Replacing Summative Peer Evaluation with Self-Managed Group: A Lesson in Illusory Superiority

Precha Thavikulwat
Towson University
pthavikulwat@towson.edu

Bosco Wing Tong Yu
The Hong Kong Polytechnic University
bosco.yu@cpce-polyu.edu.hk

ABSTRACT

Summative peer evaluation has deleterious effects that can be avoided by a process we call self-managed group (SMG). SMG requires students to specify their preferred group sizes and choose between personal performance and group performance as their preferred bases for group credit, before being assigned to default groups. Subsequently, students are enabled to re-assign themselves from the default group, subject to instructor-set conditions. SMG does require a measure of personal performance closely linked to group output. Group credit for the group assignment is based on the product of the group output score and weighted personal performance, itself computed as the preference weighted sum of each student's personal performance score and the mean personal performance score of the members of the group. The preferred-basis-for-group-credit choice is strongly affected by illusory superiority. A study of SMG used in conjunction with a computer-assisted, internet-based, one-semester game shows that when the game ended, 94.4% of participants who performed below the mean chose, to their own disadvantage, the personal performance basis for group credit. The disadvantageous effects of illusory superiority, as well as the possibility of high performing members rationally sacrificing personal advantage for group benefits are teaching points that instructors who use SMG with their games and exercises may find productive to explore with students.

Keywords: illusory superiority, peer evaluation, rational self-interest, self-managed groups, self-serving bias, social autonomy principle

INTRODUCTION

The use of student peer evaluations is well established in higher education, so much so that instructors might question why anyone should want to replace them. The usual thinking is that evaluations are informative, and that more information is always better than less information, provided that the information can be gathered without undue expense. The usual thinking presupposes that the expected use of the information does not affect the quality of the information, a presupposition that is not assuredly true when information providers have conflicting interests. In fact, group work is an environment where conflicting interests are the norm. As three undergraduates speaking about group assignments at a conference on teaching complained, "It's inevitable that a member of the team will shirk" (Glenn, 2009).

Scherpereel (2010) defines peer evaluation as "the process of using peer ratings to generate information regarding individual performance pertaining to a member of a work group (team)" (p. 239). He explains that the information can be used for either formative or summative purposes. Formative peer evaluation, conducted before or soon after groups are formed, is used to enhance the teaching of students from whom the information is obtained. In contrast, summative peer evaluation (SPE), conducted at the conclusion of group work, is used to assign credit towards grades for individual contributions to a group assignment. The peer evaluation of which we are concerned is SPE.

Considering that students are people who respond to incentives, SPE can have deleterious anticipatory, concurrent, and consequential effects. The deleterious effects are strengthened by self-serving bias (Miller & Ross, 1975), which is the tendency of people to have inflated views of their own abilities (Regan, Gosselink, Hubsch, & Ulsh, 1975).

Anticipating that they will be peer evaluated at the end of the group assignment, students are incentivized to be inauthentic (Dixon, 2002). They may collude, scheme, and misrepresent, to constitute members of the dominant majority who rate each other positively.

Concurrent with completing the evaluation forms, students are incentivized to assign ratings that benefit those they favor more at the expense of those they favor less, even when those they favor more should have contributed less than those they favor less. Avoiding cognitive dissonance (Festinger, 1957), they rationalize discrepancies.

Consequential to understanding how the evaluations have affected their course grades, students who realize a lower grade than they anticipated from exams and other objectively graded activities are incentivized to complain about the biases, even the bullying, of the peers who rated them poorly. While distinctly unpleasant, this effect may even be called for, considering that the disadvantaged are often victims of bias and bullying, and that the usual anonymity of the SPE process may effectively protect perpetuators of discriminatory acts.

Yet, neither the alternative of assigning all members of a group the same credit score for the accomplishments of the group nor that of assigning no credit towards grades for group accomplishments is satisfactory. The same-credit solution incentivizes free riding, and the no-credit solution incentivizes sloth. What is needed is a solution that incentivizes collective effort and individual initiative.

SELF-MANAGED GROUP

Consider the principle that simulations and experiential exercises should be designed to enable students to manage social problems by themselves, rather than to rely on an instructor to manage the problems for them, which we shall refer to here as the social autonomy principle, itself a special case of the learner empowerment principle that calls for giving students more choices about their learning environment (Saye, 1997). The principle is based on two considerations. First, when instructors manage problems that students can manage by themselves, instructors expend efforts for naught. Second, when students manage social problems by themselves, they gain experience from which they may develop useful skills. Peer evaluations, whether formative or summative, violate the principle, because in both cases the evaluations are submitted to the instructor for action.

