J. Engr. Education, 89(2), 133–140 (2000)

ACCOUNTING FOR INDIVIDUAL EFFORT IN COOPERATIVE LEARNING TEAMS

Deborah B. Kaufman and Richard M. Felder
Department of Chemical Engineering

Hugh Fuller
College of Engineering

North Carolina State University

ABSTRACT

An "autorating" (peer rating) system designed to account for individual performance in team projects was used in two sophomore-level chemical engineering courses in which the students did their homework in cooperative learning teams. Team members confidentially rated how well they and each of their teammates fulfilled their responsibilities, the ratings were converted to individual weighting factors, and individual project grades were computed as the product of the team project grade and the weighting factor. Correlations were computed between ratings and grades, self-ratings and ratings from teammates, and ratings received and given by men and women and by ethnic minorities and non-minorities. Incidences of "hitchhikers" (students whose performance was considered less than satisfactory by their teammates), "tutors" (students who received top ratings from all of their teammates), dysfunctional teams, and teams agreeing on a common rating were also determined. The results suggest that the autorating system works exceptionally well as a rule, and the benefits it provides more than compensate for the relatively infrequent problems that may occur in its use.

INTRODUCTION

Cooperative learning (CL) is an instructional paradigm in which teams of students work on structured tasks (e.g., homework assignments, laboratory experiments, or design projects) under conditions that meet five criteria: positive interdependence, individual accountability, face-to-face interaction, appropriate use of collaborative skills, and regular self-assessment of team functioning. Many studies have shown that when correctly implemented, cooperative learning improves information acquisition and retention, higher-level thinking skills, interpersonal and communication skills, and self-confidence (Johnson, Johnson, and Smith, 1998).

Most cooperative learning experts agree that the approach works best if team grades are adjusted for individual performance. If this adjustment is not made, students who do little or no work may receive the same credit as those who do a great deal of work, which is unfair and works against the principle of individual accountability. The students who actually do the work justifiably resent both their less industrious teammates and the instructor who appears to be sanctioning and rewarding laziness and irresponsibility. Some instructors who do not adjust team grades for individual performance argue that they are only simulating the work environment, but they are incorrect. In the professional world, individuals who do not pull their weight on work teams eventually suffer consequences far worse than low grades.

An "autorating" (peer rating) system designed to account for individual performance in cooperative learning team projects has been developed at the Royal Melbourne Institute of Technology (RMIT) by Professor Rob Brown (Brown, 1995). Team members confidentially rate how well they and each of their teammates fulfilled their responsibilities, taking the ratings from a prescribed list of nine terms ranging from "excellent" to "no show." The students are cautioned that they are rating only responsibility of performance and not academic ability or percentage contribution to the project. The instructor assigns numerical values to each rating ("Excellent" = 100, "Very Good" = 87.5, "Satisfactory" = 75,…,"No show" = 0) and computes a weighting factor for each student as the student's individual average rating divided by the team average. The square root of that number may be used instead if the instructor wishes to give less weight to the peer ratings. The student's final project grade is the product of the weighting factor and the team project grade.

The validity of peer ratings can be (and often is) questioned. Common concerns are that individuals will inflate their self-ratings; team members will agree to give everyone identical ratings to avoid conflict; and gender or racial bias and personal dislikes might influence the ratings. This study attempts to assess the validity of these concerns.

CLASS AND TEAM DEMOGRAPHICS

The RMIT autorating system was used in two consecutive sophomore-level chemical engineering courses at North Carolina State University:

CHE 205 – Chemical Process Principles. Fall 1997.

CHE 225 – Chemical Process Systems. Spring 1998.

Table 1 reports demographic data for the students in each course.

Table 1
Demographic Data

Class	N	Men	Women	Non-minorities	Minorities
CHE 205	137	70%	30%	88%	12%
CHE 225	71	70%	30%	92%	8%

N is the number of students who received final course grades. "Minorities" includes African-American students (11% of the students in CHE 205, 7% in CHE 225) and Native American students (<1% in CHE 205, 1% in CHE 225), and "non-minorities" includes Caucasian students and students of all other ethnic backgrounds enrolled in the course. (There were no students of Hispanic background in either course.)

