Exam Number	*Groups > Individuals (all)*		*Males > Females (individuals)*		*Males > Females (groups)*	
	t	*p*	*t*	*p*	*t*	*p*
Exam 1	9.70	<0.0005	1.09	<0.15	0.59	>0.25
Exam 2	9.86	<0.0005	2.06	<0.025	0.77	<0.25
Exam 3	8.67	<0.0005	3.02	<0.0025	1.45	<0.10
Exam 4	11.74	<0.0005	2.60	<0.01	0.50	>0.25
Final Exam	7.80	<0.0005	3.85	<0.0005	1.47	<0.10

Table 2. End-of-term evaluation of group exam format (95% response rate)

	% of women (n = 79)			% of men (n = 44)		
Statements	*strongly agree/ agree*	*neutral*	*disagree/ strongly disagree*	*strongly agree/ agree*	*neutral*	*disagree/ strongly disagree*
1. I like being able to take each exam in a group as well as individually.	94	5	1	82	4	14
2. Group exams are a positive learning experience for me because I get a chance to talk with my classmates about the course material.	91	8	1	79	14	7
3. I found the exams in this course less intimidating than they might otherwise have been because I knew I would get to take them in a group as well as individually.	76	21	3	61	25	14
4. I feel that the individual and group exam format is a fair way to help improve people's test scores.	97	3	0	84	7	9
5. Because of my experience with group exams in this class, I would not hesitate to sign up for another course in which this same type of exam format was used.	94	6	0	79	14	7
6. During the group exams, all the group members eventually participated in the discussion at some point.	86	6	8	84	5	11
7. During the group exams, I felt that one or two people tended to dominate the discussion and some people rarely contributed at all.	33	31	36	43	21	36
8. I appreciate the opportunity to petition for reconsideration of any test question whose answer I disagree with.	95	5	0	82	14	4
9. Ever since I was a child, I have been interested in subjects like the outdoors, rocks, dinosaurs, and wildlife.	47	32	21	66	23	11
	very strong/ strong	*medium*	*weak/ very weak*	*very strong/ strong*	*medium*	*weak/ very weak*
10. In general, I would rate the confidence I have in my ability to do well in science as:	37	50	13	66	27	7

	more	*equally*	*less*	*more*	*equally*	*less*
11. I feel _____ comfortable taking the group exam in an all-male or all-female group compared to how I would feel in a mixed-gender group.	34	66	0	20	73	7
	harder	*about the same*	*not as hard*	*harder*	*about the same*	*not as hard*
12. Knowing the breakdown of my exam score is 75% individual and 25% group, I believe I studied _____ for the exams in this class compared to a class with only individual exams.	15	80	5	7	89	4
	≥4, 3	*2, 1*	*0*	*≥4, 3*	*2, 1*	*0*
13. Before _____ of the exams in this course, I studied with member(s) of my testing group.	1	18	81	11	18	71
14. Before _____ of the exams in this course, I studied with classmates who were not members of my testing group.	23	37	40	23	25	52