In accordance with the social autonomy principle, our solution replaces SPE with self-managed group (SMG), an arrangement supported by rules executed by computer software built into a computer-assisted game or assignment-management system. The solution requires students to specify their preferred group sizes and preferred bases for group credit. Students' preferred group sizes are constrained to fall within the range set by the instructor. Their preferred basis for group credit is limited to a binary choice between personal performance and group performance.

In the briefing preceding the requirement to choose, students are told that they will be computer-assigned to groups based on their two specified preferences. The computer assigns students to groups as they register, first-come-first-served, so those who register later may find themselves placed in a group of a size lower than the size they prefer. Nonetheless, students can use the self-assignment feature of the software to re-assign themselves after registration should they be displeased with their initial assignments, subject to limitations on group size set by the instructor. The effect of the instruction is to incentivize every student to learn the meaning of group credit, to register early, and to become adept at using the self-assignment feature, in order to secure for themselves an optimal group assignment.

GROUP CREDIT

The crux of our solution is in the computation of group credit. We define group credit as credit towards grades that students receive from the measurable outcomes of group work, consisting of personal performance credit and group performance credit. Personal performance credit (p) is credit for an outcome assigned to individual members of the group. Group performance credit (g) is credit for the collective outcome. For simplicity, we assume a single outcome for each of both credits. Inasmuch as multiple outcomes can be summed to a single value, our single-outcome analysis generalizes to multiple outcomes.

For an equation of the relationships, let p be the mean $\bar{p}$ across all members of a group and S be the standard of performance that can be set relative to either p or g. The standard can be the instructor-set standard of the outcome, or the points allotted for the outcome, or the mean or highest score of that outcome. Also let n_p be the number of group members specifying preference for personal performance, and n_g be the number of group members specifying preference for group performance. Then

$$\text{Group Credit} = \left(\frac{p n_p + \bar{p} n_g}{n_p + n_g}\right)\frac{g}{S}. \tag{1}$$

Equation 1 can written more cleanly by defining the personal performance weight (ω), such that

$$\omega = \frac{n_p}{n_p + n_g}. \tag{2}$$

Thus, Equation 1 also can be written as

$$\text{Group Credit} = \frac{[p\omega + \bar{p}(1-\omega)]g}{S}. \tag{3}$$

Equation 3 can be understood in two ways depending upon the reference basis of S. If S is with reference to p, then Equation 3 means that each member's group credit is a score, g , that depends on that member's weighted p relative to S. On the other hand, if S is with reference to g, then each member's group credit is a score of weighted p that depend on the group's g relative to S.

S should be set with reference to whichever score, g or p, has higher weight towards grades. For example, if g is the score on a 40-point group assignment whereas p is the score on a 10-point individual assignment, then setting S with reference to p delimits each group member's score on the 40-point group assignment. The message is that p modifies g. Conversely, if g is the score on a 10-point group assignment whereas p is the score on a 40-point individual assignment, then setting S with reference to g delimits the students' 40-point individual assignment. Now, the message is that g modifies p.

At the limit when either p or g has zero weight, the indeterminate ratio of the zero-weight item relative to S is resolved by setting the ratio to 1. Thus, should no measurable outcome be assignable to individual members, then p, $\bar{p}$, and S will be zeros, $[p\omega + \bar{p}(1-\omega)]$ / S is set to 1, and Group Credit = g for every member of the group. The measurable outcome, p, can be frequency of attendance at group meetings, instructor-assigned points for individual papers submitted by members, judges' ratings of individual presentations by group members, or scores from a game administered such that every member receives an individual score, among others. The essential condition of p, however, is that efforts expended to raise p also tends to raise g.

For the case of social groups that support members without themselves delivering a measurable group product, such as study groups, g = S = 0 and g/S is set to 1, so Equation 3 reduces to Equation 4.

$$\text{Group Credit} = p\omega + \bar{p}(1-\omega). \tag{4}$$

EXAMPLE

For a numerical example, suppose an instructor assigns a group project to a group of four students. If the instructor sets aside 10 points for the group project, the instructor might assign the project output, which could be a group paper, a group presentation, or the outcome of a team game or exercise, the score of 8 points. If the instructor uses SPE to assign group credit for the project, the instructor might convert the SPE score of each student into an SPE ratio (r) to derive a group-credit result following Equation 5, and as shown in Table 1 for the four students.