On the first day of class the students filled out questionnaires that asked them to specify sex, ethnicity, grades in prerequisite courses (calculus, chemistry, and physics courses for CHE 205, advanced calculus and CHE 205 for CHE 225), outside interests, and times available for group work outside of class. The students were told that they could skip any questions that they felt intruded on their privacy, but only a few failed to respond to all questions. They were then grouped into teams of three or four by the instructor to assure as much as possible heterogeneity of academic ability (as measured by the prerequisite course grades), commonality of interests, and common blocks of time for meeting outside class. Table 2 shows profiles of the team composition in the two courses.

Table 2
Cooperative Learning Teams

	CHE 205	CHE 225
Total	39	18
All female	6	1
All male	20	8
Mixed gender	13	9
All non-minorities	26	12
Mixed ethnicity	13	6

Homework problem sets, due twice a week in CHE 205 and once a week in CHE 225, were completed in these teams. One student in CHE 205 had a full-time outside job and was allowed to complete assignments individually, but group work was mandatory for all other students. By the end of the semester in CHE 205, four of the 39 teams contained only two members as a result of students dropping the course.

COOPERATIVE LEARNING PROCEDURES

Team members were assigned roles that rotated from assignment to assignment. The *coordinator* organized working sessions and made sure that all team members understood their responsibilities. The *recorder* prepared the final solution set. A *checker* (or two checkers in a team of four) proofread the final solution set, verified that all team members understood both the solutions and the problem-solving strategies used to obtain them, and took primary responsibility for submitting the solution set on its due date.

The teams were periodically asked to submit assessments of how well they were functioning. They were encouraged to see the course instructor if they were having problems of any sort, and in some cases the course instructor sought out teams that reported having difficulties. Occasional mini-clinics were held in class to discuss ways of dealing with problems commonly encountered by cooperative learning teams.

After the first six weeks, the students were told that their teams would be disbanded and reformed unless all members of a team indicated confidentially that they wished to remain together, in which case they would be permitted to do so. Of the 39 teams in CHE 205, only one elected to disband and so could not. All of the teams in CHE 225 elected to remain together.

For more details about the cooperative learning model implemented in the two courses, see Felder (1995, 1998).

PEER RATING PROCEDURE

The peer rating form used in the course is shown in Figure 1. In CHE 205, the students were specifically asked *not* to rate themselves, and the form they received differed from that shown in Figure 1 in only that respect. (Self-ratings are normally not included in the peer rating procedure, but were included in CHE 225 for the purposes of this research study.) Each student received a copy of the form on the first day of each course. The form was briefly explained, and the students were told that they would fill it out at the end of the semester and that their ratings would be used to adjust their average homework grade (which accounted for 15% of their final course grade). They were not told how the adjustment would be done.

Midway through the semester, forms were handed out and the students were instructed to fill them out honestly, show them to their teammates, and discuss reasons for ratings lower than "Very Good." In the last week of the semester, after the last assignment had been turned in, they were given blank forms again and told to fill them out confidentially, sign them, and return them

to the instructor. The explanations of the purpose of the form and the meaning of the ratings were repeated, and the students were cautioned that both fairness and self-interest dictated that they submit their ratings. The instructor logged in the forms and sent e-mail reminders to those who had not submitted them.

A teaching assistant (TA) converted each verbal rating to a numerical equivalent, with "Excellent" = 100, "Very Good" = 87.5, and so on in 12.5-point decrements down to "No Show" = 0. The TA then entered the ratings on a spreadsheet and computed a weighting factor for each student as the student's individual average rating divided by the team average. The student's final homework grade was determined as the product of the weighting factor and the team average homework grade. An illustrative calculation is given in Table 3.

Table 3
Illustrative Peer Rating Analysis

Team HW grade =	**80**							**Indiv. HW Grade**
Name	**Vote 1**	**Vote 2**	**Vote 3**	**Vote 4**	**Indiv. Avg.**	**Tm. Avg.**	**Adj. Fctr.**	
John	87.5	100	75	87.5	87.5	82.8	1.06	85
Betty	87.5	100	87.5	87.5	90.6	82.8	1.09	88
Dave	62.5	75	50	75	65.6	82.8	0.79	63
Angela	87.5	87.5	87.5	87.5	87.5	82.8	1.06	85

Row 1 of Table 3 shows that John (Student 1) received ratings of "Very Good" (= 87.5) from himself and Angela, "Excellent" from Betty, and "Satisfactory" from Dave (horizontal entries for John). The entries in the "Vote 1" column indicate that John gave ratings of "Very Good" to himself, Betty, and Angela and "Ordinary" (= 62.5) to Dave. As mentioned previously, students submitted self-ratings in CHE 225 but not in CHE 205. If Table 3 represented a team in CHE 205, the main (upper left to lower right) diagonal entries in the 4x4 block of ratings would have been blank.

