

Collaborative Testing, Gender, Learning Styles, and Test Performance

William Breedlove
Department of Sociology and Anthropology
College of Charleston

Tracy Burkett
College of Charleston

Idee Winfield
College of Charleston.

Abstract

Prior research provides little systematic evidence on student perceptions of collaborative testing. The scarce information that is available is usually in the form of a few instructor or student comments that characterize in general the collaborative experience. Likewise, one finds little evidence on how reactions to collaboration differ by student characteristics and test outcomes. Knowledge of these perceptions and variations can lead to a better understanding of how collaboration works and can better equip instructors for using collaborative testing. The present study begins to fill in this evidence gap. We report the results of a post-collaborative test survey and how those responses are associated with gender, learning style differences, and test performance.

How do students collaborate on paired or group tests? How do students perceive collaborative testing? Do those perceptions differ by gender and learning styles? What beliefs and behaviors are associated with better test performance under collaborative testing? Given the large body of accumulated research on collaborative learning, it is surprising to find few answers to these basic questions. The scant evidence available suggests that collaborative testing can be a powerful technique for enhancing academic achievement and for developing skills relevant to workplace success. Knowledge of these perceptions and behaviors, and the factors that are associated with their variation, are important for a better understanding and improved application of this teaching and learning method.

In 1990, Lunsford and Ede asked how gender, race, and class may affect classroom collaboration and concluded that at the time they could only offer vague answers to their questions. It seems to us that more than fifteen years later we still have only vague or impressionistic answers. In this paper we report evidence from a larger study of collaborative testing in which we systematically surveyed students about their collaborative experiences. We present a set of empirical results against which others might make more precise comparisons of collaborative learning outcomes. We also report how student reactions to and evaluations of collaborative testing differ across gender, learning styles, and test performance. Suggestions are offered for teachers interested in using collaborative testing and questions for future research are raised.

Literature Review

Collaborative learning is probably the most widely studied of all teaching techniques. Johnson, Johnson, and Stane (2000) found that over nine hundred studies of social interdependence have been conducted over the past one hundred years. Those that fit the criteria for inclusion in their meta-analysis of collaborative learning collectively report one hundred ninety-four different comparisons of the effects of various collaborative learning techniques on measures of academic achievement. A smaller number of studies report the effects of collaborative learning on non-achievement outcomes such as intergroup relations, acceptance of diversity, self-esteem, locus of control, class attendance and participation, quality of interactions with teachers and classmates, course satisfaction, and other outcomes (see Slavin, 1995; Millis & Cottell, 1998).

Within this large body of research, though, one finds very few studies that examine collaborative testing. Collaborative testing is an extension of collaborative learning into the evaluative setting. In the large majority of instances described in the prior research, collaboration among students is used as a means to learn subject matter on which students are then tested as individuals. The relatively few studies that report on collaborative testing find the technique to be beneficial in a number of ways. Achievement is enhanced, anxiety is lowered, and satisfaction is increased (Guest & Murphy, 2000; Grzelkowski, 1987; Hanshaw, 1982; Helmericks, 1993; Ley, Hodges, & Young, 1995; Muir & Tracy, 1999; Nowak, Miller, & Washburn, 1996; Russo & Warren, 1999). There are, however, at least two major shortcomings that characterize almost all the studies of collaborative testing we located. First, the evidence they present concerning student reactions to and evaluations of collaboration is often anecdotal and expressed in qualitative statements such as “most agreed” or “the majority were satisfied.” Ignoring possible problems of self-selection bias concerning which students are likely to offer comments and the self-censoring of those comments, these kinds of general statements are not adequate for making precise comparisons across studies and may not be very useful to teachers interested in evaluating their own collaborative testing experiences. For this, instructors need an empirical baseline against which they can compare their experiences. Second, there is almost no investigation into how the association between collaboration and student behaviors, beliefs, and attitudes may differ between groups. Among the multiple dimensions that may be important, collaborative testing experiences may differ by gender, learning styles, and test outcomes.

