An Empirical Examination of the Roles of Ability and Gender in Collaborative Homework Assignments*

Jeffrey Parker
George Hay Professor of Economics
Reed College
3203 SE Woodstock Blvd.
Portland, OR 97202
Voice: 503-517-7308
Fax: 503-777-7776
parker@reed.edu

Revision 10
November 2008

An Empirical Examination of the Roles of Ability and Gender in Collaborative Homework Assignments

Abstract

This paper investigates how ability and gender affect grades on homework projects performed by assigned pairs of students in an undergraduate macroeconomics course. The assignment grade is found to depend on the ability of both students, and the relative importance of the stronger and weaker student differs in predictable ways depending on the kind of assignment. Male-male pairs earn lower grades than male-female or female-female pairs, conditional on the measured ability of the students.

1. Introduction

Professors who allow, encourage, or assign students to work together on homework projects are usually left wondering how the abilities and efforts of the students involved contributed to the quality of the paper submitted. In particular, some may be concerned about the degree to which weaker students may free-ride on the efforts of stronger ones.[1] This paper uses nine years of student performance on collaborative homework projects in an undergraduate macroeconomics course to test the effects of stronger and weaker students on homework grades. It also finds evidence that the gender composition of a collaborative pair affects its performance.

Collaborative learning and, more generally, active-learning exercises have been a major point of emphasis of the leaders in the economic-education community.[2] (For example, see Becker (1997), Keenan and Maier (1995), and the papers in Becker and Watts (1998).) However, despite evidence that these activities are pedagogically effective, Becker and Watts (1996) find that cooperative learning was used in less than twenty percent of the economics courses that they surveyed in 1995 and this changed little in a follow-up survey done in 2000 (Becker and Watts (2001)).

1.1 Related Literature

The benefits of using collaborative activities in teaching both college students and those in primary and secondary schools have received considerable attention in the education literature. A recent survey of the higher-education literature is provided by Johnson, Johnson, and Smith (1998), who examine the evidence on the effectiveness of the related concept of "cooperative learning" in higher education and conduct a meta-analysis suggesting positive effects of cooperative activities on student learning. Slavin (1995) gives a comprehensive overview of various cooperative-learning methods in primary and secondary education.

Educators in science and mathematics have examined gender effects in small group settings. Heller and Hollabaugh (1992) find the weakest performance by two-male-one-female groups on three-person collaborations on physics problems. They also find that high-performing groups seem to require at least one high-ability student. Pollock, Finkelstein, and Kost (2007) look at the effects of small-group

collaborations on the learning of male and female physics students and find no systematic evidence that such activities reduce the gender gap.

A few studies have attempted to get inside the "black box" of student collaboration. Vidakovic and Martin (2004) provide transcripts and analysis of student interaction while devising proofs in geometry. Strober, Cook, and Fuller (1997) used videotapes to examine the interactions among introductory economics students working on a problem. The authors of the latter study "did not see any obvious gender differences in the groups." (p. 267)

Despite the widely recognized benefits of student collaboration, the difficulty of assigning individual grades to collaborative work may deter some professors from using collaborative activities. In particular, the potential for free riding, in which one or more student collaborators fails to contribute to the outcome yet receives a good grade based on the efforts of his or her partners is an important potential problem. In a discussion of collaborative teaching activities in sociology, Yamane (1996) lists free riding first among the complaints directed at collaborative learning projects. Some teachers, such as Bartlett (1995) in economics, McKinney and Graham-Buxton (1993) in sociology, and Kaufman and Felder (2000) in engineering have devised ingenious mechanisms to try to overcome the free-riding problem. Furtwengler (1995) suggests an evaluation rubric for cooperative learning in higher education based on primary and secondary school studies.

1.2 Overview of the Present Study

This study uses observations on the performance on homework assignments of collaborating pairs of students in a macroeconomics course to test several models of student interaction. The grade received by the pair of students on the assignment

3

is assumed to depend on the academic ability of the two students, the difficulty of the assignment, and the gender composition of the pair. Academic ability is measured by the grade in the course, which is treated as an endogenous variable and is instrumented by the student's performance in prior courses.

The particular student-ability variables used in this paper separate the effect of the stronger student in the pair from that of the weaker student. This allows us to see the comparative effects of the stronger and weaker student on the assignment grade and allows the testing of several alternative models of collaborating-student interaction.

Webb (1997) presents several models of student interaction in collaborative projects. One model is free-riding, which she called "social loafing." Another is division of labor, in which collaborating students simply divide up the tasks and work mostly independently. These models of homework collaboration have different implications for the determination of the grade. For example, in a pure free-riding situation one would expect that the ability of the weaker student would have no effect on the grade (assuming that it is the weaker student who free-rides). In the case of a pure division of labor in which each student does half of the assignment, each student's ability should affect the assignment grade equally.

In a true collaboration in which each student contributes to a collective understanding of the project, one might expect each student to contribute in relation to his or her ability. Even if both students exert equal effort, one would expect that the stronger student would have a greater effect on the grade. On challenging assignments, the limiting factor in the assignment grade is likely to be the ability of the pair to understand the main points of the exercise. Thus under true collaboration, the stronger student will have more individual understanding to share

4

with his or her partner, thus may play a greater role in determining the assignment grade.