$$\text{Group Credit} = rg \tag{5}$$

TABLE 1
Example Using Student Peer Evaluation

Student	SPE Ratio	g	Group Credit
A	.9	8	7.20
B	.8	8	6.40
C	.7	8	5.60
D	.6	8	4.80

Suppose the instructor decides to replace SPE with SMG and to use attendance at group meetings as the measure of p. In this case, the group met six times. The instructor decides to set the standard of performance to the number of group meetings, so S = 6. One student attended all six meetings, two attended five, and one attended four, as shown in Table 2. SMG group credit is derived using Equation 3. To simplify Equation 3, let the weighted personal performance ($\hat{p}$) be defined as follows:

$$\hat{p} = p\omega + \bar{p}(1-\omega). \tag{6}$$

TABLE 2
Example Using Self-Managed Groups

Student	p	$\bar{p}$	$\hat{p}$			g	Group Credit (S = 6)		
			ω = 1	ω = 0	ω = .5		ω = 1	ω = 0	ω = .5
A	6	5	6.0	5.0	5.5	8	8.00	6.67	7.33
B	5	5	5.0	5.0	5.0	8	6.67	6.67	6.67
C	5	5	5.0	5.0	5.0	8	6.67	6.67	6.67
D	4	5	4.0	5.0	4.5	8	5.33	6.67	6.00

Then Equation 3 can be re-written as follows:

$$\text{Group Credit} = \frac{\hat{p}g}{S}. \tag{7}$$

The resulting group credit of the four students is shown in Table 2 for ω = 1, ω = 0, and ω = .5.

SWITCHING GROUPS

Our solution allows any student to exit a group at will, and to enter at will any group whose size is below the minimum of the instructor-set range of group sizes. Within the instructor-set range, students can join a group only if three conditions are met ex post:

1. The group's size does not exceed the maximum of the instructor-set range
2. The group's size does not exceed the smallest preferred group size specified by its members
3. The group's p does not fall

The first condition respects an instructor-set constraint. The second condition assures unanimous consent of group members to increasing the group's size above the minimum of the instructor-set range. The third condition assures that the entry of the new member will not have the net effect of collectively reducing the group credit of ex ante members. The three rules allow students to switch groups autonomously without causing problems that would require instructor involvement.

PROBLEMS

The issues that accompany SMG are psychological and technical. We address the psychological issue here, setting aside the technical one for another time.

Psychologically, SMG place the onus on students to understand the group-credit formula and group-switching mechanism, but students are not always attentive. Inattentiveness will induce students to choose the most available option, rather that the most rational one. Rationally, about half of every population will be below average in performance, in the usual case when the distribution of performance is approximately normal.

Psychologically, however, self-serving bias includes illusory superiority, also known as the better-than-average effect and the Lake Wobegon effect, the last on account of Garrison Keillor, a radio personality known for his spoof of a fictional town where the children are all above average (Keillor, 1989). Illusory superiority is a robust tendency, the conclusion of a meta-analysis by Zell, Strickhouser, Sedikides, and Alicke (2020).

Accordingly, we expect that that most students will choose personal performance rather than group performance as their preferred basis for group credit at registration. But inasmuch as the personal-performance choice disadvantages those whose actual performance scores are below average, the extent to which below-average performers will maintain their disadvantageous choice as the exercise proceeds towards its conclusion remains an open question. To the extent the illusory superiority of below-average performers overwhelms evident facts, to that extent the scores of SMG will diverge in favor of above-average performers.

IMPLEMENTATION

We implemented SMG in a computer-assisted, internet-based (Pillutla, 2003) business game, GEO, in which each participant receives an individual score. Participants collaborate and compete in the game, as investors of companies, as employers and employees, and as consumers bidding for products produced by the companies of the game.

Companies sell shares to investors. The investor controlling most of the shares outstanding is enabled to employ the company's manager, who in turn is enabled to employ other executives. Investors, managers, and other company executives are all roles played by participants of the game, with each participant able to play multiple roles, in the same way that a single person in everyday life can be an investor and a manager of a company at the same time, or an investor of one company and a manager of another company at the same time.

The groups of the game function as social groups. Although every participant is required to be a member of a group, the groups themselves are not necessarily the teams that produce products, because companies of the game can be managed by members of the same group, or of different groups.

Participants are scored based on their virtual consumption of the virtual products produced by the companies of the game. Consumption is enabled by government payments, dividends, interest, employment, and gains from trading shares. Those with higher income and capital gains generally consume more and therefore receive higher scores, but the relationship of income and capital gains to consumption is moderated by the participants' engagement and astuteness in bidding competitively for the products produced by the companies of the game.

