An Endogenous Measure of Class Participation

E. Anthon Eff [*]

ABSTRACT

Peer assessments can provide valuable information to instructors who are assigning class participation scores. Student peer assessments bring their own problems, however, including grade inflation and variation in the quality of assessments across students. This paper addresses these problems by employing an endogenous measure of class participation, which dampens the influence of poor quality information and makes the evaluation fairer, since it weighs more heavily the opinions of classmates that students respect. The technique is employed on example data from an undergraduate seminar course, and the results are evaluated. The technique is shown to be incentive-compatible.

Introduction

In smaller classes, teachers are not satisfied if students merely sit passively during lectures. Students are asked to work in groups, to make presentations, to respond to questions, and to participate in discussion. Teachers find that these activities help develop critical thinking and communication skills (Anderson 1998; Panitz 1999), and that they also make the classroom a more enjoyable environment—an important consideration in an age when students often view themselves as customers.

Providing incentives to encourage classroom participation is a challenge. Unless classroom participation is included in the grade, students are likely to act as free riders during group projects, and to sit silently during class discussions (Bean and Peterson 1998). But assigning a grade to a student's contribution to the class learning environment is difficult. The assessment is potentially unfair because the teacher has limited information about any given student's impact on the learning of other students. Students may, in most cases, possess more information than the teacher, at least about the contributions of those with whom they interact most closely. Because of this problem of information asymmetry, students often regard class participation grades as unfair. Teachers are likely to respond to the problem of potential unfairness either by giving all students the same grade, or by basing the grade on seemingly objective measures such as attendance, or the frequency with which a student speaks in class. These responses tend to skew incentives in an undesirable direction. For example, awarding a point for each time a student speaks up would lower the quality of student contributions.

Thus, a class participation grade that provides the proper incentives, and that is considered fair by students, must address the problem of information asymmetry. One way to reduce the problem of information asymmetry is to base the class participation grade on student peer assessment (Speck 1998). Each student typically knows which other students made the contributions that most increased her own learning. A score that considers the perspectives of all students would no longer suffer the problem of the teacher's limited perspective.

Previous research on class participation and peer assessment has focused on the correspondence between teacher assessment and peer assessment (Falchikov and Goldfinch 2000), on the determinants of

[*] Associate Professor, Department of Economics and Finance, Middle Tennessee State University, Box X050, Murfreesboro, TN 37132; phone: 615-898-2387; fax: 615-898-5045; email: eaeff@mtsu.edu. Thanks are due to Joachim Zietz, Reuben Kyle, Edwin Borras, and two anonymous reviewers for useful suggestions.

grades set by peer assessment (Persons 1998), and on the consequences of using peer assessment (Topping 1998). These studies have produced many valuable insights. There is now broad agreement that peer assessment can make an important contribution, not only to the fairness of grading, but to the quality of the learning environment (Falchikov and Goldfinch 2000), particularly in courses dealing with knowledge that is not conclusive (Gopinath 1999). Research has also identified some ways in which peer assessments can be made more accurate: a good peer assessment asks students to rate on a single well-defined general performance dimension rather than on multiple specific dimensions; a good peer assessment avoids using large numbers of peers; and a good peer assessment involves students in defining precisely what is meant by good performance (Falchikov and Goldfinch 2000). No work has been done, however, on developing quantitative techniques to make peer assessment fairer and more accurate.

This paper proposes a method that will enhance the fairness and accuracy of peer assessment. The following section discusses the inherent problems of peer assessment and then describes a technique that will minimize those problems. The proposed technique is then applied to an example data set. The paper concludes by examining the example results to evaluate how well the technique works.

Some Problems with Peer Assessment

Although peer assessment can lower the subjectivity of a class participation grade, it does have its own problems. First, peer assessment can spoil some of the camaraderie of the classroom, as students sense that their classmates are passing judgment on them. Second, students may not take assessment seriously, and may give the highest possible grade to all of their classmates, which would lead to grade inflation. A third problem has to do with the quality of peer assessments: some students, particularly those with little involvement in the class, may not know the names of their classmates and may not have a clear idea of the contributions of other students. Finally, students tend to know well only the contributions of their group mates and friends.