The next section presents a simple illustrative example of three models of collaboration and their implications for the effects of the two students' abilities on the assignment grade. Sections 3 and 4 provide background on the macroeconomics course, the assignments, and the dataset. The econometric method used to test for these effects is described in Section 5. The final sections of the paper present the results of these tests and offer tentative conclusions.

The results suggest that the relative importance of the stronger and weaker student varies among different kinds of assignments. On problem sets, the stronger student seems to affect the grade much more strongly; both students' abilities contribute more equally on more project-oriented econometric assignments. I also find significant gender effects, with male-male pairs underachieving relative to other groups.

2. Three simple models of homework collaboration

This section examines three simple, benchmark examples of collaboration on homework assignments: pure division of labor, pure free riding, and joint work with equal contribution. Under pure division of labor, the two students work independently on separate parts of the assignment, joining them together at the end for submission. In the pure free riding case, one student (I assume the weaker student) does nothing and the other (stronger) student does the entire assignment. With joint work, the two students work together exerting equal amounts of effort.

Let P and Q be two students working together on a homework project. For simplicity, suppose that the assignment comprises ten separable pieces that are either done correctly or wrong. Each piece receives a grade of 1 if done correctly and 0 if not. We index the ability of each student by the probability that the student, working alone, can complete each piece satisfactorily, which is p for student P and q for student Q. Assume $p > q$.

2.1 Pure division of labor

If the two students divide up the pieces with each doing five, then on average P will get $5p$ points on her part and Q will get $5q$ points on his, giving a total expected grade of $E(g) = 5(p + q)$. If they follow the pure division of labor strategy, their collective grade depends equally on their abilities, p and q. We can summarize the effect of each student's ability on the expected assignment grade as

$$\frac{\partial E\left(g\right)}{\partial p} = \frac{\partial E\left(g\right)}{\partial q} = 5.$$

2.2 Pure free riding

If Q free-rides completely on the more able student P, then P ends up doing all ten pieces while Q does nothing and the grade will be $10p$. Q's ability will not affect the assignment grade at all; it will depend only on p:

$$\frac{\partial E\left(g\right)}{\partial p} = 10, \quad \frac{\partial E\left(g\right)}{\partial q} = 0.$$

2.3 Pure collaboration

The more complex situation is where they work together in pure collaboration. Information on how to solve a problem is a nonrival good—only one partner needs to know it in order to get the problem right. A simple case of this collaborative interaction would be for both students to do all ten pieces, compare results, and choose the correct one (assuming they can recognize it) in the case that one got it right and one wrong. In this simple case, the only problems that they get wrong are the ones that both partners fail to solve, which occurs with probability $(1 - p)(1 - q)$. Thus, the expected assignment grade is $E(g) = 10 [1 - (1 - p)(1 - q)] = 10 (p + q - pq)$. Once again, the effects of the two students' abilities in this case are symmetric, but because $p > q$, the effects are not equal:

$$\frac{\partial E(g)}{\partial p} = 10(1 - q), \quad \frac{\partial E(g)}{\partial q} = 10(1 - p) < \frac{\partial E(g)}{\partial p}.$$

Intuitively, because Q is a less able student, there will be more pieces of the assignment that Q cannot solve, on which the assignment grade will be determined by P's ability. Note that

$$\lim_{p \to 1} \frac{\partial E(g)}{\partial q} = 0;$$

as P's ability to solve pieces approaches perfection, there is nothing left for Q to contribute. This suggests that under this simple pure-collaboration model:

- The abilities of both students affect the assignment grade.
- The ability of the stronger student will matter more than that of the weaker student.

- The greater the difference in the abilities of the students, the greater will be the difference in the magnitude of their effects on the grade.

Of course, there is much more to a typical student collaboration than simply working the problems individually and comparing answers. Students who work well together may be able to solve problems that neither could solve alone. To capture this crudely in the simply collaborative model, we could add a term S relating to the quality of synergies in the students' interaction: $E(g) = 10\,(p + q - pq) + S$. This term could depend on factors such as the gender composition of the group or the nature of the assignment.

3. Course context and assignments

The assignments examined here are weekly homework assignments in Economics 314, a macroeconomic theory course required of economics majors at Reed College.[3] From 1998 to 2006, I taught one section of Econ 314 every spring semester using the then-current edition of David Romer's *Advanced Macroeconomics* text.[4] Students were required to do weekly homework projects, usually working in pairs with a new partner each week. There were between nine and eleven homework projects during the course that varied among three basic types. The most prevalent type is theoretical problems from Romer's text. The second type is empirical exercises in which students perform (with step-by-step guidance) basic econometric analysis with each pair using different data sets, for example, data for different countries. The third type is presentation to the class of a published empirical paper. On a few occasions an assignment has combined two of the types: a few problems

and a little econometric work or some econometric work that is presented to the class.

Class size ranged from a low of seven to a high of twenty-one students over the sample period. Only when the class size was small (or on a couple of occasions when I made an error in assigning) did students work with the same partner more than once in the semester. When there was an odd number in the class, either one student worked alone (the norm for textbook problems) or one group had three members (for econometric projects or presentations). Presentation projects were often done in teams of three when the class was relatively large because of limited time available for the presentations themselves.