The weighting factor used to determine each individual's homework grade was the individual's average rating divided by the team average. A maximum weighting factor of 1.10 was imposed, so that calculated factors greater than this value were scaled down. This step was taken to preclude students receiving highly inflated homework grades by virtue of having a teammate with very low ratings. If the instructor had wished to attach less importance to the peer ratings, he would have used the square root of this quotient as the weighting factor, and if he had

wished to reduce the grade increase awarded to students on teams with low-rated members, he would have made the maximum weighting factor closer to 1. For a more detailed discussion of the autorating system, see Brown (1995).

We find it interesting that no students in either course ever asked exactly how the descriptive ratings of their teammates would be used to adjust their homework grades. (If any had asked, the instructor would have told them.) Students apparently assume that the ratings will be used in some qualitative manner if they are used at all; it apparently never occurs to them that the descriptive terms ("Excellent," "Very Good," etc.) will be converted to numbers and used to make quantitative adjustments to team grades.

NOMENCLATURE AND DATA ANALYSIS

The following nomenclature will be used in reporting results. *IER* (individual average effort rating) denotes the average numerical peer rating a student received from his or her teammates (not including self-ratings), and *GER* (group average effort rating) denotes the average rating for all team members. *Average test grade* is a weighted average of a student's three individual test grades (weighted at 20% per test) and final examination grade (weighted at 40%), all tests having been graded on a 0–100 basis.

PGPA (prior grade-point average) is a student's cumulative grade-point average scaled to a 0–100 basis for semesters up to but not including the one that included CHE 205 or CHE 225. The scaling formula is [GPA(0-100) = 12.5 x GPA(A=4) + 50]. *Normalized test grade* is the difference between a student's average test grade and his or her PGPA. Loosely speaking, the normalized test grade is a measure of performance relative to grades in prior courses: the higher the normalized grade, the better the performance relative to pre-course expectations. PGPAs for sixteen students in CHE 205 were unavailable for various reasons, and so these students were omitted from statistical tests involving normalized test grades.

In CHE 205, twelve of the 137 students did not submit peer ratings. Six of these students were male, six were female, and two were minorities. In three groups, only one student submitted ratings. These students and groups were excluded from analyses involving IER and GER values. In CHE 225, all students submitted peer ratings for their teammates. One student did not submit a self-rating.

All reported levels of significance are derived from nonparametric Wilcoxon rank-sum tests unless otherwise noted, with "statistically significant" defined as $p<0.1$. Pearson

correlations were used to test for association between average student ratings and student performance in the class.

RESULTS

Correlations between ratings and grades

Peer ratings correlated positively with average test grades in CHE 205, as shown in Figure 2 [R=0.54, p=0.0001]. The correlations between IER and average test grade were even stronger for women (R=0.76, p=0.001) and minority students (R=0.79, p=0.005). In CHE 225 the correlation between peer ratings and test performance was weaker but still statistically significant (R=0.32, p=0.005), and the correlations for female and minority populations were not statistically significant. These results might mean that the students who performed best in the course tended to be the most diligent and responsible in carrying out assigned tasks, or they could indicate that the students were not all following the instructions to make team citizenship (as opposed to ability or percentage contribution) the basis of the peer ratings.

Of students entering CHE 205 with a PGPA less than 3.0, those with IER>80 earned an average normalized grade of –12.3 and those with IER<70 earned a normalized grade of –28.9. The difference is significant at the 0.06 level (1-tailed test). The implication is that the performance relative to expectations of good team citizens exceeded that of poor team citizens. The negative values of the normalized grades have no intrinsic significance but simply reflect the arbitrary formula used to convert the PGPA from a 0–4 basis to a 0–100 basis.

Correlations between self-ratings and ratings from teammates in CHE 225

The average IER in CHE 225 was 89.1 and the average self-rating was 90.0, a statistically insignificant difference. Similarly, self-ratings of male, female, minority, and non-minority students were not statistically different from ratings they received from teammates. Contrary to expectations, inflated self-ratings proved to be less common than *deflated* self-ratings. Four students (6% of the class) gave themselves ratings at least one level higher than the highest rating they received from a team member. Three of these students were non-minority males and the other was a non-minority female. Ten students (14% of the class) gave themselves ratings at least one descriptor value *lower* than the lowest rating they received from their teammates. Eight of these ten students were male, and none was a minority student.