A participant's game score is the sum of the participant's personal performance credit and group credit. The groups are social group, so Equation 8, which modifies Equation 3, applies.

$$\text{Game Score} = p + \text{Group Credit} = p + [p\omega + \bar{p}(1-\omega)] = p(1+\omega) + \bar{p}(1-\omega). \tag{8}$$

EXPECTATIONS

Although participants can switch groups entirely by themselves, whether they do so early in the game or not depends upon the class activities that the instructor expects of the groups and on the participants' understanding of how to avail themselves of the game's group-switching feature. If the instructor requires each group to deliver a class presentation on its scheduled day, for example, switching groups will be more difficult, because some students may not be available to present on some days and because some days, usually the later days, may be preferred by many students over other days. After the game has substantially commenced, however, group-switching should be rare, because personal relationships will have developed among group members that are socially costly to break and reset when a member switches from one group to another. Formally,

H1: Few or no participant switches groups past the first quartile of the game.

At the start of the game, all groups are purely social groups, rather than teams brought together to produce a collective product. As such, we expect that self-serving bias will dominate over collective concerns to cause the great majority of participants to specify personal performance, rather than group performance, as their preferred basis for group credit. Formally,

H2: At the start of the game, few to no participant specifies group performance as their preferred basis for group credit.

As the game proceeds, rational self-interest dictates that participants whose personal performance scores are below the mean personal performance score of their groups should specify group performance as their preferred basis for group credit, thereby lowering the group's ω from about 1.0 to about .5 and thus raising their own game scores, in accord with Equation 3. Even so, we expect that some below-mean participants will retain their specified personal performance choices until the end of the exercise, because inattentiveness and illusory superiority dominates over rational self-interest. Formally,

H3: At the end of the game, the number of participants specifying group performance as their preferred basis for group credit is fewer than the number whose personal performance score is below the mean of their groups.

TABLE 3
Number of Participants by Preferred Group Sizes

Quartile	Preferred Group Size	
	3	4
0	2	31
1	1	32
2	2	31
3	2	31
4	1	32

METHOD

We report here on a study in which the game was administered to 33 college undergraduates, 57.6% female, enrolled in a one-semester global economic environment course at an Asian university. Participants choices of preferred group sizes were limited to three and four. The game advanced through 120 periods. Offers for products, shares, and employment were processed continuously. Production, however, was batched for each period; salaries, interest, and taxes were paid each period; and financial statements were made available for each period.

RESULTS

Dividing the 120 periods of the game into 30-period quartiles, the number of participants by preferred group sizes from the start of the game (quartile 0) to the end (quartile 4) is shown in Table 3. Across all quartiles, 93.4% to 97.0% of participants preferred size 4 over size 3.

The number of participants who switched groups peaked in quartile 1, at 72.7%, as shown in Table 4. Only 10.0% of participants switched groups in quartile 2 and none did in later quartiles, confirming expectation (H1: Few or no participant switches groups past the first quartile).

TABLE 4
Number of Participants by Group Switching Action

Quartile	Switch	No Switch
0	1	32
1	24[a]	9[a]
2	3[a]	30[a]
3	0	33
4	0	33

[a] $\chi^2(1) = 25.1$, $p = .000$.

The number of groups by actual group sizes is shown in Table 5. Largely consistent with preferences, 77.8% of all groups are of size 4 across all quartiles. Although the distribution of group sizes is unchanged between quartile 0 and quartile 1, group membership changed substantially due to participants switching groups in quartile 1, as shown in Table 4.

TABLE 5
Number of Groups by Actual Group Sizes

Quartile	Actual Group Size			
	1	2	3	4
0	0	1	1	7
1	0	1	1	7
2	2	0	1	7
3	2	0	1	7
4	2	0	1	7

The complexity of group switching is notable. From quartile 1 to quartile 2, two members of the single three-member group exited the group to form a new three-member group with one member of the single two-member group, leaving the remaining one member of each original group to be the sole member of the two original groups. The two participants left in single-member groups ended the game with the lowest and fifth-lowest personal performance scores. Neither of these two participants could join the three-member group because one member of that group had specified a preferred size of 3, effectively blocking the expansion of the group.

The number of participants by preferred basis for group credit and personal performance scores relative to group means are shown in Table 6. No participant specified group performance as their preferred basis for group credit in quartile 0, confirming expectation (H2: At the start of the game, few to no participant specifies group performance as their preferred basis for group credit).