Mitigating student discomfort with peer assessment requires careful management by the teacher. Students should be involved in the determination of the criteria on which they are assessed, so that they understand those criteria and think they are fair (Falchikov and Goldfinch 2000). They should also be informed that organizations increasingly require employees to participate in peer assessments—such as the "360 degree feedback" (Edwards and Ewen 1996) evaluations—and that they need to accustom themselves to accepting the judgment of their peers (Gopinath 1999).

The problem of grade inflation can be addressed by requiring each student to assign a complete, ordinal ranking to every other student (i.e., no ties allowed). This approach might seem unreasonable to some students, since ties would often be appropriate. However, students who should be tied will in fact typically have similar scores when all results are averaged. Students tend to rank themselves too highly, and perhaps particularly weaker students, whose conflict of interest problem is greatest (Gopinath 1999; Falchikov and Boud 1989; Burke 1969). It seems best to eliminate self-ranks altogether.

The problem of variation in the quality of rankings is more serious. Some students will produce high-quality rankings, since they have been deeply involved in the class and are aware of their classmates' contributions. Other students may randomly assign scores to their classmates, since they are unclear about who did what in the class. In particular, the students deserving the lowest class participation scores tend to produce the lowest quality peer evaluations. Resolving this problem requires some method of weighting ranks by the score of the person doing the ranking. It appears reasonable to count a rank from a high-scoring student more than a rank from a low-scoring student, since the former is more likely to assign an accurate rank.

The final problem occurs because a class is never a homogenous group whose members interact equally with each other. A class can be better thought of as a set of smaller cliques, although some students may bridge cliques and others may be isolated. It is therefore understandable that students assign high ranks to their friends, since friends contribute much more to a particular student's learning experience. In addition, students typically have an even more limited perspective than the teacher, and are likely to have good information only about the classmates with whom they interact closely. Resolving this problem requires that the testimony of friends be weighted more than the testimony of those outside a student's clique.

The following section develops a technique that mitigates these problems with peer assessment.

A Peer Assessment Methodology

The three main problems that a peer assessment methodology must solve are: grade inflation, variation in the quality of peer assessments, and variation in the amount of information a student has about each of her peers. The first problem can be resolved by requiring students to rank each peer with no ties allowed. The second and third problems can be resolved by weighting ranks in the following way:

$$c_j = \sum_{i=1}^{n} r_{ij}\alpha_i \tag{1}$$

where the class participation score c_j of the j^{th} student is the sum of the ranks r_{ij} assigned by the $i=1,...,n$ students, each rank weighted by a weight α_i that is unique to the student *assigning* the rank r_{ij}. When student i is a high-ranking student, then the weight α_i should be high, since the testimony of a high-ranking student should be particularly accurate. When student i is particularly liked or respected by student j, then the weight α_i should also be high, since the testimony of a friend is likely to be more informed than the testimony of someone who does not know student j well.

The following equation meets the weighting requirements suggested above:

$$c_j = \frac{1}{\lambda}\sum_{i=1}^{n} r_{ij}r_{ji}c_i \ , \qquad r_{jj} = 0 \ , \ \forall j \ , \tag{2}$$

where λ is some constant, r_{ij} is the rank student i assigns to student j, r_{ji} is the rank student j assigns to student i, and c_i is the class participation score of student i. Self-ranks r_{jj} are set to zero. Note that the rank assigned by a classmate i to student j will be weighted more heavily either when student j esteems classmate i (r_{ji} will be high) or when the class as a whole esteems classmate i (c_i will be high). The specification therefore meets the aims of amplifying both the best quality information and information that comes from classmates in student j's own clique.

This equation can be expressed in matrix form as

$$\lambda c = Ac \tag{3}$$

where **A** is an n x n symmetric matrix of the ranks $r_{ij}r_{ji}$, and **c** is an n x 1 vector of the class participation score. λ is an eigenvalue and **c** is an eigenvector of matrix **A** (Strang 1980: 181). Each element $r_{ij}r_{ji}$ in **A** can be interpreted as the mutual respect between students i and j. Thus, a particular student j with a high score c_j in equation (2) must have established mutual respect with other students i, and particularly with those that have a high score c_i. The class participation score **c** is thus endogenous, in that the score of each student j is a function of the scores of all other students. [1]

