

## Peer Evaluation in a Mandatory Cooperative Education Environment

Matthew W. Ohland and Cynthia J. Finelli  
Clemson University / Kettering University

### Abstract

Previous studies with a majority white student population [1] and with a majority black population [2] demonstrated that peer ratings to assess individual performance of team members showed no effects relating to gender but significant effects relating to race. For this study, the student population is majority white, but is from Kettering University, an institution that requires participation in a cooperative education program from the first freshman term. We found in this case no difference among peer ratings based on either gender or race, but the small sample size of female and minority students in the Kettering student population makes further investigation necessary to confirm this result. Supporting the observed trend, we note that the students in the Kettering population are rated more uniformly in teamwork contribution by their peers than in the previously cited studies—there is less score variation. We hypothesize that the required cooperative education experience of the Kettering students has improved their teamwork skills and that the peer ratings correctly reflect this. Further study is proposed to evaluate this hypothesis.

### I. Introduction

Establishing individual accountability in team assignments is of primary importance in every recognized approach to cooperative learning. It is unfair for those who fail to do their share of the work to receive the same grade as the rest of the group, a phenomenon called “hitchhiking.” Brown [3] describes one method of measuring the relative contribution of each team member using a peer-evaluation or “autorating” system to assign individual grades based on a group grade. For this technique to be effective, groups should be assigned and coached by the instructor according to the established practices of cooperative learning. In the study described here, the peer evaluations completed by students do not actually influence student grades, but are used as formative and summative feedback for students.

This paper expands a body of research studying a peer evaluation instrument in a variety of contexts. Kaufman, Felder, and Fuller [1] used a modified form of Brown’s instrument at North Carolina State University. (The acronym “KFF” is used here to refer to this paper.) In the majority white population of the KFF study, no gender bias was detected, and minority students on average received lower ratings and gave higher ratings than non-minority students, with the differences being statistically significant in one class but not in another class. Layton and Ohland [2] confirmed the same conclusions in a study at North Carolina A&T State University, in a majority black population—there was still no gender bias, and minority students (mostly black) were again found to receive lower ratings and give higher ratings than non-minority students. (“L&O I” refers to this study.) In an effort to eliminate the observed racial bias, the instrument for peer evaluation and the administration of it were modified in the present study and in a second study by Layton and Ohland (The study parameters for the second study by Layton and Ohland (“L&O

II”) are included in some tables below for reference; this study is included in these same proceedings [4].)

The use of this instrument at Kettering University provides opportunity to verify its usefulness under a broad set of conditions. The racial and gender profile of this study is similar to that of the KFF study, but there are significant institutional differences. Kettering is a primarily undergraduate private college that offers Bachelor Degrees in five engineering disciplines, four science and mathematics areas, and management. All of the undergraduate students participate in a mandatory cooperative education program from their first freshman semester whereby they alternate twelve-week semesters on campus and at their worksite. The institutions studied previously, North Carolina State University and North Carolina A&T State University, do not have a required co-op experience.

## II. Class and team demographics

Peer evaluations from two separate electrical engineering courses offered at Kettering University are included in this study. The first is ECE 412–Digital Signal Processing, a senior-level elective course, and the second is ECE 440–Electromagnetic Fields II, a senior-level required course. For these courses, 75% and 40% respectively of the overall course grade was based on team assignments. The racial profile of the Kettering students is similar to that of the KFF study, while the gender profile is similar to that of L&O II. Table 1 presents the demographic distribution of each of the studies mentioned. The data from both courses in this study are identified by the acronym “O&F” in the tables that follow.

Table 1: Demographic data.

<i>Study</i>	<i>N</i>	<i>Men</i>	<i>Women</i>	<i>Non-minorities</i>	<i>Minorities</i>
O&F	62	82%	18%	92%	8%
L&O II	70	85%	15%	10%	90%
L&O I	70	73%	27%	13%	87%
KFF	208	70%	30%	89%	11%

Here, *N* is the number of students receiving final course grades. “Minorities” includes African-American, Hispanic, and Native American students (if any); “non-minorities” includes students of all other ethnic backgrounds.