Figures

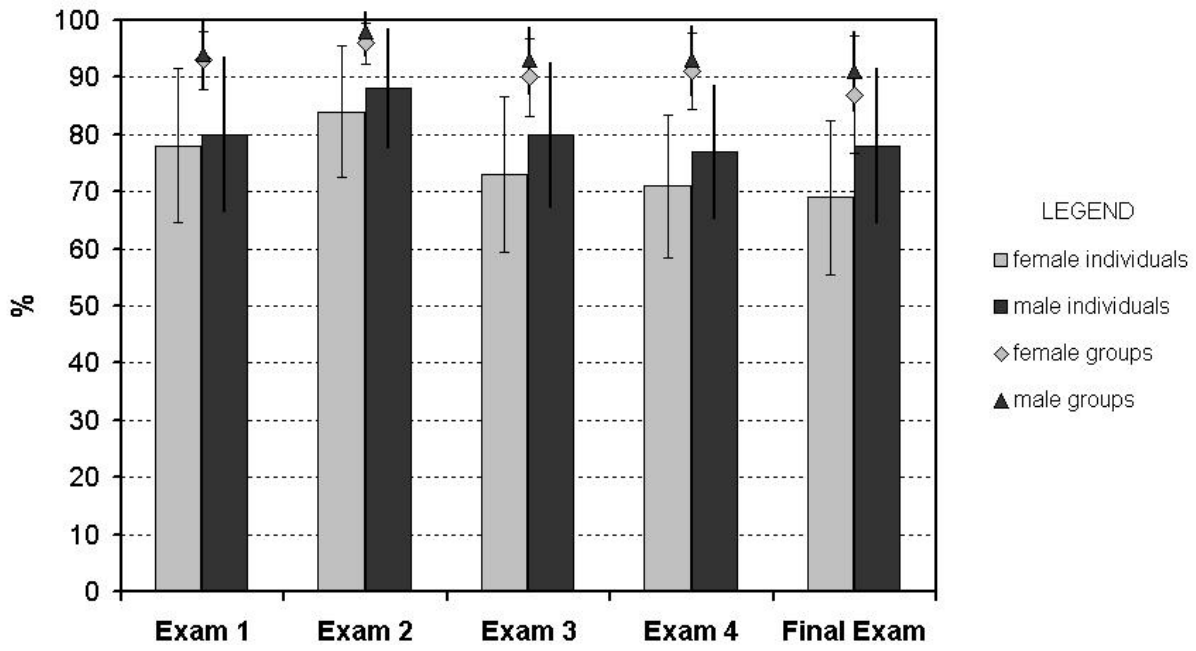
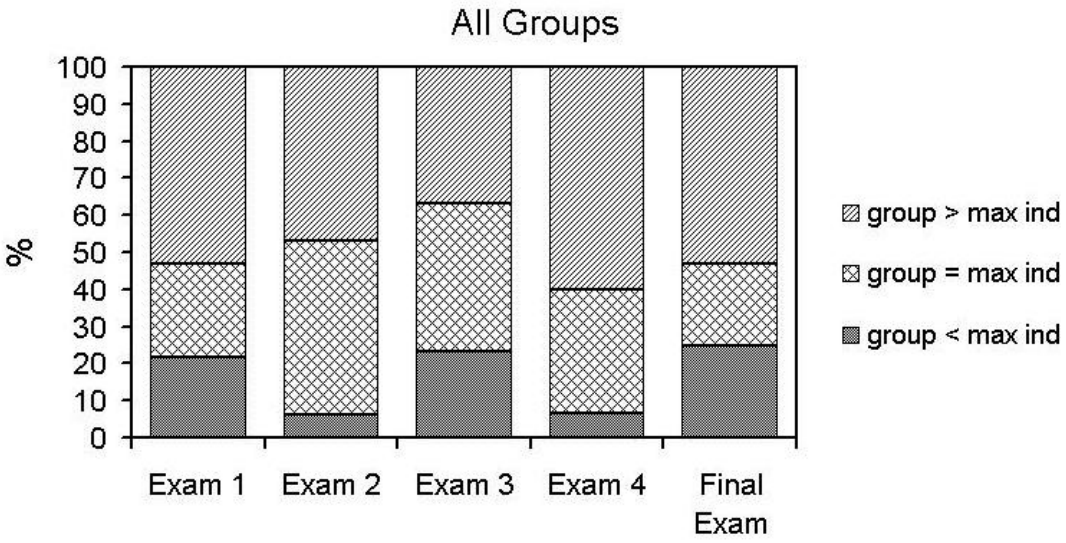
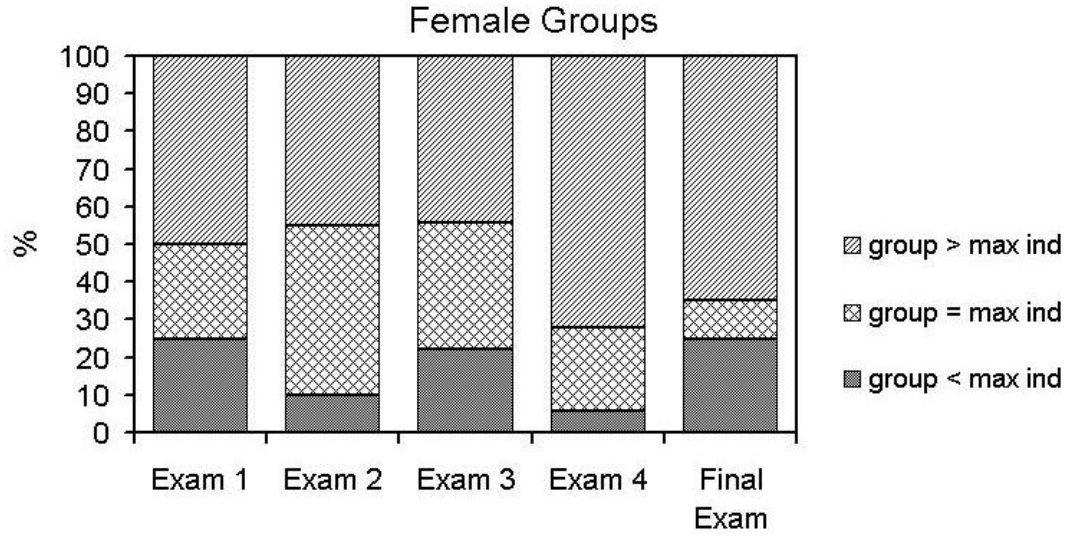