A good ranking scheme would provide an incentive for students to guess correctly the final class participation score, that is, their own score would be higher if they assigned high ranks to the true high scorers in the class. From equation (2) it can be seen that a student could collude with another student, each agreeing to give the other a high rank, so that $r_{ij}r_{ji}$ would become quite large. However, if the other student had a low rank c_i, the total contribution to the score c_j would be small. Collusion would only contribute significantly to the score c_j if student i had a high score c_i—in which case student j would be doing exactly what the peer evaluation intends to accomplish: assigning high ranks to the best students. All students therefore have an incentive to assign the correct ranks to other students. [2]

---

[1] Equation (3) has long been used by sociologists and social psychologists, beginning with Seeley (1949) and Katz (1953), to calculate network *centrality* scores. Bonacich (1987) provides an oft-cited explication of the technique. Tam (1989) describes eigenvector-based centrality measures as *endogenous centrality*. Sociologists have used endogenous centrality scores to produce measures of status, prestige, or power (Wasserman and Faust 1994: 205-210). To my knowledge, there exist no previous studies in which endogenous centrality scores have been used to improve the accuracy of peer assessment data.

[2] As a reviewer correctly pointed out, while the testimony of a friend is likely to be more informed, it is also likely to be upward biased. Nevertheless, the upward biased ranking of a friend will only be of much advantage when the friend herself is highly ranked. And one should note that this friend has an incentive to provide an unbiased ranking, or her own score will suffer. In addition, while applying a higher weight to the testimony of friends has the potential to make the measure *less accurate*, it also—in the eyes of the student—makes the measure *fairer*, since the measure weights most heavily the testimony of those the student indicates she respects.

An Application of the Proposed Methodology

This section gives an example of how the class participation score was calculated for an undergraduate seminar course with 21 students. Each class period, groups consisting of two students are responsible for a topic. They hand out written material and make a brief presentation. Students will have unique information about the performance of students with whom they do group work, since preparation for the presentation takes place outside the view of the instructor and other classmates. Fifty percent of the course grade is based on an end-of-the-semester peer evaluation, employed to assess class participation. The peer evaluation requires that each student identify herself, and then rank all 21 students without ties; 1 is the best score and 21 the worst.[3] Table 1 presents the results of the evaluation. Each cell is the rank given by the row student to the column student. Thus, the number '10' in the second row of the first column is the rank given by student 'B' to student 'A.'

TABLE 1. RANK ASSIGNED BY ROW STUDENT TO COLUMN STUDENT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
A	3	5	19	18	12	13	17	9	4	1	14	15	20	21	6	16	8	7	2	11	10
B	10	3	18	19	11	13	5	8	9	4	16	14	17	20	6	21	7	12	1	2	15
C	7	6	1	8	17	18	19	20	5	21	2	9	10	16	14	15	13	3	4	11	12
D	17	8	21	15	11	10	9	2	5	6	14	13	20	19	7	1	3	16	18	12	4
E	4	14	18	19	8	12	11	3	2	13	6	5	20	21	10	7	1	9	15	16	17
F	16	7	10	15	4	1	3	5	9	17	8	7	20	18	6	14	2	13	19	12	11
G	5	7	13	4	11	9	14	2	18	5	19	6	20	8	10	12	1	15	16	17	3
H	17	5	19	11	7	9	3	2	18	6	14	13	21	20	12	10	4	16	8	15	1
I	4	13	19	18	5	17	6	8	3	1	11	15	21	20	2	16	12	9	10	14	7
J	3	6	21	18	17	16	14	9	4	1	5	13	20	19	2	15	8	7	10	11	12
K	18	11	19	13	8	16	5	1	15	10	12	9	21	20	17	4	2	14	7	6	3
L	9	13	12	10	7	14	5	2	19	3	17	18	21	20	4	11	1	15	8	16	6
M	4	5	6	7	8	3	2	9	10	11	12	13	14	1	15	16	17	18	19	20	21
N	14	13	12	11	10	9	8	7	6	5	4	3	2	1	15	16	17	18	19	20	21
O	4	8	21	18	5	10	13	12	3	2	6	14	19	20	1	17	11	7	15	9	16
P	14	4	10	17	7	13	3	2	18	11	15	19	21	20	6	16	1	9	8	12	5
Q	13	8	15	16	14	7	17	2	6	18	12	11	21	19	5	4	1	10	9	3	20
R	8	7	21	18	17	16	11	5	10	3	14	15	19	20	6	4	2	9	13	1	12
S	11	5	20	17	16	14	15	7	13	10	12	9	19	21	2	18	6	3	1	4	8
T	9	3	20	13	12	16	10	21	17	18	11	14	19	15	8	7	4	5	2	1	6
U	7	8	21	13	12	9	3	2	14	15	16	18	20	19	10	11	4	6	5	17	1
Avg	10	8	17	14	11	12	9	7	10	9	11	12	19	18	8	12	6	11	10	11	11

Notes: 1 is the best score, 21 is the worst.