Gender effects on performance and ratings

Men entered CHE 205 and CHE 225 with higher PGPAs than women and earned higher test averages and higher normalized grades than women in both classes. In CHE 205, the differences between the average grades earned by men and women are marginally significant (Table 4). Men gave their teammates slightly higher ratings and received slightly higher ratings from their teammates than did women in both classes (Table 5). The differences are not significant.

Table 4
Gender Differences in Performance

	CHE 205			CHE 225		
	Women	Men	p	Women	Men	p
PGPA	89.7 (N=38)	91.3 (N=83)	0.25	91.1 (N=21)	92.8 (N=49)	0.34
Average test grade	71.6 (N=41)	78.1 (N=96)	0.06	78.2 (N=21)	82.2 (N=50)	0.27
Normalized grade	–17.6 (N=38)	–12.6 (N=83)	0.09	–12.9 (N=21)	–10.9 (N=49)	0.64

Table 5
Gender Differences in Average Ratings Received and Given

	Received by (IER)			Given by		
	Women	Men	p	Women	Men	p
CHE 205	86.0 (N=89)	87.5 (N=246)	0.67	85.7 (N=91)	87.6 (N=244)	0.76
CHE 225	87.7 (N=59)	89.7 (N=145)	0.19	87.5 (N=59)	89.7 (N=145)	0.81

The difference between male and female self-ratings in CHE 225—89.1 and 88.7, respectively—is not significant. Group-average ratings (GER) were slightly higher for mixed-gender teams than for same-sex teams in CHE 205 (89.8 for mixed-gender teams vs. 86.7 for same-sex teams, p=0.09), and the opposite result was observed in CHE 225 (90.0 for same-sex teams, 87.3 for mixed-gender teams, difference not significant).

Effects of ethnicity on performance and ratings

Minority students entered CHE 205 and CHE 225 with grade-point averages significantly lower than those of their non-minority counterparts and on average earned lower test grades in the two courses (Table 6). The differences were statistically significant in CHE 205.

Table 6
Academic Performance by Ethnic Groups
(M = minority, NM = non-minority)

	CHE 205			CHE 225		
	M	NM	p	M	NM	p
PGPA	87.2 (N=15)	91.3 (N=106)	0.08	91.1 (N=6)	92.4 (N=64)	0.47
Average test grade	62 (N=16)	78 (N=121)	0.005	77.8 (N=6)	81.3 (N=65)	0.21

In both classes, minority students on average received lower ratings and gave higher ratings than did non-minority students (Table 7). The differences are significant in CHE 225. In CHE 205, non-minority students received similar ratings from both minority and other non-minority students. Non-minority students gave higher ratings to other non-minority students than to minority students; the difference is not statistically significant in CHE 205 and highly significant in CHE 225. Minority students in CHE 205 gave higher ratings to other minority students than to non-minority students, but the difference was not significant. With no more than one minority student per team in CHE 225, minority students did not have the opportunity to rate other minority students in the class.

Table 7
Average Ratings Received and Given by
Minority(M) and Non-Minority(NM) Students

	CHE 205			CHE 225		
Average ratings given	N	Rating	p	N	Rating	p
By non-minorities	297	86.9	.81	186	88.6	.05
By minorities	38	88.5		18	94.4	
To non-minorities	294	87.7	.38	186	90.3	.0004
To minorities	41	82.9		18	77.1	
By non-minorities to non-minorities	262	87.6	.19	168	89.8	.0008
By non-minorities to minorities	35	81.4		18	77.1	
By minorities to non-minorities	32	87.9	.42	18	94.4	—
By minorities to minorities	6	91.7		—	—	

Minority students also gave themselves lower self-ratings in CHE 225 than did non-minority students (87.5 versus 89.1, respectively), but the difference was not significant.

Students in ethnically heterogeneous teams had roughly the same average PGPA as students in teams that did not include minority students in both CHE 205 (90.1 vs. 91.1, p=0.58)

and CHE 225 (93.3 vs. 92.1, p=0.69). In CHE 205, students in the heterogeneous teams earned lower normalized grades than did students in teams that did not include minority students, with the difference being statistically significant (Table 8). The heterogeneous teams in CHE 225 did slightly better than the homogeneous teams in terms of normalized grades. The average GER for heterogeneous teams was somewhat lower than that for all non-minority teams in both courses, but the differences were not significant.

