All Together Now? College Students' Preferred Project Group Grading Procedures

Jody R. Hoffman and Steven G. Rogelberg
Bowling Green State University

Three hundred sixty undergraduate students were randomly assigned to evaluate 1 of 12 versions of a hypothetical college syllabus from a course in which student project groups were used. As hypothesized, students reported the highest course-enrollment intentions and most positive perceptions of the group grading procedure when both individual and group performance on the group project was evaluated. In addition, students with low grade point averages (GPAs) exhibited stronger course-enrollment intentions than students with high GPAs when group work accounted for a high percentage (50%) of students' course grades and when group performance was evaluated. Contrary to hypotheses, students did not prefer courses in which they had input into the grading of their group work. Implications for the use of group grading procedures are discussed, in addition to future research directions.

Increasingly, instructors choose to introduce some form of group work in their classrooms (e.g., Johnson & Johnson, 1989), often reporting a number of benefits resulting from such work, including decreased student absenteeism, increased student preparation and effort (Dinan, 1995), improvements in student learning and achievement (Johnson & Johnson, 1989), improved interpersonal relationships, and increases in student retention and self-esteem (Johnson & Johnson, 1987). With the increased use of group work, however, comes the question of how (if at all) such work should be graded. Although some instructors report a philosophical opposition to grading group work, others have attempted to design group work grading procedures that have positive effects on student learning and motivation, similar to the effects individual grading can have (e.g., Karau & Williams, 1993; Zeidner, 1992). The present study examines group work grading procedures from the college student’s perspective, attempting to determine which grading procedures are most positively perceived and produce the strongest course-enrollment intentions.

Previous Research Examining Students' Reactions to Group Grading Procedures

Reports of college students’ reactions to group grading systems have focused largely on survey methodology (generally of students in course with group) and anecdotal accounts. In field studies, for example, Morahan-Martin (1996) and Basu and Middendorf (1995) found the use of other groups’ and instructor’s assessments of group performance to be perceived as fair by students. Individual accountability is also thought to be preferred by students. For example, Bykerk-Kauffman (1996) reported anecdotally that students are relieved when told that grading systems will include some means of holding individuals responsible for their work in groups. In a survey of business students at a single college, Alexander and Stone (1997) found the assignment of the same grade to all students who had worked on a group project to be perceived as neither fair nor unfair. Varco-Shea, Darlington, and Turnbull (1996) reported anecdotally that students were dissatisfied with a requirement that no student in a group could earn the same peer evaluation as another student.

A classic experimental field study of college student group work and evaluation preferences was performed by Deutsch (e.g., 1947). This study stands out from other work in that controlled comparisons could be made between two different forms of group evaluation systems.
Specifically, Deutsch randomly assigned college students to cooperative and competitive group work situations within the context of an actual course. In the cooperative groups, grades were assigned to the group as a whole; group performance (e.g., number of puzzles solved or quality and quantity of ideas when discussing and developing recommendations for solving a human relations problem) was the sole basis for evaluation. The group with the highest performance at the end of the semester was excused from a term paper. In the competitive groups, individual contributions to the group’s performance were ranked, apparently in an exclusive manner (i.e., no two students within the same group could receive the same rank). The individual with the highest performance within each group was excused from a term paper at the end of the semester. Interestingly, the results of a survey administered at the end of the semester indicated that the majority of students in each condition declared a preference for the grading system they had participated in, suggesting the importance of prior experience.

In sum, although existing evidence suggests that college students exhibit preferences for group grading procedures, to our knowledge, previous work has not systematically assessed college students’ reactions to a variety of specific group grading practices within the context of a single study. In the light of the potential importance of group grading procedures and the small number of studies examining these procedures, the present study examines factors that may affect students’ perceptions of group grading systems and their intentions to enroll in a course using such systems.

Grading-System Attributes

Our review of the extant literature on group grading procedures identified two major attributes of group-grading systems, which we refer to as evaluation target and evaluator (see Lejk, Wyvill, & Farrow, 1996, for a partial review of grading procedures). Evaluation target refers to the aspect of work in groups that is evaluated to determine grades (generally the performance of individual group members; the performance of the group; or both, the last being by far the most common in practice, based on our review). Evaluator refers to the individual(s) (i.e., instructors vs. instructors with input from group members) responsible for determining group grades.

To develop hypotheses for students’ preferences with regard to these attributes, we drew on Karau and Williams’s (1993) collective effort model (CEM). Building on an expectancy theory framework, this model was primarily developed to explain individual motivation, including motivational deficits (e.g., social loafing), within group contexts. Karau and Williams suggested that individual effort (and resulting performance) on group tasks is a function of the extent to which such performance is likely to result in valued outcomes (i.e., maximize expected utility—in this case, grades for group work).