For this study, teams of three or four students were formed at the beginning of the course. Student grades in a pertinent prerequisite course were compiled by the instructor prior to the first day of class, and students were categorized into one of three ability levels based on this grade. Lists of students in each ability category were generated and (without revealing the manner in which the lists were created) students were asked to make teams of three students by selecting one person from each list to constitute a team. Teams of four, allowed only when necessary, have an extra person from one of the lists. At mid-semester, teams were told that they could disband and form new teams if at least two teams in the course so desired; this did not happen in either of the courses. The final composition of the teams is shown in Table 2.

Table 2: Team composition.

<i>Study</i>	Number of teams in each category						
	Total	All female	All male	Mixed gender	All minority	Mixed ethnicity	All non-minority
O&F	20	0	10	10	0	5	15
L&O II	17	0	11	6	11	6	0
L&O I	21	1	11	9	12	9	0

### III. Team activities and the cooperative learning environment

Both courses in this study required team homework assignments. In addition, the first course (ECE 412) included three team laboratory assignments and a final team project, while the second course (ECE 440) included a final team project. Students were expected to meet outside of normal class hours to complete the assignments, and they were requested to submit a single assignment reflecting the effort of all team members for each of the team assignments.

Cooperative learning has been shown to have strong benefits in the classroom [5]. In the courses described here, use of cooperative learning was a primary goal, and the instructor was candid with her students about this from the beginning of the course. Cooperative learning was informally included in the courses by interspersing short active learning tasks throughout the scheduled class time. During these activities, students interacted with others sitting near them. As the semester progressed, students tended to sit with members of their formal teams and to interact with those same students during the course, but this was not necessarily true. Cooperative learning was more formally used through the extensive team activities in these courses. The degree to which cooperative learning was achieved varies between the courses, but the five basic elements of cooperative learning were included in the following ways:

Positive interdependence. A high percentage of course grade (from 40–75%) was based on team activities in each course. For each team assignment, all members of the team were given a single grade; thus, for the individual team members to succeed, they all had to work together. In addition, to encourage cooperation in preparation for exams, the second course included bonus points based on the performance of all members of a particular team. For example, if each member of a team earned greater than 80% on the first test, then all members of that team had one point added to their final course grade.

Individual accountability. Since all members of a team received a single grade for each team assignment, students felt a certain degree of “peer pressure.” Some of the students indicated that they didn’t want to “let their teammates down.” Further, individual tests and the final exam ensured that each student mastered the course material.

Face-to-face promotive interaction. This was not guaranteed, but it clearly improved during the semester. Material submitted for the team assignments evolved from problems completed by separate team members hastily stapled together to complete, consistent assignments obviously requiring a significant team effort.

Use of teamwork skills. It was mostly assumed that students used effective teamwork skills when completing the team assignments. Because of Kettering’s mandatory cooperative education pro-

gram, most students have industry experience, particularly working with diverse teams, prior to enrolling in any of these courses. As such, the students seem to work effectively in teams.

Group processing. By periodically assessing the team's effectiveness using the peer-rating procedure, students reflected on the performance of their own team and were able to refocus the team's activities if necessary.

#### IV. Peer rating procedures

At various times throughout each course, students were asked to complete a peer-rating form to evaluate the performance of individuals on the team. Students were told that the data was for informational purposes only – the instructor did not use the ratings in any official grade assignment. The form included a rating system with nine possible ratings. A definition of each rating was included on the form, as were examples of how each rating could be satisfied. Students were asked to rate all team members, including themselves, regarding the degree to which that person fulfilled his/her responsibilities in the team assignments. The forms were to be completed independently, during the class period, and each person signed their form. The peer evaluation form is identical to the one used by KFF and L&O II, with instructions given per the recommendations made in L&O I. These instructions intend to focus student evaluations on team contribution rather than ability or any other measure.

Data from the peer-ratings completed at intermediate points during the semester served to educate the students about the peer evaluation procedure, to allow the students to reflect on the evaluation criteria, and to alert the instructor to teams that were having difficulties. Data from forms completed at the conclusion of the course are analyzed in this study.