Table 8
Effects of Ethnic Heterogeneity of Teams on Normalized Grades and GERs

	Normalized grades: non-minority teams	Normalized grades: mixed ethnicity teams	p	GER: non-minority teams	GER: mixed-ethnicity teams	p
CHE 205	–12.1 (N=79)	–17.6 (N=42)	0.002	88.4 (N=25)	86.6 (N=12)	0.73
CHE 225	–11.8 (N=47)	–10.9 (N=24)	0.67	90.0 (N=12)	88.1 (N=6)	0.71

Incidence of identical ratings

In two (≈ 6%) of the CHE 205 teams and two (≈11%) of CHE 225 teams for which more than one team member submitted ratings, all peer ratings were identical. For example, each student in Team 1 of CHE 205 received a rating of 100 from each of his teammates. The incidence of individual students giving identical ratings to their teammates is of course much higher, as shown in Table 9. Men were more likely than women and minority students more likely than non-minority students to give identical ratings.

Table 9
Percentages of Students Giving Teammates Identical Ratings

	CHE 205	CHE 225
Students giving teammates identical ratings	40% (N=48)	49% (N=34)
Female students giving teammates identical ratings	29% (N=9)	38% (N=8)
Male students giving teammates identical ratings	44% (N=39)	53% (N=26)
Minority students giving teammates identical ratings	54% (N=7)	67% (N=4)

Non-minority students giving teammates identical ratings	37% (N=41)	47% (N=30)

Use of ratings to identify hitchhikers, dysfunctional teams, tutors, and effective teams

Hitchhiker is a cooperative learning term for a student who shirks his or her responsibilities in a team. Unless measures are instituted to assure individual accountability, hitchhikers receive the same grades as the more industrious team members who do the bulk of the work, and so get a "free ride." Educators who have reservations about cooperative learning often cite the possibility of successful hitchhiking as a drawback of the approach.

Peer rating provides a mechanism for identifying hitchhikers in a course. Many students are inclined to cover for teammates who occasionally miss team meetings or fail to contribute to problem solutions; however, we believe they are unlikely to give good ratings to students who chronically fail to participate in team efforts. Granted, shirking responsibility is only one of several possible causes for low ratings: they may also be received by students who attempt to dominate their teammates or by bright students who do all of the work themselves and refuse to involve their teammates in the effort. However, our experience is that *consistently* low ratings are most likely to be given to students who are perceived as failing to pull their weight on the team.

In this study, we use the term "hitchhiker" to denote students whose average peer ratings are less than 75—i.e., students whose citizenship is rated as less than satisfactory by their teammates. As shown in Table 10, the incidence of hitchhikers in both classes was very low. Roughly 7% of the students in each class received less than satisfactory ratings from their teammates. The average test scores for 10 of the 14 students in this category were below a "C" level (the level required to advance to the next course in the curriculum). Although no causal relationship can be inferred, irresponsibility in teamwork clearly correlates with poor academic performance.

Table 10
Incidence of Hitchhikers

	$70 \leq$ Test Average ≤ 100	Test Average < 70
CHE 205	< 1% (N=1)	6.1% (N=8)
CHE 225	4.3% (N=3)	2.9% (N=2)

Teams with less-than-satisfactory group average ratings (GER<75) or for which two or more team members fail to submit ratings may be termed *dysfunctional*. The three teams listed in Table 11 fit this criterion. Two of the three teams also received failing average test grades.

Table 11
Incidence of Dysfunctional Teams

Class	Team Number	GER	Team-average test grade
CHE 205	11	69.8	57.7
CHE 205	35	No ratings submitted	40.6
CHE 225	18	68.1	73.5

Students who received average ratings of 100 from their teammates may be characterized as *tutors*, the presumption being that they went beyond their required duties and provided substantial assistance to their teammates. (Possible exceptions to this characterization might be students in the few teams where everyone was rated "Excellent" by everyone else.) *Effective teams* are defined as those for which the GER is greater than 90. The incidences of tutors and effective teams in both courses are shown in Table 12.

Table 12
Incidences of Tutors and Effective Teams

	CHE 205	CHE 225
Number of students with IER=100	18% (N=25)	11% (N=8)
% of teams with at least 1 member with IER = 100	46% (N=18)	33% (N=6)
% of teams with GIER > 90	33% (N=13)	44% (N=8)

CONCERNS ABOUT PEER RATINGS

In this section we examine common concerns about peer ratings in light of the results of this study.