By the last quartile, only 6.1% of the participants specified group performance as their preferred basis for group credit even though 54.5% of them would have raised their own game score had they done so, confirming expectation (H3: At the end of the

TABLE 6
Number of Participants by Preferred Basis for Group Credit and Personal Performance Scores Relative to Group Means

Quartile	Preferred Basis for Group Credit		Personal Performance Scores Relative to Group Means			Statistics[a]	
	Group Performance[a]	Personal Performance[a]	Below[a]	Mean	Above[a]	$\chi^2(1)$	*p*
0	0	33					
1	0	33	19	0	14	23.9	.000
2	3	30	17	2	14	13.5	.000
3	2	31	14	2	17	11.0	.001
4	2	31	18	2	13	17.8	.000

[a]$\chi^2(1)$ based on group and personal performance choices versus below and above scores.

game, the number of participants specifying group performance as their preferred basis for group credit is fewer than the number whose personal performance score is below the mean of their groups). Remarkably, of the 18 participants who ended the game scoring below their group means, 17 ended the game choosing, to their disadvantage, personal performance as their preferred basis for group credit, a rational-error choice by 94.4% of the below-average population.

CONCLUSION

The results confirm all expectations. Few participants switched groups past the first quartile of the game. No participant specified group performance as their preferred basis for group credit at the start of the game. Only 6.1% of participants specified group performance as their preferred basis for group credit by the end of the game, even though about half would have raised their own game score by doing so. About 94.4% of participants who ended the game with personal performance below their group means disadvantaged themselves by choosing personal performance as their preferred basis for group credit, attesting to the remarkable strength of illusory superiority.

SMG resolves the difficulties of SPE by replacing a conflict-promoting judgmental process with a student-managed process. The SMG procedures described enable students to select their preferred group sizes within the range set by the instructor, to specify their preferred basis for group credit, and to switch groups at any time, subject to instructor-set limitations.

Moreover, SMG enable instructors to draw attention to illusory superiority. Forewarned, students may be forearmed, but whether forewarning suffices is an empirical question that might be clarified by further studies.

If SMG is applied in a setting wherein groups are charged with producing a collective product, so that Equation 3 applies, then specifying group performance as their preferred basis for group credit by those whose personal performances exceed their group means is an act of sacrifice that can nevertheless be rational. The apparent sacrifice of the higher performers who choose group performance over personal performance may raise the motivation of the lower performing members of the group, giving rise to an increase in group performance that may be enough to override the loss of the sacrifice.

Thus, SMG also enables instructors to explore with their students the conditions when rational self-interest favors sacrificing personal advantage for group benefits. Students may gain more from studying these conditions than from the nominal subject matter of the discipline for which the game or experiential exercise is designed. After all, the true purpose of higher education is not primary to teach students about a subject matter, but rather to teach students how to think and act rationally using the tools of a subject matter. The principle that simulations and experiential exercises should be designed to enable students to manage social problems by themselves embodies this purpose.

REFERENCES

Dixon, R. J. (2002). Toward greater authenticity: A case for divergent simulations. *Simulation & Gaming, 33*, 360-366. DOI: 10.1177/104687810203300310.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.

GEO. [Developed by P. Thavikulwat] Towson, MD (601 Worcester Road, Towson, MD, USA). Available from http://wp.towson.edu/precha/

Glenn, D. 2009, June 8. Students give group assignments a failing grade. *The Chronicle of Higher Education*. Available http://chronicle.com/article/Students-Give-Group/47303/

Keillor, G. (1989). Faith: Stories from the collection more news from Lake Wobegon [CD sound recording]. St. Paul, MN: High Bridge.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*, 213-225.

Pillutla, S. (2003). Creating a Web-based simulation gaming exercise using PERL and JavaScript. *Simulation & Gaming, 34*, 112-130.

Regan, J. W., Gosselink, H., Hubsch, J., & Ulsh, E. (1975). Do people have inflated views of their own ability? *Journal of Personality and Social Psychology*, 31, 295-301.

Saye, J. (1997). Technology and educational empowerment: Students' perspectives. *Educational Technology Research and Development, 45*, 5-25.

Scherpereel, C. M. (2010). Peer to peer: A new tool for student peer evaluation. *Developments in Business Simulation & Experiential Learning, 37*, 238-250. Available from the *Bernie Keys Library*, https://absel.org.

Zell, E., Strickhouser, J. E., Sedikides, C., Alicke, M. D. (2020). The better-than-average effect in comparative self-evaluation: A comprehensive review and meta-analysis. *Psychological Bulletin, 146*, 118-149. Available http://dx.doi.org/10.1037/bul0000218