Belenky, Clinchy, Goldberger, and Tarule (1997) argue that women and men differ in their ways of knowing. Women have a preference for connected, socially-based knowledge. Their learning style is more cooperative and less individualistic and competitive. Men, on the other hand, are oriented toward individualistic and competitive learning environments. Consequently, women prefer collaborative learning more so than men. Lundeberg and Moch (1995) found support for this thesis. Ocker and Yaverbaum (2001), however, report that men in their study were more comfortable with collaborative team assignments.

Learning styles are the characteristic ways that students obtain, store, and retrieve information (Felder & Henriques, 1995). Though different measures of learning styles identify somewhat different

learning style dimensions, most measures include a dimension that contrasts active and reflective learners. The active-reflective contrast is similar to the Jung's and Myers-Briggs' extravert-introvert personality types (Felder & Silverman, 1988). Active learners prefer to process information through discussion or doing something with the information. Reflective learners process information through moments of quiet introspection. Due to their learning style preferences, active learners work well in groups while reflective learners may be hindered by such learning situations. Group work that does not allow for a period of reflection, such as collaborative testing, may be more of a hindrance than an aid to reflective learners.

Lastly, it is important to examine how student perceptions and behaviors correspond to their actual test performance. These associations suggest which kinds of behaviors lead to higher test scores and if perception of collaborative testing is dependent on test performance.

Data and Methods

The data is drawn from two independent experiments on collaboration that took place in consecutive spring semesters at a liberal arts college. Subjects were students in five sections of an introduction to sociology course. One hundred fifty-nine students, eighty-seven in the first experiment and seventy-two in the second, completed all the instruments required for inclusion in the analysis. Fifty-one percent of our students were freshmen and another thirty-five percent were sophomores. Sixty-one percent were female. Thirteen percent were non-white.

At the beginning of the semester, students were told that they had the opportunity to participate in a research study on learning and that participation was voluntary. In the first experiment, those who chose to participate would be allowed to take their second course exam with a randomly assigned same-sex partner and then complete a post-collaborative survey. They would take their first and last exams by themselves. Partners for the second exam would be assigned the day of the exam so students would not have the opportunity to engage in collaborative learning with their partner or build rapport with their partner beyond what might have been built during prior class meetings. Almost all students chose to participate. We administered a post-collaborative survey at the beginning of the class immediately following the collaborative test day and before students received their grades. The survey contained both

closed-ended and open-ended questions. The closed-ended questions are scored 5 = “strongly agree” to 1 = “strongly disagree”.

In the second experiment, a new set of students were similarly instructed about the testing and partnering process. Unlike the first experiment, they would take all three exams with their partner. Partners were determined early in the semester thereby allowing subjects to prepare collaboratively and build rapport before their exam. For the analysis in this paper, we examined the data for the second exam only for both semesters. The context of the second experiment is, of course, different and combining data from the two semesters may be cause for concern. Our post-collaborative survey data shows, however, that very few students actually worked together in preparation for any of the three exams. This difference, therefore, may not be significant. Further, while the students in the second experiment had the experience of a first collaborative exam and post-test survey, their second exam post-test survey responses are highly similar to those of the students in the first experiment. Lastly, we find no significant difference in the average second exam scores for the two sets of subjects. These pieces of information suggest that the context was not so different as to invalidate combining the data from the two experiments.

Learning styles are measured by Solomon and Felder's *Index of Learning Styles Questionnaire*. This data is available for the first experiment only. They indicate that just over three quarters of the students prefer an active learning mode while the remainder prefers a reflective mode.

Test performance is measured as percent correct of seventeen multiple-choice questions that were common to each instructor's exam. These questions covered subject matter that all instructors agreed their students had exposure to either through class discussion, lecture, or reading. Students averaged eighty-one and eighty-two percent correct on these seventeen items in the first and second experiments respectively.

Analysis and Discussion

Table 1 (see below) presents the percentage of subjects agreeing or strongly agreeing to nine closed-ended questions about their collaborative experiences. How did subjects work together? Numbers 1 and 2 show that almost all settle on a common strategy for answering questions and almost eighty-five

percent work at a similar pace. Although they were not instructed in how to test collaboratively, it appears that this novel testing situation presented no significant problems with the question-answering process for the majority of subjects. Numbers 3 and 4 indicate that about ninety percent both helped and received help in understanding from their partner. These high percentages suggest that the exam was truly a cooperative, shared experience. They also suggest that retention of learned information should be higher since explaining something to another person is associated with higher retention. The association between collaborative testing and retention may be a fruitful area for future research.