Students will agree to give one another identical ratings

Usually the first concern raised about peer rating methods is that most teams will agree among themselves to give everyone a rating of "excellent" or "very good." (With the autorating system, it makes no difference which rating they settle on, since the grade adjustment factor will

be 1.0 regardless of their choice.) The fact that only 6% of the student teams in one course in this study and 11% in the other course did so (Table 9) suggests that this concern may be unfounded.

Moreover, there is nothing wrong with team members reaching such an agreement. Their doing so is likely to be an indication that the team was working well, with everyone pulling his or her weight throughout the semester. If they all receive identical peer ratings they will all earn the team project grade, which is exactly what should happen.

Students will inflate their self-ratings

A second common concern is that students will inflate their self-ratings to give themselves an advantage when the project grades are computed. The study results show that this concern may also be largely unfounded, although isolated instances of inflation are certainly possible. The average self-rating in CHE 225 was 90.0 and the average rating from teammates was a statistically indistinguishable 89.1. Similarly, self-ratings of male, female, minority, and non-minority students were not statistically different from ratings received from teammates. Roughly 6% of the CHE 225 students gave themselves at least one rating higher than any of the ratings they received from their teammates. None of these students earned a higher course grade as a consequence of their ratings.

A greater concern than inflation of self-ratings may be *deflation*. Fourteen percent of the CHE 225 students gave themselves lower ratings than they received from any of their teammates. One of them claimed that while his teammates believed he was well prepared and cooperative, he himself knew he could have done better. Fortunately, the course grades received by these students were not affected by their modest self-ratings. As long as peer ratings are not given excessive weight in course grading, a situation in which an inflated or deflated self-rating affects a course grade is unlikely to arise.

Students will give ratings based on personal prejudices

In both CHE 205 and CHE 225, women on average received lower ratings from their teammates and gave lower ratings to their teammates than did men (Table 5), but the differences were not statistically significant. Minority students on average received lower ratings and gave higher ratings than non-minority students, with the differences being statistically significant in CHE 225 (Table 7). In CHE 205 minority students gave higher ratings to other minority students and non-minority students gave higher ratings to other non-minority students, but neither difference is significant, and in CHE 225 non-minority students gave significantly higher ratings

to other non-minority students than to minority students (Table 7). No ratings were given by minority students to other minority students in this course.

The potential for gender and racial bias influencing peer ratings certainly exists; however, the data provide no basis for concluding anything about such an influence in this study. Lower ratings received by women and minority students could have resulted from any of a number of causes, including the following ones:

1. The students with lower ratings were generally less diligent and/or responsible than their teammates, and so the ratings were fully justified.
2. The students with lower ratings were weaker academically than their teammates. Although the ratings were supposed to reflect only team citizenship, stronger team members would be most likely to make the greatest contributions to the team effort and so to get higher ratings.
3. The students with lower ratings tended to be relatively passive in team sessions and so were perceived as contributing less to the team than their more vocal teammates contributed.
4. The results were due to statistical chance and would not be replicated in other studies.
5. The students with lower ratings were the victims of prejudice.

Although we cannot confirm or refute any of these explanations on the basis of the available data, we have cause to believe that the second and third factors may have played substantial roles in accounting for gender and ethnic differences in ratings that were not the result of pure statistical chance. For whatever reasons, women on average earned lower grades than men and minorities earned lower grades than non-minorities (Tables 4 and 6). The correlation between ratings and test grades was extremely high for minorities ($R = 0.79$), which provides a strong argument for the second of the listed explanations. Moreover, research studies have shown that members of minority cultures (and women in engineering still constitute a minority culture) tend to play more passive roles in mixed teams (see Felder *et al*., 1995), suggesting the applicability of the third explanation. In short, while biased peer ratings cannot be ruled out, other explanations for the observed results seem more likely.

Students will complain about having their grades affected by peer ratings

There is almost nothing an instructor can do in a class—lecture or actively involve students, assign teamwork or give only individual assignments, give unadjusted team grades or use peer ratings to adjust the grades for individual effort—that will eliminate all student

complaints. The complaints become a matter of serious concern only if they are widespread or involve charges of academic policy violations against the instructor.