Students 'C' and 'N' assigned high ranks to themselves and to the few other students they knew, but their involvement in the class was minimal, and they ranked most of the class by simply assigning sequential numbers as they filled out the form. Student 'M' did not take the evaluation seriously and assigned sequential numbers to most of the class, including herself. The rows represented by these three students are the poorest quality peer evaluations. One can see by looking at columns 'C', 'M', and 'N' that these three students also received the poorest scores from their peers. These data conform to the general proposition that students with higher scores will have higher quality evaluations.

Two other features of the data are noteworthy. First, nine students assigned themselves as the top student in the class and 13 students placed themselves as one of the top three; clearly, self-evaluations are upwardly biased. Second, two students (F and G) became confused during the evaluation and tied peers, despite the instructions to assign a different number to each peer.

To calculate the class participation scores **c**, the ranks are reordered so that the best ranks have the highest numbers. Self-ranks are discarded by setting the diagonal of the matrix to zero. Since students did not all have the same self-rank, and since some students mistakenly assigned ties, the ranks in each row are renumbered, to make the row sums identical for each row. Next, the matrix is multiplied by its transpose, so that each cell is the product of rank given and rank received. Finally, the absolute value of the eigenvector corresponding to the largest eigenvalue is extracted.[4] The eigenvector is standardized by

[3] Experience with a dozen or so peer evaluations suggests that students more readily comprehend the ranking scheme when asked to assign low numbers to the best performers; perhaps because ranking schemes used in contests employ this ordering (first place, second place, etc.). Before calculating the eigenvectors, however, the ordering must be reversed so that high numbers are assigned to the best performers.

[4] If all elements of **A** have the same sign, then the eigenvector **c** corresponding to the largest eigenvalue will have elements all of the same sign (Mizruchi, Mariolis, Schwartz, and Mintz 1986).

dividing it by its own largest value, so that the scores can be interpreted as a percentage of the class participation score that is received by the highest-scoring student.

TABLE 2. CALCULATED CLASS PARTICIPATION SCORES

Student	[1] c_j (Eq. 2)	[2] c'_j (Eq. 4)	[3] $\frac{(c_j - c'_j)}{c'_j}$ (percent)	[4] θ_i (Eq. 5) (percent)	[5] $\frac{(\theta_i - 1/21)}{1/21}$ (percent)
A	84	77	9	6.0	26
B	92	89	3	7.4	56
C	21	32	-35	0.5	-89
D	44	50	-12	1.7	-64
E	73	72	2	4.8	2
F	62	61	1	3.3	-30
G	74	83	-10	5.2	10
H	99	96	4	9.1	91
I	79	74	6	5.3	11
J	91	81	13	7.3	52
K	57	66	-14	2.9	-39
L	61	64	-5	3.3	-30
M	14	22	-34	0.3	-93
N	18	26	-29	0.5	-90
O	94	87	8	7.7	61
P	64	64	1	3.6	-24
Q	100	100	0	9.2	93
R	80	72	11	5.4	14
S	84	73	15	5.9	25
T	79	66	20	5.3	12
U	75	72	4	5.0	5

Notes: c_j and c_j' are expressed as the percentage of the score of the highest-scoring student. [1] endogenous measure of class participation; [2] naïve approach, where all students are assumed to provide equally accurate peer assessments; [3] percentage change of endogenous measure from naïve measure; [4] the percent of the information used in creating the endogenous measure that comes from each student's peer assessment; [5] the percentage change in information used from each student, comparing endogenous measure with naïve measure.

The first column in Table 2 reports the standardized class participation scores. Student Q has the highest class participation score, and all other students' scores are expressed as the percent of the score attained by Q. Students C, M, and N had the lowest scores, and these students also turned in the poorest quality peer evaluations. At this point, the instructor must use her judgment to curve these scores, creating a final class participation grade.

The following section examines the results to get some insight into how the weighting scheme modifies the raw peer assessment data. In particular, the section attempts to assess the degree to which the weighting scheme singles out the information in the peer assessments provided by the best students, and to assess the strength of the incentive students have to assign the correct rankings to their classmates.