With regard to the evaluation-target attribute, this model suggests that students’ expected utility will be greatest when individual performance within groups is targeted in the grading of group work. More specifically, Karau and Williams (1993) have argued that the relationship between individual performance and individual outcomes (i.e., course grades) is more direct than that between group performance and individual outcomes (i.e., course grades), because students have more control over their own behavior and performance than those of a group. Thus, to the extent that students’ individual performance is incorporated into the group project, expected utility will be greater.

Hypothesis 1. Group-grading systems that incorporate individual-performance evaluations will produce more positive evaluation-system perceptions and

---

1 We are thankful to the reviewers of this article for this suggestion.
higher course-enrollment intentions than systems that do not incorporate individual-performance evaluations (i.e., those systems in which group performance is the sole basis of grades).

With regard to the evaluator attribute, we expected that students would prefer group grading procedures that incorporate their input into instructors’ evaluations. Karau and Williams’s (1993) CEM suggests that student input into instructors’ assignment of grades will increase their expectancy that valued outcomes (i.e., higher grades) will be obtained, as opposed to courses in which they have no input into group work grades. Indeed, research in other domains (e.g., performance appraisal) has found that the incorporation of appraisees’ evaluations (or “voice”) in assessments of performance is preferred by those being evaluated (Greenberg, 1986; Taylor, Tracy, Renard, Harrison, & Carroll, 1995), partially because such systems are perceived by employees to result in more accurate evaluations (Lind & Tyler, 1988).

Hypothesis 2. Grading procedures that incorporate both students’ and instructors’ evaluations of group work will be preferred by students, as opposed to grading procedures that use only instructors to evaluate group work. Specifically, instructors who incorporate students’ evaluations of their group’s (or group members’) performance will produce more positive perceptions of the grading system and higher course-enrollment intentions in a course using such a system.

Grade Point Average

Group researchers have found a strong relationship between individual achievement and preferences for individual or group based work and rewards. High-achieving group members are less attracted to, and more likely to leave, groups when incentives are group based, whereas the opposite holds for low-achieving members (e.g., Cable & Judge, 1994). In the context of Karau and Williams’s (1993) CEM, low-achieving students would be expected to prefer group-based grading systems because they stand to achieve the best outcomes (i.e., grades) in such a system. On the other hand, high-achieving group members would prefer individually based systems, because such systems would maximize outcomes for such members. Some previous research supports this hypothesis. For example, high-achieving group members may feel inequity when paired with less able group members (e.g., DeMatteo, Eby, & Sundstrom, 1998).

Hypothesis 3. Students with high GPAs will prefer a course in which group project grades are based on their individual performance in the group. Students with lower GPAs, on the other hand, will prefer a course in which group project grades are based on the group’s performance.

Grading-System Context

One contextual factor of potential importance in a group grading situation is the percentage of course grades that is based on group work. For example, a situation in which 50% of course grades are based on group work (as opposed to a more minimal percentage, such as 20%) may amplify reactions to group grading systems. More specifically, Karau and Williams’s (1993) CEM suggests that a higher percentage will increase the valence of the outcome (i.e., grade), increasing the motivational force associated with each grading procedure. For example, low-achieving students’ expected negative reactions to individually based grading procedures will be amplified when 50%, as opposed to 20%, of their course grades are dependent on such a system.

Hypothesis 4. Significant interactions between the percentage of course grades that is based on group work and (a) evaluation target, (b) evaluator, and (c) GPA will be found.

Method

Participants

Three hundred sixty students from a medium-sized midwestern university participated in this study. Participants were recruited from introductory psychology courses at this university and completed the study during class time. Participation was voluntary, and experimental credit was provided to those who took part (no students refused to participate). The majority of participants were female (70.3%) first-year undergraduate students (70.7%). The mean age of participants was 19.63 (SD = 3.00). Most (96.7%) reported that their current introductory psychology course was the only psychology course they had ever taken. Ninety-eight percent had worked in a group as part of a college course, whereas 98% had worked in a group at some time during high school.
**Design**

We used a $2 \times 2 \times 3$ between-subjects design. The levels of the first factor, the percentage of students' course grade that was based on group work, were 20% (i.e., a minority) and 50% (i.e., a majority). These levels were chosen to ensure maximal impact of this variable within the bounds of realistic values. The levels of the second factor, evaluator(s), were instructors only and instructors with input from group members. The final factor, evaluation target, consisted of three levels: (a) individual based (i.e., individual performance within the group was the sole basis of evaluation); (b) group based (i.e., group performance was the sole basis for evaluation); and (c) hybrid (i.e., with components of both individual- and group-based methods). All factors were fully crossed, resulting in 12 between-subjects experimental conditions.