#### V. Data analysis

The raw data corresponding to the nine possible ratings of the peer evaluation form were converted to a numeric value from 0 to 100. For the current study, forms evaluating all team members were completed by 62 students. After eliminating student self-ratings, a total of 133 ratings are studied. Comparisons of ratings by gender and by race/ethnicity are given here. All reported levels of significance are derived from a nonparametric Wilcoxon (Mann-Whitney) rank-sum test [6], with statistical significance defined by  $p < 0.05$ . Ratings given by students in this study are summarized by gender in Table 3.

Table 3: Ratings by gender for this study.

Average ratings given	N	Rating	p
By men	111	87.4	0.34
By women	22	90.3	
To men	111	87.7	0.81
To women	22	88.6	
By men to men	91	87.2	0.87
By men to women	20	88.1	
By women to men	20	90.0	0.57
By women to women	2	93.8	

There is no demonstrated gender bias of either practical or statistical significance. Ratings given and received by minorities and non-minorities in this study are summarized in Table 4.

Table 4: Ratings by race/ethnicity.

Average ratings given	N	Rating	p
By minorities	8	87.5	0.80
By non-minorities	125	87.9	
To minorities	9	88.9	0.62
To non-minorities	124	87.8	
By minorities to minorities	–	–	–
By minorities to non-minorities	8	87.5	
By non-minorities to minorities	9	88.9	0.62
By non-minorities to non-minorities	116	87.8	

Again, there are no differences of either practical or statistical significance. Note that no minorities rated other minorities in this study. It is relevant to investigate any differences in final course grade that may exist. Tables 5 and 6 show the average course grade by race/ethnicity and by gender, respectively, for the present study and for the others described earlier.

Table 5: Grades by race/ethnicity for this study.

Study	Course	Ethnicity	Average grade	p
O&F		Non-minorities	87.3	0.55
		Minorities	88.9	
L&O II		Non-minorities	80.3	0.02
		Minorities	64.9	
L&O I	MEEN 440	Non-minorities	84.0	0.0003
		Minorities	76.8	
KFF	CHE 205	Non-minorities	78.0	0.005
		Minorities	62.0	
	CHE 225	Non-minorities	81.3	0.21
		Minorities	77.8	

Table 6: Grades by gender.

Study	Gender	Average grade	p
O&F	Men	87.0	0.14
	Women	89.7	
L&O II	Men	67.7	0.49
	Women	61.3	

In KFF, non-minorities have significantly ( $p=0.005$ ) higher test grades than minorities in one course (CHE 205) and higher, but not significantly higher ( $p=0.21$ ), test grades in the other course (CHE 225). Similarly, in both L&O I and L&O II non-minority students received higher grades. In this study, however, no significant grade differences were observed either by race/ethnicity or by gender. Again, please note that the sample size of minorities and women is small (11 women and 5 minorities). We hypothesize, however, that since the grades received by

women and men and those received by racial majority/minority students are so similar, that this trend will persist in further study. While comparison of student grades at different institutions (especially in different courses) is always problematic, it should be noted that in this case, it is particularly inappropriate—Kettering has a grading policy in which a numeric grade of 93 is required to make a letter grade of “A,” 85-92 receives a “B,” 78-84 receives a “C,” and 70-77 receives a “D.” While it is not the goal of this study to compare grades across institutions, it bears mentioning here to ensure that this policy difference does not affect our analysis. Fortunately, t-test results are independent of reference frame, so scaling and shifting the original data set (as would be necessary to adjust Kettering grades to the same scale used by other institutions) have no effect on the outcome.

## VI. Conclusions and recommendations

To date, the results of this study are in agreement with the objectives desired using this peer evaluation instrument and with the instructions given to encourage evaluation of team contribution rather than ability. While in both the KFF and the L&O I studies there was a racial/ethnic bias, this study shows none. A causal relationship between the choice of instrument and administration and the lack of bias is impossible to prove with small sample sizes. It is noted that there is significantly less score variability in this study (standard deviation = 11.7) than in either L&O I (standard deviation = 14.8) or L&O II (standard deviation = 13.3). This reduced variation makes it more likely that different sub-populations (separated by race/ethnicity or by gender) will have similar scores. We suggest two plausible explanations for why there is less variation in this case:

For the first scenario, Kettering students, because of the mandatory cooperative education, are experienced at working in teams and are better at it. This causes the range of variation of team performance to be restricted. A second possibility is that the reduction in the variability of the peer evaluation of team performance mirrors a reduction in the variability of grades. The distribution of final grades at Kettering had an average of 87 and a standard deviation of 6.3 points. In contrast, the values of the average and standard deviation of the grade distribution for L&O I ( $79 \pm 9$ ) and L&O II ( $67 \pm 15$ ) indicate both significantly lower grades and significantly higher variability. The compression of the range of numeric grades assigned to students Kettering does not account for all of the reduction in variability.