Sensitivity Analysis

Comparison of Endogenous Method with Naïve Approach

One can specify equation (1) so that all $\alpha_i = 1$. This specification is a naïve, egalitarian approach, where one assumes that all peer assessments are equally accurate, and simply sums the ranks received by student *j* without weighting:

$$c'_j = \sum_{i=1}^{n} r_{ij} \tag{4}$$

The second column in Table 2 presents the results of the naïve approach given in equation (4). The third column calculates the percent difference between the naïve approach and the endogenous method presented in equation (2). One can see that scores of the lowest scoring students are reduced substantially by the endogenous method. Top ranking students experience little change, but the scores of students in the middle fall by as much as 14% and rise by as much as 20%—all of this due to the effect of the weighting scheme. Thus, the fairness adjustment occurs on the lower, not higher, end.

Influence of Each Student on Classmates' Scores

Equation (5) presents a measure of the influence each student has in determining the scores of classmates. The numerator in equation (5) is the summed contribution of student i to each of her classmates. The denominator is the summed contribution of all students to each of their classmates. θ_i can be interpreted as student i's percent contribution to the average classmate's score. The fourth column in Table 2 reports θ_i for each student. The ranks assigned by student Q, the highest scoring student, make up 9.2% of the average classmate's score. The ranks assigned by student M, the lowest scoring student, make up only 0.3% of the average classmate's score.

$$\theta_i = \frac{\sum_{j=1}^{n} r_{ij} r_{ji} c_i}{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij} r_{ji} c_i} \tag{5}$$

In the naïve, egalitarian approach of equation (4), each student contributes 4.76% (1/21) of the information to the average classmate's score. The fifth column in Table 2 compares θ_i with 4.76% to see how much the endogenous method of equation (2) amplifies or dampens the contribution of each student. The contribution of Q, the highest ranking student, grows by 93%, while the contribution of M, the lowest ranking student, falls by 93%.

The endogenous method clearly performs as intended: the assessments given by the highest-scoring students are weighted much more heavily than the assessments given by the lowest-scoring students. The method accomplishes this weighting without costly screening time on the part of the instructor. There is no need to examine the scores, determine which are of the highest quality, and adjust weights accordingly. The weighting is done automatically.

The endogenous method would fail to generate superior scores only if the best students turn in the lowest quality peer assessments. In this unlikely case, the method would weight heavily information of poor quality. For two reasons, the best students will almost always provide the highest quality peer assessments. First, to be selected by classmates as the highest ranking, they would have certain characteristics, such as conscientiousness and empathy, that would make them more aware of the contributions of classmates and more likely to take the assessment seriously. Second, the specification of the weights in equation (2) is such that the best students earn the highest score by successfully identifying each other. Were a high-ranking student to fail to identify correctly other high-ranking students, then her own score would fall, and her influence on the scores of other students would diminish.

Strength of Incentive to Assign Accurate Ranks to Classmates

It is possible to assess the strength of the incentive to provide correct ranks by simulating changes in the ranks assigned by a single fairly high-ranking student, say student B. Rewrite equation (2) so that ranks assigned by student B to her classmates are randomized:

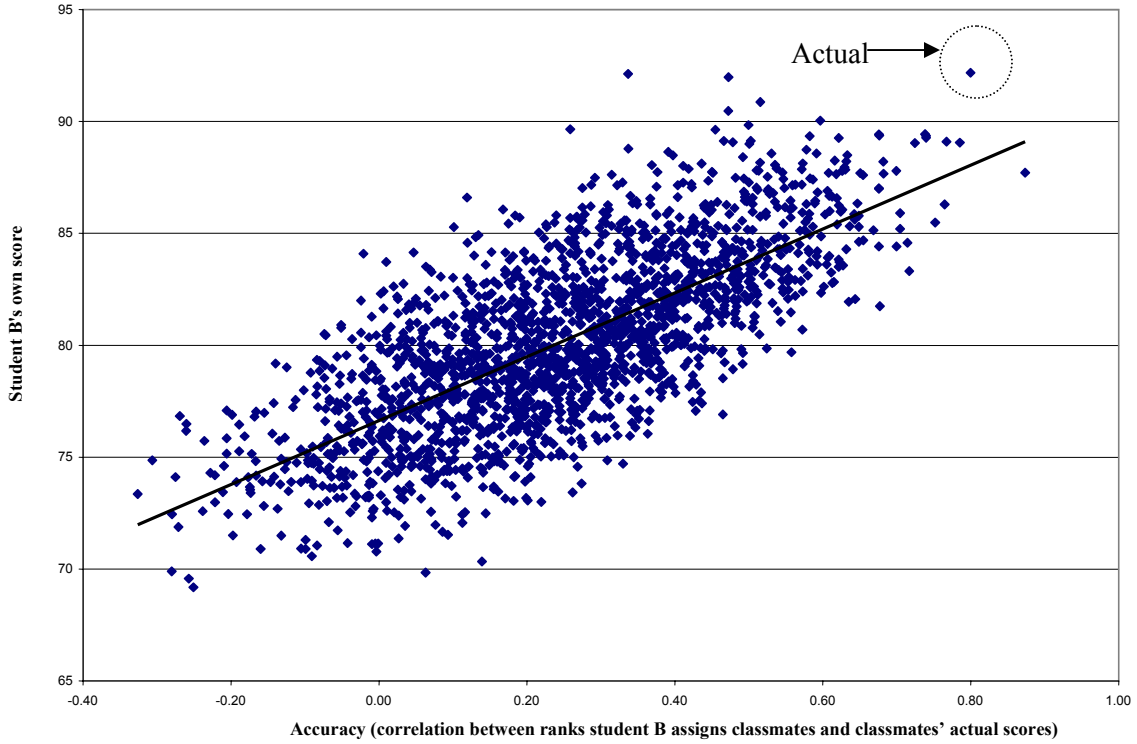
$$\tilde{c}_B = \frac{1}{\lambda}\sum_{i=1}^{n} r_{iB}\tilde{r}_{Bi} c_i \ , \qquad r_{BB} = 0 \tag{6}$$

The data are in all respects the same as before, *except* that the ranks assigned by student B are now random. The ranks that other students assign student B are *not* affected. The accuracy of the ranks student B randomly assigns to her classmates ($\tilde{r}_{Bi}$) can be measured by the correlation between $\tilde{r}_{Bi}$ and c_i. When the assignment is accurate, the correlation will be large and positive.

Figure 1 reports the results of 2,000 simulations. The horizontal axis measures the correlation between $\tilde{r}_{Bi}$ and c_i; positive values indicate that student B correctly assigns ranks to her classmates. The vertical

axis measures the score earned by student *B* ($\widetilde{c}_B$). The pattern of the results from the 2,000 simulations show that when student B correctly assigns ranks to her classmates, her own score increases. When she assigns ranks that are wrong—that is, negatively correlated with the classmates score—her own score decreases.

FIGURE 1: CORRESPONDENCE BETWEEN ACCURACY WITH WHICH STUDENT *B* RANKS CLASSMATES AND STUDENT *B*'S OWN CLASS PARTICIPATION GRADE: THE RESULTS FROM 2,000 SIMULATIONS. (TRENDLINE R^2=0.518)



Comparison of Random Ranks with Actual Ranks

Figure 1 clearly shows that students have an incentive to assign ranks correctly to their classmates. Failure to assign the correct ranks will lower their own grades. But there is another, rather surprising, result shown in Figure 1. The actual performance of student B is shown circled in the upper right hand corner of the graph: her class participation grade is 92.17 and the correlation between the ranks she actually assigned and the grades her classmates actually earned is 0.7997. *Not once*, in 2,000 simulations, did the random rankings manage to improve her class participation grade. Evidently, the ranks she assigned to her classmates served to optimize her own score, at least relative to the 2,000 random vectors.[5]

Table 3 explores further this remarkable result. Two thousand simulations are done on each student, in order to see if any others resemble student B. The last column reports the percentage of the 2,000 simulations that resulted in a class participation score higher than the actual score. For 9 of the 21 students, *none* of their 2,000 random ranking vectors are superior to the rankings they chose for themselves. For six students, at least 10% of their 2,000 random ranking vectors are superior to the ranks they chose for themselves.

[5] The number of possible rankings is 21!= 51,090,942,171,709,400,000. Two thousand simulations is of course an infinitesimal fraction of the possible rankings.