Partner assignments were made by the instructor in a mostly non-systematic way. There were two exceptions to random assignment: On the first project (an econometric one), I attempted to pair students with no econometric or statistical background with ones who had some experience with statistical software. On the second, which was the first set of mathematical textbook problems, I tried to pair students with weaker math background with ones having completed more mathematical coursework. After these first two assignments, there was no pattern to assignments other than avoiding partners who had worked together already. In some years, the assignment submitted the week after spring break was done individually by all students, due to the difficulty of coordinating group work during the weekends before and after break. Regardless of the assignment of partners, student teams were allowed—indeed, encouraged—to seek help from other teams (and from the instructor) and to work in larger groups. To the extent that collaboration extended beyond the assigned pair, the relationship between

assignment grade and the abilities of the two students submitting the assignment is likely to be weakened.

The projects were graded on a 10-point scale with each partner receiving the same grade. In keeping with Reed College tradition, the grade was not written on the returned paper or otherwise communicated directly to the students, although the comments and corrections written on the paper allowed them to infer the instructor's assessment of the overall quality of their work. The level of difficulty of assignments and the grading scale remained comparable over the nine years of the sample. Assignments were graded by the instructor knowing the identity of the students, raising the possibility of experimenter bias. However, the grading criteria on the problem sets and econometric assignments are highly objective (e.g., is the problem solved correctly or are the econometric results accurate), suggesting that this problem may not be severe. Grading on the presentation assignments was inevitably more subjective because oral presentation skills (and the instructor's perception of them) may play an important role alongside demonstrated understanding of the paper being presented.

The weekly projects counted for about 20 percent of the course grade. The remainder of the grade was based on in-class and take-home exams (at least two of each) given during and at the end of the semester. To compensate in grading for the possibility of free-riding on group assignments, I asked each student at the end of the semester to identify the one or two students who were the most helpful to work with and those who were not helpful. Final grades of these students were adjusted where there was consensus among partners that the student was excellent or poor.

4. Data

The unit of observation is a homework project submitted by a team of students. Although students sometimes worked alone or in groups of three, the econometric analysis below examines only projects done by teams of two students. The dependent variable is the grade (on a 10-point scale) received on the assignment.

Of the 607 homework projects submitted by the 131 students taking macroeconomics over the nine-year sample, 387 involved textbook problems (300 done by two-student teams and 87 done by single students), 167 involved econometric tasks (150 done by two-student teams, 12 done individually, and 5 done by teams of three students), and 65 involved making presentations (30 done by two-student teams, 2 done individually, and 33 done by teams of three). Some projects involved more than one activity, so the numbers do not add up to 607. This study examines the 474 projects that were completed by teams of two students.[5]

As suggested in section 2, I expect that the principal determinant of the grade on a homework assignment would be the levels of ability and effort of the students working on the assignment. A good indicator of ability and effort for any individual student is the student's grade in the macro course itself, but this variable is endogenous. To deal with endogeneity, I calculate a predicted final course grade for each student based on a first-stage regression of course grade on a set of predetermined instruments. The instruments examined include the student's overall and economics-course grade-point average prior to the semester in which he or she took the macroeconomics course, whether or not the student had taken the corresponding course in microeconomics, the student's gender, and the total number

of courses, number of economics course, and number of math courses completed prior to taking macroeconomics.

5. Econometric method

I use i to index individual students, j to index assignments through the semester, and t to index years of the sample. Let $g_{i_1,i_2,j,t}$ be the grade received by students i_1 and i_2 on the jth assignment of year t.[6] This is the dependent variable of the model. $Z_{i_1,i_2,j,t}$ is a row vector of other variables affecting the assignment grade, such as gender dummies and dummies for the particular assignment j and year t. Let y_i be the macroeconomics course grade received by the ith student, an endogenous indicator of student ability. X_i is a row vector of predetermined characteristics of the ith student influencing his or her course grade.

The econometric model of the assignment grade is

$$g_{i_1,i_2,j,t} = \beta_0 + \beta_{mx} \max\left(y_{i_1}, y_{i_2}\right) + \beta_{mn} \min\left(y_{i_1}, y_{i_2}\right) + Z_{i_1,i_2,j,t}\gamma + \varepsilon_{i_1,i_2,j,t}, \tag{1}$$

where β_{mx} measures the effect on course grade of the stronger student in the pair and β_{mn} measures the effect of the weaker student. Because y is measured by the course grade, it is endogenous and must be instrumented.

In the estimation of (1), the actual course grades are replaced by the predicted course grades from

$$y_i = \alpha_0 + X_i\delta + u_i, \quad i = 1, 2, \ldots, n, \tag{2}$$

where n is the overall number of students. Equation (2) is estimated by ordinary least squares using the sample of all students who completed the macroeconomics

course during the nine-year sample.[7] Based on the first-stage regression, I calculate a predicted value

$$\hat{y}_i = \hat{\alpha}_0 + X_i\hat{\delta}$$

for each student in the sample. I then compare the predicted course grades of the two students involved in each assignment observation $\hat{y}_{i_1}$ and $\hat{y}_{i_2}$, construct the appropriate variables for the maximum and minimum predicted course grade, and estimate the second-stage regression (1) by ordinary least squares.

Although it is valid to calculate second-stage regression coefficients by OLS, the usual OLS standard errors are not valid for the same reasons as in the second stage of a two-stage least-squares procedure. However, the standard two-stage least squares estimator of the standard error cannot be used here either because it would not take account of the duplication of individual students' predicted course grades, which occurs because each student does more than one homework assignment in the sample, or of the process of choosing the maximum and minimum on each assignment.