As far as we know, no academic policy at any university precludes using peer ratings as part of a teamwork assessment process. Widespread objections to peer ratings or to group work in general might arise in some circumstances, but they did not arise in the courses described in this study (nor have they arisen in any other course in which the instructor used cooperative learning). In fact, only one negative comment about the autorating system appeared in course evaluations in either CHE 205 or CHE 225. At the end of the first semester several students whose homework grades were lowered by peer ratings complained, but when questioned they admitted that they missed several team meetings and came to others unprepared, and they acknowledged that they had been warned about the possible consequences of such behavior. We believe that the lessons in responsibility learned by these students through the autorating system could eventually prove more valuable than anything else they may have learned in the course.

If anything, the use of the autorating system *reduces* the number of complaints typically made by students in a cooperative learning class. By far the most common complaint is about hitchhikers—students who are not contributing to their team but continue to get credit for the team's work. When students know that the hitchhikers will *not* receive the same grade as the workers, they are much less inclined to complain about the unfairness of cooperative learning. Moreover, although we cannot prove it, we believe that the occurrence of hitchhiking will be reduced by the knowledge that there will be a penalty for those who try it.

CONCLUSIONS AND RECOMMENDATIONS

A modified version of a peer rating system developed at the Royal Melbourne Institute of Technology was used in two sophomore chemical engineering courses in which the students completed homework assignments in cooperative learning teams. The students were instructed to rate their teammates and (in one of the courses) themselves on the basis of team citizenship, as opposed to academic ability or percentage of the work actually done by each student. The rating system is easy to administer and to use for individual grade adjustments. It provides a modest reward to students who go above and beyond the minimum required individual effort in teamwork and effectively identifies "hitchhikers" and keeps them from getting full credit for work done by their more responsible teammates. Following are the principal observations and conclusions of the study.

- Differences between student self-ratings and average ratings received from teammates were insignificant. The common faculty concern that many students will inflate their own ratings to give themselves an advantage over their teammates was not borne out by this study. In fact, *deflated* self-ratings were more commonly observed than inflated ones.
- Only two teams in each class (out of 39 teams in CHE 205 and 18 teams in CHE 225) submitted identical ratings for all team members. This result allays the common faculty concern that most teams will agree on a common rating if peer ratings are used.
- Peer ratings exhibited significant positive correlations with test grades: i.e., the students who were rated highest on team citizenship tended to do better in the course than students who received lower ratings. Students who show up prepared for work sessions and contribute actively to the team process—the principal requirements for high peer ratings—thus do better on tests than students who do not, although there is no basis for inferring a causal relationship.
- In both courses, women received lower ratings from their teammates and gave lower ratings to their teammates than did men, but the differences were not statistically significant. Non-minority students on average gave lower ratings to minority students than to other non-minority students, with the difference being statistically significant in the second course. Gender and racial bias could be a factor in these results, but alternative explanations suggested in the paper are considered more likely.

The study results thus show that most of the concerns frequently raised about peer ratings in cooperative learning may be unfounded, with a possible exception being the potential influence of personal prejudice in assigned ratings. The latter issue is an appropriate subject for further study.

We believe that as successful as the peer rating system used in this study was, more might be done to make peer ratings as effective as they could be. Above all, we recommend providing the students with more guidance and practice in assigning ratings than we provided in CHE 205 and CHE 225. In their excellent reference on cooperative learning in higher education, Millis and Cottell (1998) present a peer evaluation form that assigns numerical ratings to four different components of effective teamwork: attending meetings on a regular basis, making an effort at assigned work, attempting to make contributions and/or to seek help within the group when needed, and cooperating with the group effort. We are currently using results from a modified version of this form to derive quantitative definitions of the RMIT system terms

("Excellent"…"No Show") that should make peer ratings more objective and less subject to personal feelings. We also suggest taking some time in class to present several team scenarios, have the students fill out rating sheets for the hypothetical team members, and then discuss the ratings and reach consensus on what they should be.

REFERENCES

Brown, R.W. (1995). Autorating: Getting individual marks from team marks and enhancing teamwork. *1995 Frontiers in Education Conference Proceedings*. Pittsburgh, IEEE/ASEE, November 1995. For a reprint, contact Rob Brown at rwb@rmit.edu.au.

Felder, R.M. (1995). A longitudinal study of engineering student performance and retention. IV. Instructional methods and student responses to them. *J. Engr. Education, 84*(4), 361–367. This paper may be viewed at <http://www2.ncsu.edu/unity/lockers/users/f/felder/public/Papers/long4.html>

Felder, R.M. (1998). A Longitudinal Study of Engineering Student Performance and Retention. V. Comparisons with Traditionally-Taught Students. *J. Engr. Education, 87*(4), 469–480. This paper may be viewed at <http://www2.ncsu.edu/unity/lockers/users/f/felder/public/Papers/long5.html>

Johnson, D.W., R.T. Johnson, and K.A. Smith. (1998). Active learning: Cooperation in the College Classroom. Edina, MN: Interaction Book Co.