Responses in numbers 5 and 6 reflect psychological or emotional outcomes. Ninety percent believe that a partner made the exam less stressful. Because prior research shows that high levels of test anxiety can reduce test performance, the self-reported lowering of stress seen here should be associated with higher test scores. Subjects also say that having a partner made them feel more confident. This too could lead to higher test scores. On the other hand, having a partner may account for why no more than seventy-seven percent of students say they prepared more thoroughly for the exam. Perhaps the knowledge that they would have a partner led some to slack off or become free-riders. Numbers 8 and 9 reflect subject beliefs about the utility of collaboration. Ninety-four percent agree that collaboration is good for preventing "stupid" mistakes. Our data does not allow us to investigate what kinds of mistakes they mean by "stupid" or the extent to which they are catching such mistakes versus what must be, in their understanding, "non-stupid" mistakes. Future research might examine this mistake-catching function of collaborative testing. Additional research might also examine why and how some students are preparing more thoroughly than others. What is the motivating force behind this behavior? Lastly, ninety-three percent report that collaborative testing is better than individual testing. This is consistent with the impressions and anecdotal evidence of prior studies. But, since seven percent did not agree, we feel that it is important that teachers keep this practice voluntary and seek to understand why some students find individual testing preferable.

The open-ended questions asked subjects to identify what was good and what was bad about collaborating on the exam. The two most frequently mentioned good aspects concerned issues of interacting with others and confidence. About thirty percent wrote that sharing, talking, and the process of working with someone were beneficial. Approximately twenty percent mentioned feeling more confident,

yet only three percent wrote that it made them study harder.

Forty-five percent wrote that nothing was bad about collaborating. The most commonly cited drawbacks were negative feelings about negotiating disagreements and second-guessing themselves and time pressures. These drawbacks were mentioned by approximately twenty-two and twelve percent of respondents, respectively. Less than ten percent felt pressures associated with being partly responsible for another person's grade and only about five percent felt that their partner was unprepared. Lastly, less than five percent believed that noise was a problem.

Table 1
Post-collaborative Survey Responses, Closed-ended Questions

	Percentage Agreeing or Strongly Agreeing
1. Partners used a similar strategy to answer questions.	96
2. Partners worked at a similar pace.	84
3. I helped my partner understand.	92
4. My partner helped me understand.	89
5. Having a partner made taking the exam less stressful.	90
6. Having a partner made me more confident.	89
7. I prepared for the exam more thoroughly.	77
8. Collaborative testing prevents "stupid" mistakes.	94
9. Overall collaborative testing is better than individual testing.	93

We next examined the associations among gender, learning styles, test performance, and our survey results. The zero-order correlations are shown in Table 2. For gender, we find a significant correlation with agreement that having a partner made taking the test less stressful. It is our male students who are more likely to agree. If lower stress leads to higher test scores, and in our data these variables are positively and significantly correlated, it is male students who are, on average, more likely to

receive this benefit of collaboration. Contrary to earlier findings, we find no association between gender and preference for collaboration over individual work. Indeed, except for one outcome, we find no significant gender differences.

The learning style correlations show that active learners are more likely than reflective learners to agree that their partner helped them understand, that a partner made them more confident, that they prepared more thoroughly, that stupid mistakes were prevented, and to rate collaborative testing as better than individual testing. Although preparing more thoroughly is not significantly correlated with test performance in our data, it is significantly associated with confidence. It is perhaps not surprising that active learners would rate collaborative testing higher than reflective learners. Active learners have a preference for engaging the material in some active way and for group work. Collaborative testing provides an opportunity for actively discussing and debating the material. Reflective learners prefer to work alone, to have an internal conversation reflecting on the material, and may find active engagement in a testing situation to be distracting. Teachers should take learning styles into account and advise students that collaborative testing may not be to everyone's liking.

Finally, we note that higher test scores are significantly and positively correlated with helping a partner understand, having a partner that helped them understand, agreeing that a partner lowered stress and increased confidence, and agreeing that collaborative testing is better than individual testing. If it is true that being able to teach something requires more understanding than is needed to simply answer questions about that same thing, then it makes sense that students who said they helped their partner understand also scored higher on the exam. Those students had a better understanding of the material as shown in their ability to help others understand. Teachers using collaborative testing might have students engage in some practice teaching of each other in order to deepen their understanding and increase their ability to help each other in collaborative situations.