Instead, standard errors are calculated using a bootstrap procedure that simulates the first-stage regression of course grades for individuals, the calculation of the predicted course grades from this equation, the determination of which of the two predicted course grade is larger, and the estimation of the second-stage regression using the max and min predicted grades as regressors. See the appendix for a detailed discussion of the bootstrap methodology and a comparison of computed standard errors under alternative assumptions.

The instruments used in equation (2) include the student's cumulative grade point average at the beginning of the semester in which he or she took macro, the

student's grade point average in prior economics courses, and a dummy variable indicating whether the student had completed the corresponding microeconomic theory course. A gender dummy and measures of the number of economics and mathematics courses completed had no marginal effect on the macro grade and were omitted.[8]

6. Results

Summary statistics for the variables in the first-stage course-grade regression, in which the unit of observation is the individual student, are shown in Table 1. Those of the project-grade regression, in which the unit is the assignment paper submitted by a pair of students, are in Table 2.

The first-stage regression of course grade on pre-course academic characteristics is shown in Table 3. Because of the extremely small coefficients and low statistical significance of the coefficients on prior math, total, and econ courses and the gender dummy, the regression in the third column is used to instrument the course grade. The dummy for having taken microeconomics is retained because its coefficient is plausible and large enough to influence the prediction and its p value is 0.075 in a one-tailed test. The equation F statistic for this regression is 25.2, suggesting that the instruments are strong.

The results in Table 3 suggest that the most important predictor of success in macroeconomics is the student's overall college GPA when he or she enters the class. Grades in prior economics courses also help predict the macro grade, but have a smaller and less statistically significant effect. As noted above, having taken the microeconomic theory class has a nearly significant effect with a point estimate of

one-half of a plus/minus grade (e.g., improving the predicted grade halfway from B to B+).

The second-stage regression has assignment grade as the dependent variable. The regressors are maximum and minimum (predicted) course grade of the two students, dummies for the gender composition of the two-student group, and a full set of assignment/year interaction dummies to control for differences in the degree of difficulty of assignments both within and between years.

Table 4 presents the estimated coefficients on the two predicted course grades and dummies for the gender composition of the two-student group. The omitted group is male-female. Column (1) shows the pooled regression of all homework projects; columns (2) through (4) show the results for the three specific project types.

In the pooled sample, the grade on the collaborative project responds most strongly to the quality (as represented by predicted course grade) of the stronger student. The weaker student's quality also plays an important role, though its effect is much smaller. However, the relative importance of the two students' quality measures seems to vary substantially across types of assignments. On the assignments involving problems from the Romer text (the largest category), the results are even more striking than those of the pooled sample: the stronger student dominates the determination of the assignment grade. On the assignments involving econometric tasks, the two students' quality levels contribute much more equally (the hypothesis that the two effects are equal cannot be rejected). The small number of two-person presentation-type assignments in the sample makes it difficult to assess the results in column (4), but on these assignments it appears that the better student once again dominates the determination of the grade.

15

One notable and consistent result in Table 4 is that male-male groups tend to underperform relative to the grade that would be expected based on their predicted course grades. Though the negative coefficient occurs for all types of assignments, the effect is statistically significant only for the problem sets and the pooled sample. Note that this is not a reflection of any tendency for male students to do less well in the course as a whole. (Recall that the gender variable had a coefficient that was essentially zero in the course-grade regression of Table 3.) Instead, it appears that males perform worse on problem sets on which they are assigned to work with other male students. Female students seem to do about equally well when working with other females as when working with males: the coefficient of the female-female dummy variable in Table 4 is always small and statistically insignificant.

7. Interpretation

The econometric results strongly reject the hypothesis of pure free riding for the pooled sample and for both problem-based assignments and econometric assignments. The ability of the weaker student clearly matters in these cases. [9]

For the problem sets, the evidence supports the true collaboration model, in which the ability of the stronger student affects the assignment grade more strongly than that of the weaker student. This is consistent with the nature of these assignments. Textbook problems in this course are mathematical in nature and the students generally find them both difficult and time-consuming. A pair is likely to get a problem right as long as at least one of them can (and does) figure out how to solve it, which matches the assumptions of the simple illustrative collaborative model in section 2.

The simple collaborative model also implies that the stronger student's effect should be more dominant as the difference in abilities is larger. Table 5 tests this implication of the model. Let $d_{i_1,i_2} = \max\left(y_{i_1}, y_{i_2}\right) - \min\left(y_{i_1}, y_{i_2}\right)$ be the difference between the predicted grades of the stronger and weaker student on an assignment. Column (1) shows the coefficients of the regression for those 111 pairs in the full sample for which one student's predicted grade exceeded that of the other by a full grade point. Clearly, the stronger student's ability dominates the assignment grade in this case, as predicted by the model. The bottom row of the table shows that we can decisively reject the null hypothesis that the effects are equal for these pairs.

Column (2) of Table 5 shows the comparable regression for the pairs of students whose predicted grades differ by less than one grade point. For these cases, the effects of the two students are nearly identical and the hypothesis of equality is not rejected.[10]

The collaboration model in section 2 suggested that the course grade should follow the model $E(g) = 10\ (p + q - pq) + S$, where p corresponds to the maximum predicted grade and q to the minimum. Column (3) of Table 5 estimates this equation directly by including an interaction term consisting of the product of the max and min predicted grades. Consistent with the predictions of the model, the max and min predicted grades have equal effects and the cross-product term has a negative effect. However, the hypothesis that the coefficient on the cross-product term is the negative of the coefficients on the max and min predicted grades is decisively rejected. Thus, the evidence of Table 5 provides considerable support for, if not complete consistency with, the pure collaboration model.