Millis, B.J., and P.G. Cottell, Jr. (1998). *Cooperative learning for higher education faculty*, p. 198. Phoenix: Oryx Press.

PEER RATING OF TEAM MEMBERS

Name_____ **Team #**_____

Please write the names of all of your team members, INCLUDING YOURSELF, and rate the degree to which each member fulfilled his/her responsibilities in completing the homework assignments. The possible ratings are as follows:

Excellent	Consistently went above and beyond—tutored teammates, carried more than his/her fair share of the load
Very good	Consistently did what he/she was supposed to do, very well prepared and cooperative
Satisfactory	Usually did what he/she was supposed to do, acceptably prepared and cooperative
Ordinary	Often did what he/she was supposed to do, minimally prepared and cooperative
Marginal	Sometimes failed to show up or complete assignments, rarely prepared
Deficient	Often failed to show up or complete assignments, rarely prepared
Unsatisfactory	Consistently failed to show up or complete assignments, unprepared
Superficial	Practically no participation
No show	No participation at all

These ratings should reflect each individual's level of participation and effort and sense of responsibility, not his or her academic ability.

Name of team member	**Rating**
_____	_____
_____	_____
_____	_____
_____	_____

Your signature: _____

©R.M. Felder, 1997.

Figure 1. Peer rating form

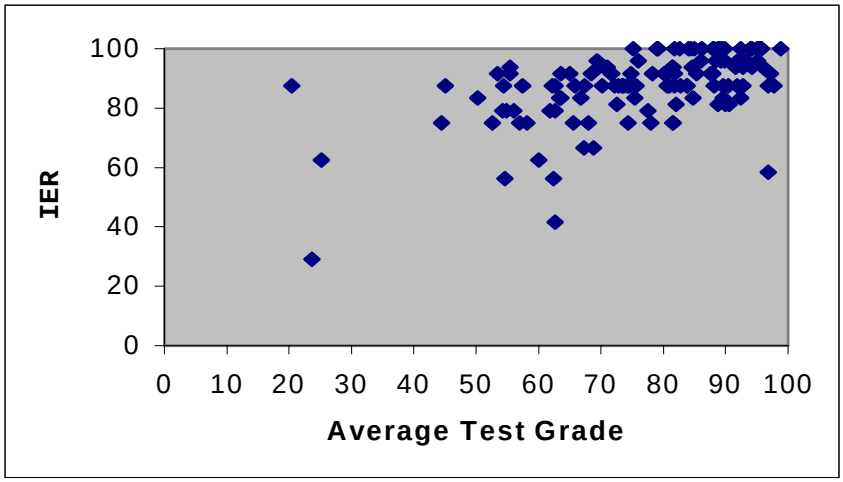

Figure 2. Correlation plot of IER versus average test grade in CHE 205

DEBORAH KAUFMAN

Deborah Kaufman is currently a doctoral student in the Department of Chemical Engineering at North Carolina State University. She received her B.S. in Chemical Engineering from Cornell University. Her dissertation is in the area of bioseparations, and she is also participating in N.C. State's Preparing the Professoriate mentorship program. She will begin a job with the Process Sciences group at Genentech in the Spring of 2000.

RICHARD FELDER

Richard Felder is Hoechst Celanese Professor Emeritus of Chemical Engineering at North Carolina State University and Faculty Development Codirector of the NSF-sponsored SUCCEED Coalition. He received his B.Ch.E. in Chemical Engineering from the City College of New York, and his M.A. and Ph.D. in Chemical Engineering from Princeton University. He is co-author of *Elementary Principles of Chemical Processes* (Wiley, 1978, 1986, 2000) and codirector of the National Effective Teaching Institute. Many of his papers on education-related topics can be found at < *www2.ncsu.edu/effective_teaching/* >.

HUGH FULLER

Hugh Fuller is Director of Educational Assessment for the North Carolina State University College of Engineering. Prior to taking this position, he was Director of Institutional Research and Director of the N.C. State Academic Skills Program. His current interests include assessing attitudes toward engineering and confidence levels of first-year engineering students.