Table 2
Collaborative Testing, Gender, Learning Style, and Test Performance

	Gender	Learning Style	Test Performance
1. Partners used a similar strategy to answer questions.	-.097	-.019	.041
2. Partners worked at a similar pace.	-.129	-.148	.082
3. I helped my partner understand.	-.047	.177	.332**
4. My partner helped me understand.	-.135	.255*	.180*
5. Having a partner made taking the exam less stressful.	-.220**	.144	.231**
6. Having a partner made me more confident.	-.135	.459**	.214*
7. I prepared for the exam more thoroughly.	.041	.227*	-.008
8. Collaborative testing is good for preventing "stupid" mistakes.	-.069	.281**	.113
9. Overall collaborative testing is better than individual testing.	-.093	.273*	.245**

^a Data from the first experiment only.

Note: The nine survey items are score 5 = strong agree to 1 = strong disagree. Gender is coded Female = 1, Male = 0. Learning style is coded Active = 1, Reflective = 0.

* = $p < .05$ ** = $p < .01$ two-tailed tests

Conclusion

The results of this study provide an empirical baseline against which others can compare their collaborative testing outcomes. They also indicate some significant differences in beliefs and behaviors related to the collaborative testing experience across gender, learning style preference, and test performance. These differences should be taken into account when using a collaborative testing format. Overall, student perception indicates strong support for collaborative testing.

References

- Belenky, M., Clinchy, B., Goldberger, N., & Tarule, J. (1997). *Women's ways of knowing: the development of self, voice, and mind*. New York: Basic Books.
- Felder, R., & Henriques, E. (1995). Learning and teaching styles in foreign and second language education. *Foreign Language Annals*, 28, 21-31.
- Felder, R., & Silverman, L. (1988). Learning and teaching styles in engineering education. *Engineering Education*, 78, 674-681.
- Grzelkowski, K. (1987). A journey toward humanistic testing. *Teaching Sociology*, 15, 27-32.
- Guest, K. & Murphy, D. (2000). In support of memory retention: a cooperative oral final exam. *Education*, 121, 350-354.
- Hanshaw, L. (1982). Test anxiety, self-concept, and the test performance of students paired for testing and the same students working alone. *Science Education*, 66, 15-24.
- Helmericks, S. (1993). Collaborative testing in social statistics: toward *gemeinstat*. *Teaching Sociology*, 21, 287-297.
- Johnson, D., Johnson, R., & Stanne, M. (2000). Cooperative learning methods: a meta-analysis. Retrieved September 2003, from <http://www.clcrc.com/pages/cl-methods.html>
- Ley, K., Hodges, R., & Young, D. (1995). Partner testing. *Research and Teaching in Developmental Education*, 12, 23-30.
- Lundeberg, M., & Moch, S. (1995). Influence of social interaction on cognition: connected learning in science. *Journal of Higher Education*, 66, 312-335.
- Lunsford, A., & Ede, L. (1990). *Singular texts/plural authors*. Carbondale, IL: Southern Illinois University Press.
- Millis, B., & Cottell, P. (1998). *Cooperative learning for higher education faculty*. Phoenix, AZ: Oryx Press.
- Muir, S., & Tracy, D. (1999). Collaborative essay testing. *College Teaching*, 47, 33-36.
- Nowak, L., Miller, S., & Washburn, J. 1996. Team testing increases performance. *Journal of Education for Business*, 71, 253-256.
- Ocker, R., & Yaverbaum, G. (2001). Collaborative learning environments: exploring student attitudes and

satisfaction in face-to-face and asynchronous computer conferencing settings. *Journal of Interactive Learning Research*, 12, 427-448.

Russo, A., & Warren, S. (1999). Collaborative test taking. *College Teaching*, 47, 18-22.

Slavin, R. (1995). *Cooperative learning*. Boston: Allyn and Bacon.

Solomon, B., & Felder, R. (2001). *Index of learning styles questionnaire*. Retrieved August 2001, from <http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ilsweb.html>