While the problem sets and the pooled sample show evidence in support of the pure collaboration model, the evidence from the econometric assignments is consistent with the division of labor model. The econometric-oriented assignments in the course are typically less challenging, though not necessarily less work. Anecdotal evidence suggests that students have often divided the econometric tasks, perhaps each taking one of two assigned countries or each focusing on one or two parts of the project. Since these assignments involve less "figuring out" and more hands-on econometric work, the grade on these problems is likely to reflect effort more than enlightenment, so there is less opportunity for the information understood by the stronger student to spill over in a non-rivalrous way to the weaker one. It is not surprising that the ability of the lesser student plays an equally important role here and that the division of labor model fits the data well.

As noted above, the presentation assignments are both fewer (for groups of two) and more subjectively graded. Public-speaking skills, which might not be highly correlated with the course grade, are likely to play an important role in the grade alongside analytic ability. Nonetheless, the results are overall quite similar to those for the problem sets: the measured ability of the better student seems to be the most important factor determining the assignment grade. The negative estimated coefficient on the predicted grade of the weaker student is not statistically significant and, in such a small sample, is likely spurious.

The partner-interaction term S is proxied here by the gender dummies. As noted above, Heller and Hollabaugh (1992) find that two-male-one-female groups tend to underperform on physics assignments. Thomas-Hunt and Phillips (2004) argue that men often do not respond to women's expertise in collaborative projects and Lee (1993) finds that female computer-science students participate less when

males are present in their collaborative groups. These results suggest that mixed groups may not benefit fully from the expertise of their female members, which implies that MF groups might be expected to perform worse than same-sex groups.

However, I do not find results consistent with this behavior among Reed College economics students. In the regressions reported above, I find that male-male groups consistently tend to be the worst-performing groups. This negative effect of all-male groups is statistically significant on problem sets and overall. Moreover, in a regression for mixed groups only, the ability level of the female group member has a larger and more statistically significant effect on the project grade (0.48) than the effect of the male (0.36), which directly contradicts the hypothesis of male dominance in group work based on traditional gender roles.

I can only speculate on the underlying causes of these results. They lend statistical support to casual evidence I have collected over the years that male students often work better and harder when working with a female partner. Another possible explanation for the low performance of all-male groups is the "asking for directions" effect. If, as conventional wisdom implies, men are more reluctant to ask for assistance, then groups having at least one female may have sought and received more help on the projects both from their peers and from the instructor. While I have no data on which groups asked for help, personal recollection supports this interpretation.

8. Conclusion

This study uses data from a macroeconomics course at Reed College to test alternative models of student interaction in determining the grades on collaborative

homework assignments. The results for the pooled sample of for the problem sets in particular are consistent with a model in which both students contribute to the assignment grade in proportion to their abilities. The ability levels of both students have an effect, but the effect of the stronger student's ability is larger. On econometric, data-oriented assignments, the results are more consistent with a model in which the students divide the labor of doing the assignment and work separately.

The hypothesis of pure free riding on collaborative homework assignments among macroeconomics students at Reed is strongly rejected. This may reassure instructors who might be deterred from using collaborative assignments by free-riding concerns. However, one must be cautious about extrapolating these results to institutions and courses with different student cultures. A strong ethic of "learning for its own sake" pervades Reed's student culture and undoubtedly makes free riding less common than it would otherwise be. Moreover, grades are not disclosed to students either on their assignments during the class or on a report card at the end (although they can obtain them by asking), which contributes to a culture in which grades are emphasized less than knowledge. Given these characteristics, one might expect less free-riding at an institution like Reed than in environments where grades are more visible and where learning for its own sake is taken less seriously.

As the instructor of the course from which the data were drawn, I am not surprised by the results. I interact with the students extensively while they are working on problem sets and other homework. There is a great deal of collaborative learning on these assignments; students who catch on to the problems more quickly (sometimes with my assistance) routinely help their partners and sometimes other classmates to understand the problems better. While such collaboration muddles the

signal conveyed by the grade, as a teacher I believe that there are great benefits to this kind of collaborative learning, despite the possible loss in the clarity of the homework grade as an evaluation tool. While the present study does not provide direct evidence of the benefits of collaboration in terms of student understanding, it does suggest that the nature of collaboration in the class is often of the "pure collaboration" kind that is likely to have positive educational benefits.

This study also provides some guidance to instructors in understanding how the grades on collaborative projects reflect the abilities of the contributing students. As noted above, free riding is one of the most common reservations expressed about collaborative work. This paper strongly rejects pure free riding among students in a core theory course at a selective liberal-arts college. Instructors in environments to which the Reed College example may extend might thus feel reassured about the nature of students' collaboration.