The independent variables were manipulated through instructions embedded in a hypothetical course syllabus for an introductory clinical psychology course. In addition to three exams, the hypothetical course required the completion of a group research project. The latter was described as:

> a research study on a topic in clinical psychology of your choice. Your group will plan, conduct, and analyze this study. One class every 2 weeks will be devoted to working on these projects. At the end of the semester, your group will hand in a single paper summarizing your research project and its findings.

The syllabus also contained information about the percentage of the student's course grade that would be based on this group project (20% or 50%), in addition to the percentage based on the three individual exams, and described the target of evaluation. Note that the evaluator variable (“instructor” or “instructor, taking into consideration group members’ anonymous evaluations”) was embedded in this description. The individually based condition was presented as follows:

> Your individual group project grade will be based on the [evaluator’s] overall evaluation of the effort you contributed to the project. The instructor will base her evaluation on her observations of your performance during the class sessions in which groups work on their projects (i.e., every 2 weeks). The overall quality of the final group paper will not be considered when assigning grades.

The group based condition was described as follows:

> Your group project grade will be based on the [evaluator’s] assessment of the overall quality of your group’s project. More specifically, the [evaluator] will grade the paper submitted by your group at the end of the semester, and this overall grade of project quality will then be assigned to each group member, regardless of the individual effort each contributed to the project.

The hybrid condition was described as follows:

> Your group project grade will be based on the [evaluator’s] assessment of both your individual effort within the group and the overall quality of the project. The instructor will base her evaluation of your individual effort on her observations of your performance during the class sessions in which groups work on their projects (i.e., every 2 weeks). The overall quality of the group project will be determined by the [evaluator’s] assessment of the paper turned in by the group at the end of the semester. Thus, you may or may not receive the same grade as your group members—each member of your group will receive the same grade if they contribute the same amount of effort, as determined by the [evaluator].

**Measures**

**Intention to enroll.** Participants were presented with three items that assessed their intentions to enroll in the course (“I would register for this course,” “I am interested in taking this course,” and “I would take this course only if it was a graduation requirement”). The third item was reverse coded. These items were modified from a scale developed by Turban and Keon (1993) designed to assess job seekers’ intentions to apply for a job with a particular company. Participants responded to these using a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree).

**Perception of group grading system.** We assessed participants’ perceptions of the group grading system with three items (“I like the way group project grades are assigned,” “The way grades are assigned for the group project is very appropriate,” and “The way grades for the group project are assigned is fair”). A 5-point scale accompanied these items (as described above).

**Demographic and background information.** Respondents’ age, gender, GPA (“What is your college GPA for all courses that you have completed to date?”), field of study, and year of study were assessed at the end of the study through self-report, in addition to their experi-
ence with group work and group grading practices in a university setting.

**Manipulation Check**

Three items assessed the effectiveness of the experimental manipulation. The first referred to the evaluator manipulation ("The instructor solicited group members’ input before assigning group project grades," with response options yes and no). The second item assessed the effectiveness of the percentage manipulation ("The group project described in the syllabus accounts for half of each student’s final course grade," with response options yes and no). The final item assessed the effectiveness of the evaluation-target manipulation ("What was evaluated to determine group project grades?" with response options “individual performance,” “project quality,” or “both”). The majority of participants (97% of the sample) correctly answered all of the manipulation check items. Individuals who incorrectly answered one or more manipulation check items were not removed from data analysis; note that subsequent exclusion of these participants did not affect (i.e., change) the results of data analyses.

**Procedure**

First, students were told that they would be participating in a study of their perceptions of, and reactions to, a particular method used to determine individual students’ grades in a college course. Participants were instructed to answer all questions carefully and honestly and were reminded of the anonymity of their responses. Students were told to assume that they were registering in courses for the upcoming fall semester. One of their options was the introductory clinical psychology course. The instructor had given them the course syllabus to help them decide if they would register for the course.

Participants were then randomly assigned to experimental conditions. Participants were presented with a packet containing (a) the syllabus describing the particular experimental condition to which they had been randomly assigned and (b) the measures (dependent measures, manipulation checks, and demographic and background information, in that order). Participants were asked to read through the syllabus carefully and answer the questions that followed on the next page.