TABLE 3. COMPARISON OF 2,000 SIMULATIONS WITH ACTUAL SCORES

Student *j*	c_j (Eq. 2)				2,000 Simulations	
	[1] Score	**[2]** rank	**[3]** Reciprocity [Corr(r_{ij},r_{ji})]	**[4]** Accuracy [Corr(r_{ji},c_i)]	**[5]** Percent of times $\tilde{c}_j < c_j$	**[6]** Percent of times $\tilde{c}_j > c_j$
A	84	6	0.421	0.866	100	0
B	92	4	0.448	0.800	100	0
C	21	19	-0.137	0.070	8	92
D	44	18	-0.127	0.626	90	10
E	73	13	0.415	0.601	99	1
F	62	15	0.217	0.452	100	1
G	74	12	0.100	0.485	66	34
H	99	2	0.582	0.555	99	1
I	79	10	0.279	0.734	100	0
J	91	5	0.576	0.788	100	0
K	57	17	-0.251	0.544	45	55
L	61	16	0.112	0.700	98	2
M	14	21	0.415	-0.137	67	33
N	18	20	0.362	-0.200	52	48
O	94	3	0.544	0.621	100	0
P	64	14	0.194	0.748	93	7
Q	100	1	0.470	0.584	82	0
R	80	8	0.437	0.842	99	1
S	84	7	0.512	0.827	100	0
T	79	9	0.670	0.498	100	0
U	75	11	0.331	0.747	97	3

Notes: The c_j score is expressed as the percentage of the score of the highest-scoring student; ranks are such that 1 is the best student, 21 the worst. Last two columns may not sum to 100 due to rounding.

The rankings chosen by a classmate can raise her own score in two ways. First, assume she ranks highly those who rank her highly through some form of reciprocity agreement, then the product $r_{ij}r_{ji}$ will be high. Second, consider that her rankings are accurate in the sense that she ranks highly those who the class ranks highly. Then the product $r_{ji}c_i$ will be high. The correlation between r_{ij} and r_{ji}, reported in the third column of Table 3, is a measure of the strength of *reciprocity*. Column four reports the correlation between r_{ji} and c_i, a measure of the *accuracy* with which student *j* assigns her ranks. The six students whose rankings of classmates are most often inferior to random rankings have either low reciprocity or low accuracy, or both. Students C, D, G, and K have low reciprocity and students C, M, and N have low accuracy.

But it is not remarkable that students with very limited involvement in a course—students such as C, M, and N—would not be able to identify the best students, and would not have established ties of mutual respect. What is remarkable is that most students unconsciously evaluated others in exactly the way that would most benefit themselves, despite the fact that they knew nothing about the weighting scheme. This is the essence of the third advantage of the proposed ranking method: it is natural. Students that assign ranks to classmates with good will and good information will be rewarded for their behavior. Thus, not only is there an incentive to assign correct ranks, but students find it quite easy to respond to that incentive. This is an important consideration, for what good are the correct incentives, if the behavior motivated by the incentives is difficult to implement?

Summary and Conclusions

Teachers have incomplete information on student performance, particularly when groups are responsible for presentations or other class projects. A peer evaluation can provide additional information in such circumstances, allowing the instructor to assign a more accurate class participation grade.

Nevertheless, peer evaluations have certain intrinsic problems, and this paper presents a method to mitigate these problems. The method requires each student to assign a unique rank to all classmates. To make the scores more accurate and fairer, the ranks are weighted. High weights are assigned to ranks given by classmates that either the graded student holds in high esteem or the class as a whole holds in high esteem. The procedure works to dampen the influence of poor quality information and it also makes the evaluation fairer, since it weighs more heavily the opinions of classmates that the student indicates she respects.

The procedure was used to evaluate example data from an undergraduate seminar peer assessment, and the results were examined further in a sensitivity analysis. Three conclusions emerged. First, the peer assessments produced by the highest scoring students are weighed much more heavily than the peer assessments produced by the lowest scoring students. Second, the highest scoring students have a strong incentive to assign the correct ranks to their classmates, since failure to assign ranks correctly lowers their own class participation grade. Third, the assigning of correct ranks is very easy if a student acts with good will.

In addition, the results from the example data show that compared to a naïve, unweighted ranking, the proposed endogenous measure does not much affect the scores of the highest-scoring students. Students in the middle, however, are reshuffled—some improving their scores, others falling—depending on the degree to which their best scores come from the best students. Students at the bottom fell further, since the best students uniformly assigned them low scores.