Tables

Table 1. Summary statistics by student

	Mean	Standard Deviation	Minimum	Maximum
Final grade	3.13	0.87	0.00	4.00
Predicted grade	3.13	0.59	1.49	4.04
Male dummy	0.58	0.50	0	1
Micro dummy	0.49	0.50	0	1
Prior GPA	3.13	0.45	2.00	3.89
Prior econ GPA	3.21	0.73	0.00	4.00
Prior total courses	14.5	5.2	3.5	30.0
Prior econ courses*	2.7	1.6	0.0	10.0
Prior math courses	1.9	2.1	0.0	15.0

$N = 131$ except Final grade ($N = 126$), which excludes withdrawals.
*Counts only courses successfully completed. One student had taken one economics course but failed it, which is the single observation with zero values of both prior econ courses and prior econ GPA.

Table 2. Summary sample statistics by assignment

	Mean	Standard Deviation	Minimum	Maximum
All Assignments (*N* = 474)				
Grade*	8.25	1.18	1.0	10.36
Max predicted course grade	3.47	0.38	1.71	4.04
Min predicted course grade	2.79	0.58	1.49	3.95
Male-male	0.31	0.46	0	1
Female-female	0.18	0.38	0	1
Problem Set Assignments (*N* = 300)				
Grade*	8.13	1.28	1.0	10.36
Max predicted course grade	3.45	0.38	2.20	4.04
Min predicted course grade	2.78	0.58	1.49	3.94
Male-male	0.30	0.46	0	1
Female-female	0.17	0.37	0	1
Econometric Projects (*N* = 150)				
Grade	8.50	1.03	5.0	10.0
Max predicted course grade	3.48	0.40	1.71	4.04
Min predicted course grade	2.78	0.58	1.49	3.81
Male-male	0.32	0.47	0	1
Female-female	0.21	0.41	0	1
Presentation Projects (*N* = 30)				
Grade	8.32	0.64	7.0	9.0
Max predicted course grade	3.57	0.32	2.82	4.04
Min predicted course grade	2.91	0.62	1.54	3.95
Male-male	0.23	0.43	0	1
Female-female	0.17	0.38	0	1

*Extra credit problems allowed an assignment grade to exceed 10 for three observations.

Table 3. Course-grade regressions

	Dependent variable: Macroeconomics course grade		
	(1)	**(2)**	**(3)**
Prior college GPA	0.829***	0.832***	0.829***
	[0.238]	[0.237]	[0.237]
Prior econ GPA	0.359**	0.357**	0.345**
	[0.166]	[0.166]	[0.160]
Dummy for micro	0.212*	0.213*	0.161
	[0.127]	[0.127]	[0.111]
Prior math courses	0.006	0.006	
	[0.032]	[0.032]	
Prior total courses	0.004	0.004	
	[0.015]	[0.015]	
Prior econ courses	-0.033	-0.033	
	[0.044]	[0.043]	
Dummy for male	-0.012		
	[0.112]		
Constant	-0.682	-0.693	-0.649
	[0.565]	[0.553]	[0.516]
Observations	126	126	126
R-squared	0.48	0.48	0.47
Robust standard errors in brackets * significant at 10%; ** significant at 5%; *** significant at 1% (two-tailed)			

Table 4. Determinants of homework grade

	Dependent variable: Assignment grade			
	(1)	**(2)**	**(3)**	**(4)**
	All	Problems	Econometric	Presentation
Max predicted course grade	0.890***	1.108***	0.455**	1.501**
	[0.135]	[0.202]	[0.197]	[0.699]
Min predicted course grade	0.230***	0.229**	0.289**	−0.320
	[0.089]	[0.118]	[0.136]	[0.401]
Male-male groups	−0.210**	−0.286**	−0.078	−0.224
	[0.102]	[0.132]	[0.157]	[0.379]
Female-female groups	−0.015	−0.023	−0.167	0.215
	[0.118]	[0.172]	[0.185]	[0.405]
Observations	474	300	150	30
R-squared	0.44	0.45	0.40	0.55
Non-parametric bootstrap standard errors in brackets All equations have constant term and dummies for each assignment/year * significant at 10%; ** significant at 5%; *** significant at 1% (two-tailed)				

Table 5. Effects of the difference in abilities

	(1)	(2)	(3)
Dependent variable: Assignment grade			
	$d > 1.0$	$d < 1.0$	Full sample
Max predicted course grade	2.282*** [0.135]	0.397*** [0.138]	1.957*** [0.559]
Min predicted course grade	0.057 [0.085]	0.414*** [0.085]	1.768** [0.742]
Cross-product of max & min			−0.432* [0.215]
Observations	111	363	474
R-squared	0.71	0.49	0.45
Test of H_0: $\beta_{mx} = \beta_{mn}$	$F = 6.17$ $p < 0.02$	$F = 0.002$, $p > 0.95$	$F = 0.28$ $p \cong 0.61$
Test of H_0: $\beta_{mx} = \beta_{mn} = -\beta_{cross}$			$F = 7.63$ $p < 0.01$
Non-parametric bootstrap standard errors in brackets. Reported p values for F tests are based on quantiles of bootstrap stimulations. Each equation includes a constant term, gender dummies, and dummies for each assignment/year. * significant at 10%; ** significant at 5%; *** significant at 1% (two-tailed)			

Appendix: Bootstrap standard errors

Each replication k of the bootstrap algorithm can be summarized by the following sequence of steps:

1) Generate a simulated error term $\tilde{u}_i^k$ for each student i. This can be done in either of two ways:

   a) parametrically using a normal distribution with mean zero and variance equal to the estimated error variance from the OLS regression of (2), or

   b) non-parametrically by re-sampling from the residuals of the OLS regression of (2).