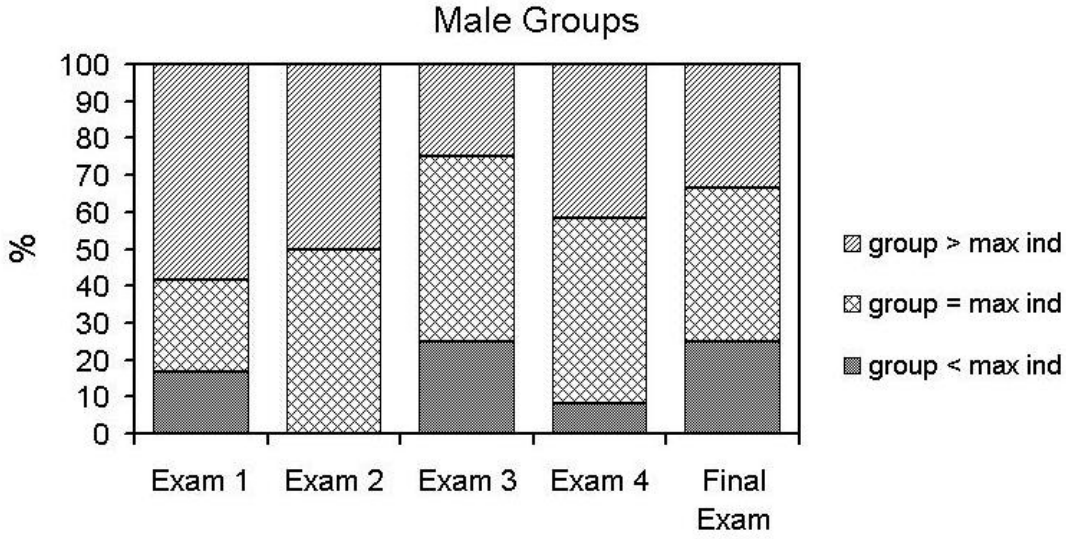
Figure 1. Mean scores for male and female individuals and groups on all five exams. Vertical lines show ± one standard deviation. For the final exam, only results from the half that covered new material and was tested using a group exam are reported here. Men's individual scores were significantly higher than women's on all but the first exam, but male and female group scores were not significantly different (see Table 1 for the statistics).

[Figure 2 on next page]

Figure 2. Percentage of groups (female, male, and both sexes together) earning a score greater than, equal to, or less than the maximum individual score within the group (labeled "max ind") on each of the five exams. Again, only the scores from the new material on the final exam are reported here. When groups earn better scores than the highest scoring individual within the group, cooperative learning appears to be successful; when the opposite is true about the scores, groups are not functioning as effectively. When the group and maximum individual scores are equal, groups may be relying primarily on their best-prepared member to provide the answers instead of pooling everyone's knowledge together.


All Groups
100
90
80
70
60
50
40
30
20
10
0
%
Exam 1
Exam 2
Exam 3
Exam 4
Final Exam
group > max ind
group = max ind
group < max ind


Female Groups
100
90
80
70
60
50
40
30
20
10
0
%
Exam 1
Exam 2
Exam 3
Exam 4
Final Exam
group > max ind
group = max ind
group < max ind


Male Groups
100
90
80
70
60
50
40
30
20
10
0
%
Exam 1
Exam 2
Exam 3
Exam 4
Final Exam
group > max ind
group = max ind
group < max ind