While the methods presented here do help elicit more accurate peer assessments from students, there are nevertheless limitations and unresolved issues. It may occasionally happen that a popular student undeservedly receives a high peer assessment. Though it seems unlikely that group members would reward a conspicuous free rider, no matter how popular, the present study does not attempt to compare peer rankings with "objective" measures of student performance, and therefore cannot judge the extent to which popularity may bias peer assessments. Carefully designed experiments could potentially address this question.

Peer assessments work well only in relatively small class settings, where students know each other by name, and are most valuable in courses with extensive group work outside of class. In cases where students do not interact outside of class, there is less justification for peer assessments—an alert teacher is likely to have fairly complete information regarding student contribution to the learning environment. Students must be continually reminded that their performance will be assessed by peers, and must be reminded to keep track of their peers' contributions, to ensure that the assessment acts as an incentive for improved student participation. Finally, for the assessment to be accurate, students must approach peer assessment with good will. That good will can be earned by involving students in specifying the criteria that should be considered during the ranking, and by showing them that every effort is being made to increase the fairness of the assessment.

Ideally, assessment should provide continuous and helpful feedback to students, so that they know how to reshape their behavior. As courses offer more computer-based components, it becomes easier to require that students enter rank scores for their classmates, and these scores can be calculated without any data entry work on the part of the teacher. Ranking exercises could potentially be conducted every class period or every week, so that students would have continuous feedback and continuous motivation to improve their standing in the class.

References

Anderson, Rebecca S. 1998. "Why Talk about Different Ways to Grade? The Shift from Traditional Assessment to Alternative Assessment." *New Directions for Teaching and Learning*. 74(Summer): 5-16.

Bean, John C. and Dean Peterson. 1998. "Grading Classroom Participation." *New Directions for Teaching and Learning*. 74(Summer): 33-40.

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology*. 92(5): 1170-1182.

Burke, Ronald J. 1969. "Some Preliminary Data on the Use of Self-Evaluations and Peer Ratings in Assigning University Course Grades." *The Journal of Educational Research*. 62: 444-448.

Edwards, Mark Robert, and Ann J. Ewen. 1996. *360° Feedback: The Powerful New Model for Employee Assessment and Performance Improvement.* New York: AMACOM.

Falchikov, Nancy and David Boud. 1989. "Student Self-Assessment in Higher Education: A Meta Analysis." *Review of Educational Research*. 59(4): 395-430.

Falchikov, Nancy and Judy Goldfinch. 2000. "Student Peer Assessment in Higher Education: A Meta Analysis Comparing Peer and Teacher Marks." *Review of Educational Research*. 70(3): 287-322.

Gopinath, C. 1999. "Alternatives to Instructor Assessment of Class Participation." *Journal of Education for Business*. 75(1):10-14.

Katz, L. 1953. "A New Status Index Derived from Sociometric Analysis." *Pyschometrika.* 18: 39-43.

Mizruchi, Mark S., Peter Mariolis, Michael Schwartz, and Beth Mintz. 1986. "Techniques for Disaggregating Centrality Scores in Social Networks." In Tuma, Nancy Brandon (editor) *Sociological Methodology* 1986, pp. 26-48. Washington, D.C.: American Sociological Association.

Panitz, Theodore. 1999. "The Motivational Benefits of Cooperative Learning." *New Directions for Teaching and Learning*. 78(Summer): 59-67.

Persons, Obeua S. 1998. "Factors Influencing Students' Peer Evaluation in Cooperative Learning." *Journal of Education for Business*. 73(4): 225-229.

Seeley, J.R. 1949. "The Net of Reciprocal Influence: A Problem in Treating Sociometric Data." *Canadian Journal of Psychology.* 3: 234-240.

Speck, Bruce W. 1998. "Unveiling Some of the Mystery of Professional Judgment in Classroom Assessment." *New Directions for Teaching and Learning*. 74(Summer): 17-31.

Strang, Gilbert. 1980. *Linear Algebra and Its Applications*. New York: Academic Press, Inc.

Tam, Tony. 1989. "Demarcating the Boundaries between Self and the Social: The Anatomy of Centrality in Social Networks." *Social Networks*. 11: 387-401.

Topping, Keith J. 1998. "Peer Assessment between Students in College and University." *Review of Educational Research*. 68(3): 249-276.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.