2) Create a simulated course grade $\tilde{y}_i^k = \hat{\alpha}_0 + X_i\hat{\delta} + \tilde{u}_i^k$, where $\hat{\alpha}_0$ and $\hat{\delta}$ are OLS estimates of the coefficients of (2). Actual course grades can take on only the feasible values 4.0, 3.7, 3.3, 3.0, 2.7, 2.3, 2.0, 1.7, 1.0, and 0.0, corresponding to the course grades of A, A−, B+, B, B−, C+, C, C−, D, and F. Thus, one could choose either to:

   a) round the simulated course grade to the nearest feasible value, or

   b) leave the simulated course grade as a continuous variable.

3) Regress the simulated course grade $\tilde{y}_i^k$ on X_i to get first-stage estimates $\tilde{\alpha}_0^k$ and $\tilde{\delta}^k$.

4) Compute the simulated instrumented course grade for each student as $\hat{\tilde{y}}_i^k = \tilde{\alpha}_0^k + X_i\tilde{\delta}^k$.

5) For each student pair i_1 and i_2, compare $\hat{\tilde{y}}_{i_1}^k$ and $\hat{\tilde{y}}_{i_2}^k$, assigning the larger to the max variable and the smaller to the min variable.

6) Generate a simulated error term $\tilde{\varepsilon}^{k}_{i_1,i_2,j,y}$ for each assignment/year/student

pair. This could be done either

a) parametrically using a normal distribution with mean zero and variance

equal to the estimated error variance from the second-stage regression of

equation (1) using predicted values for the y variables, or

b) non-parametrically by re-sampling from the residuals of this second-stage

regression of (1).

7) Create a simulated assignment grade for each assignment pair

$\tilde{g}^{k}_{i_1,i_2,j,t} = \hat{\beta}_0 + Z_{i_1,i_2,j,t}\hat{\gamma}$, where $\hat{\beta}_0$ and $\hat{\gamma}$ are the estimated values from the

second-stage regression of (1). There are two null hypotheses of interest, that

values of β_{mx} and β_{mn} are equal and that the respective values are zero. The

dependent variable is calculated under both null hypotheses: setting the

coefficients to zero and setting them to the constrained common point

estimate $\hat{\beta}_r$ of β_{mx} and β_{mn}. Results are not qualitatively different. Standard

errors and tests in Table 4 and the tests of $\beta = 0$ in Table 5 are based on $\beta_{mx} =$

$\beta_{mn} = 0$; the tests of coefficient equality in Table 5 are based on $\beta_{mx} = \beta_{mn} = \hat{\beta}_r$,

with the appropriate extension for the test involving the cross-product.

8) Regress $\tilde{g}^{k}_{i_1,i_2,j,t}$ on $\max\left(\tilde{y}^{k}_{i_1}, \tilde{y}^{k}_{i_2}\right)$, $\min\left(\tilde{y}^{k}_{i_1}, \tilde{y}^{k}_{i_2}\right)$, and $Z_{i_1,i_2,j,y}$, saving the

estimated coefficients $\tilde{\beta}^{k}_{mx}$, $\tilde{\beta}^{k}_{mn}$, $\tilde{\beta}^{k}_0$, and $\tilde{\gamma}^{k}$.

The above steps are repeated 1000 times (indexed by k). The bootstrap standard

errors are the standard deviations of the estimated coefficients over the replications.

Where p values are required, they can be computed using the quantiles of the

simulated distribution of the estimated coefficients.

As noted above, there are two sets of alternative assumptions that may be adopted in calculating the bootstrap standard errors. One can either round the simulated course grades to feasible values or not, and one can use parametric or non-parametric error terms. In practice, the results of all of the bootstrap calculations of standard errors were close to one another and also close to the second-stage robust OLS standard errors. Table A-1 compares the standard error estimates for the coefficients of the full-sample equation using several alternatives. Table 4 reports non-parametric bootstrap standard errors imposing discontinuity of course grades by rounding; the statistical significance of the coefficients would not be strongly affected by choosing another option.

Table A-1. Alternative standard error estimates for full sample

		SE of β_{mx}		SE of β_{mn}	
OLS robust standard error		0.171		0.098	
Bootstrap standard error with:		$H_0: \beta = 0$	$H_0: \beta = \hat{\beta}_r$	$H_0: \beta = 0$	$H_0: \beta = \hat{\beta}_r$
Parametric	Rounded	0.151	0.180	0.095	0.106
Parametric	Not rounded	0.134	0.167	0.091	0.099
Non-parametric	Rounded	0.135	0.166	0.089	0.094
Non-parametric	Not rounded	0.126	0.151	0.083	0.088

References

Bartlett, Robin L. 1995. A Flip of the Coin--A Roll of the Die: An Answer to the Free-Rider Problem. *Journal of Economic Education* 26 (2):131-139.

Becker, William E. 1997. Teaching Economics to Undergraduates. *Journal of Economic Education* 35 (3):1347-73.

Becker, William E., and Michael Watts. 1996. Chalk and Talk: A National Survey on Teaching Undergraduate Economics. *American Economic Review* 86 (2):448-453.

———. 2001. Teaching Economics at the Start of the 21st Century: Still Chalk-and-Talk. *American Economic Review* 91 (2):446-451.