If the first case above were the correct explanation, one might expect that the peer ratings of Kettering students would be significantly higher than those of the students in the L&O studies, yet this is not the case. It is possible, however, that the populations of the L&O studies have different perspectives regarding what comprises an “excellent” team contribution. The Kettering students may hold their better-performing classmates to a higher standard based on their co-op experience. This is an issue that merits a reliability investigation [7]—if, in fact, Kettering students have both a restricted range of peer ratings and a different standard of measurement, we should expect the reliability of the instrument to break down when used in the Kettering environment. This is recommended as a subject for further study.

If the second possibility were true, however, one should observe a correlation of peer ratings and course grades in L&O I and L&O II. This correlation is not significant, despite the fact that in

those studies the peer rating is actually used to calculate a portion of the grade (which necessarily contributes to some level of positive correlation). To further examine the plausibility of the second case above, the instrument should be administered to additional populations in which there is significant grade variability. If it can be shown that there is no significant relationship between peer ratings and course grades, then one could be more confident in supporting the first explanation.

We recommend continued verification of both the validity and reliability of this peer evaluation instrument. It is desirable to have an instrument that is both valid and reliable when used with a broad range of student populations. The addition of further cohorts to this study will enhance the population of women and racial minorities to levels that will establish confidence in the results.

## VII. Acknowledgements

We gratefully acknowledge Richard Layton of Rose-Hulman Institute of Technology for his contribution to the foundation of this work.

## References

- [1] Kaufman, D.B., R.M. Felder, and H. Fuller. (1999). Peer ratings in cooperative learning teams. In proc. *ASEE Annual Conference*. ASEE, Charlotte, NC, June 1999.
- [2] Layton, R.A. and M.W. Ohland (2000). Peer evaluations in teams of predominantly minority students. In proc. *ASEE Annual Conference*. ASEE, St. Louis, MO, June 2000.
- [3] Brown, R.W. (1995). Autorating: Getting individual marks from team marks and enhancing teamwork. In proc. *Frontiers in Education Conference*. IEEE/ASEE, Pittsburgh, PA, November 1995.
- [4] Layton, R.A. and M.W. Ohland (2001). Peer ratings revisited: Focus on teamwork, not ability. *ASEE Annual Conference*. ASEE, Albuquerque, NM, June 2001.
- [5] Johnson, D.W., R.T. Johnson, and K.A. Smith (1998). *Active Learning: Cooperation in the College Classroom*, Second Edition, Interaction Book Co., Edina, MN.
- [6] Shavelson, R.J. (1988). *Statistic Reasoning for the Behavioral Sciences*, 2nd ed. Allyn and Bacon, Inc., Boston, MA, p. 608.
- [7] Ohland, M.W. and R.A. Layton (2000). Comparing the reliability of two peer evaluation instruments. In proc. *ASEE Annual Conference*. ASEE, St. Louis, MO, June 2000.

## MATTHEW W. OHLAND

Matt Ohland received his Ph.D. in Civil Engineering with a minor in Education from the University of Florida in 1996. He began as an Assistant Professor in Clemson University's General Engineering program in Fall 2000 after serving as the Assistant Director of the NSF-sponsored SUCCEED Engineering Education Coalition and as an NSF postdoctoral fellow. His research is primarily in freshman programs and educational assessment.

## CYNTHIA J. FINELLI

Cindy Finelli is an Associate Professor in the Department of Electrical and Computer Engineering at Kettering University. She received her Ph.D. in Electrical Engineering: Systems from the University of Michigan in 1993. Her technical research includes wavelet analysis for predictive maintenance and biomedical signal processing of electrocardiograms, electroretinograms, and blood pressure signals. She is also active in the field of engineering education and is the first director of the Kettering University Center for Excellence in Teaching and Learning.