———, eds. 1998. *Teaching Economics to Undergraduates: Alternatives to Chalk and Talk*. Cheltenham, U.K.: Elgar.

Furtwengler, Carol B. 1995. Practical Methods for Assessing Cooperative Learning in Higher Education. In *Interactive Learning in the Higher Education Classroom: Cooperative, Collaborative, and Active Learning Strategies*, edited by H. C. Foyle. Washington, DC: National Education Association.

Heller, Patricia, and Mark Hollabaugh. 1992. Teaching Problem Solving through Cooperative Grouping. Part 2: Designing Problems and Structuring Groups. *American Journal of Physics* 60 (7):637-644.

Johnson, David W., Roger T. Johnson, and Karl A. Smith. 1998. Cooperative Learning Returns to College: What Evidence Is There That It Works? *Change* 30 (4):26-35.

Kaufman, Deborah B., and Richard M. Felder. 2000. Accounting for Individual Effort in Cooperative Learning Teams. *Journal of Engineering Education* 89 (2):133-140.

Keenan, Diane, and Mark H. Maier. 1995. *Economics Live: Learning Economics the Collaborative Way*. New York: McGraw-Hill.

Lee, Mihwa. 1993. Gender, Group Composition, and Peer Interaction in Computer-Based Cooperative Learning. *Journal of Educational Computing Research* 9 (4):549-577.

McKinney, Kathleen, and Mary Graham-Buxton. 1993. The Use of Collaborative Learning Groups in the Large Class: Is It Possible? *Teaching Sociology* 21 (4):403-408.

Pollock, Steven J., Noah D. Finkelstein, and Lauren E. Kost. 2007. Reducing the Gender Gap in the Physics Classroom: How Sufficient Is Interactive Engagement? *Physical Review Special Topics Physics Education Research* 3 (1):010107.

Romer, David. 2006. *Advanced Macroeconomics*. New York: McGraw-Hill/Irwin.

Slavin, Robert E. 1995. *Cooperative Learning: Theory, Research, and Practice*. 2nd ed. Boston: Allyn and Bacon.

Strober, Myra H., Allen Cook, and Kasi Allen Fuller. 1997. Making and Correcting Errors in Student Economic Analysis: An Examination of Videotapes. *Journal of Economic Education* 28 (3):255-271.

Thomas-Hunt, Melissa C., and Katherine W. Phillips. 2004. When What You Know Is Not Enough: Expertise and Gender Dynamics in Task Groups. *Personality and Social Psychology Bulletin* 30 (12):1585-98.

Vidakovic, Draga, and William O. Martin. 2004. Small-Group Searches for Mathematical Proofs and Individual Reconstructions of Mathematical Concepts. *Journal of Mathematical Behavior* 23 (4):465-492.

Webb, Noreen M. 1997. Assessing Students in Small Collaborative Groups. *Theory into Practice* 36 (4):205-213.

Yamane, David. 1996. Collaboration and Its Discontents: Steps toward Overcoming Barriers to Successful Group Projects. *Teaching Sociology* 24 (4):378-383.

Endnotes

[1] By free-riding I mean free-riding in effort: the weaker student exerts little or no effort toward a joint project. An alternative definition would be free-riding in evaluation, whereby a weak student receives a higher grade by virtue of having worked with a stronger student. This study finds free-riding in evaluation, but little evidence of free-riding in effort.

[2] "Cooperative learning," "collaborative learning," and "group learning" have been defined in subtly distinct but varying ways in the education literature of various disciplines. As will be clear below, this study examines homework activities performed jointly by groups of two or more students, possibly interacting with other groups. This model of learning embodies elements of cooperative, collaborative, and group activity.

[3] A link to the current syllabus and assignments for Econ 314 can be found at http://academic.reed.edu/economics/parker.

[4] The current edition at the time of writing is Romer (2006).

[5] In two-student groups, maximum and minimum student abilities completely specify the ability levels of the students working on the project. If three-student groups were included, either the middle-quality student would be left out or the equation would have to be respecified in a way that would not work for two-student groups.

[6] One should not think of this as a complete four-dimensional array; that would only be true if every student worked (separately) with every other student on every project in every year. For any j and t, only those pairs i_1 and i_2 who worked together on that particular assignment would have entries in the sample.

[7] Of the 131 students completing at least one homework project, only 126 completed the course and received a grade. The five students who withdrew from the course are excluded from the estimation of equation (2).

[8] A variable for prior grade point average in mathematics might also be a useful predictor, but some students had fulfilled the calculus prerequisite through transfer work or in high school and had taken no math courses at the college prior to taking macro. These students had no prior math GPA.

[9] Although the coefficient of the weaker student is statistically zero for the presentation assignments, the size of the sample is too small to take this result as reliable.

[10] Attempting to subdivide the sample further encounters problems of collinearity between the maximum and minimum grade, which reduces the reliability of the coefficient estimates. For the two sub-samples reported in Table 5, the correlation between the max and min predicted grades is 0.51 for observations with $d > 1$ and 0.78 for those with $d < 1.0$. Correlation coefficients for sub-samples with $0.5 < d < 1.0$ (140 observations) and $d < 0.5$ (223 observations) are 0.93 and 0.94, respectively, with estimates of 0.559 and 0.554 for β_{mx} and β_{mn} in the former sub-sample and -0.234 and 0.878 in the latter. In neither case are the coefficient estimates significantly different from zero or from each other.