**Results**

**Descriptive Statistics and Correlations**

Sample sizes, means, standard deviations, and correlations among the scales and demographic–background information variables are presented in Table 1. Note that these descriptive statistics were computed across experimental conditions. Means and standard deviations within each experimental condition for each dependent variable are presented in Table 2.

**Control Variables**

To determine the necessity of the inclusion of control variables (e.g., gender or age) in analyses, the distribution of each individual-difference variable across the 12 experimental conditions was examined. No significant differences in the distribution of age, gender, or other demographic variables were found across experimental conditions. Additionally, exploratory analyses indicated that these variables did not interact with the independent variables in this study, nor did they relate to the dependent vari-

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Descriptive Statistics and Correlations Among Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>n</td>
</tr>
<tr>
<td>----------</td>
<td>----</td>
</tr>
<tr>
<td>1. Intent to enroll</td>
<td>360</td>
</tr>
<tr>
<td>2. Perceptions of group evaluation system</td>
<td>360</td>
</tr>
<tr>
<td>3. Grade point average</td>
<td>355</td>
</tr>
<tr>
<td>4. Age</td>
<td>359</td>
</tr>
<tr>
<td>5. Gender</td>
<td>360</td>
</tr>
</tbody>
</table>

Note. Where applicable, values of coefficient alpha are presented along the diagonal. * p < .05. ** p < .01.
Table 2
Means and Standard Deviations for the Dependent Variables by Experimental Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Intention to enroll</th>
<th>Grading system perceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Evaluation target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybrid</td>
<td>3.29</td>
<td>1.05</td>
</tr>
<tr>
<td>Individual</td>
<td>2.94</td>
<td>1.03</td>
</tr>
<tr>
<td>Group</td>
<td>2.52</td>
<td>.97</td>
</tr>
<tr>
<td>Percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>2.42</td>
<td>1.00</td>
</tr>
<tr>
<td>50%</td>
<td>3.30</td>
<td>1.09</td>
</tr>
<tr>
<td>Evaluator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructor only</td>
<td>2.94</td>
<td>1.08</td>
</tr>
<tr>
<td>Instructor and group</td>
<td>2.80</td>
<td>1.18</td>
</tr>
<tr>
<td>Grade point average</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3.07</td>
<td>1.07</td>
</tr>
<tr>
<td>High</td>
<td>2.75</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Variables. Thus, these variables were not included as control variables in analyses.

Analyses and Discussion

Table 3 shows the results of a multivariate analysis of variance testing the effects of the four independent variables on intentions to enroll and grading-system perceptions. Note that GPA was treated as a dichotomous variable (high vs. low) by performing a median split. As indicated by Table 3, a significant two-way interaction occurred between the GPA and percentage variables. No other interactions were significant. Inspection of univariates shows that the GPA × Percentage interaction was significant only for intentions to enroll. Tukey's honestly significant difference (HSD) tests indicated that students with low GPAs had higher enrollment intentions for the 50% (M = 3.55, SD = 1.01) as opposed to 20% (M = 2.38, SD = 0.80) condition. Students with high GPAs, on the other hand, showed no significant preference for the 50% (M = 2.47, SD = 1.20) as compared with 20% (M = 3.02, SD = 1.11) conditions. This interaction was congruent with our expectations that lower achieving students would anticipate larger benefits from working in groups and be more likely (than high-achieving students) to want to enroll in a course using groups.

Table 3 also shows that the percentage and evaluation-target variables were significant. Inspection of univariates indicates that these effects were significant for both dependent variables. Thus, students had more positive perceptions of the group grading system and stronger enrollment intentions when 50%, as opposed to 20%, of their course grades were based on group work (see Table 2 for means and standard deviations for each of the independent variables). Tukey’s HSD tests indicated that all mean comparisons were significant for the conditions within the evaluator variable. Thus, students had the most positive perceptions and strongest enrollment intentions with regard to the hybrid grading system, followed by the individually targeted procedure. The least preferred system was the group-targeted one.

The preference of students for grading procedures that target individual performance to
some extent, as opposed to those that completely ignore individual performance, is consistent with Karau and Williams's (1993) CEM, which suggests that such procedures will result in higher (predictions of) expected utility. Incorporation of individual-performance assessments also decreases the impact that social loafing has on other students' grades, as such loafing is a common complaint made by students who work in groups (Lejk et al., 1996). The stronger preferences for hybrid procedures, as opposed to procedures that solely evaluate individual performance (i.e., effort), may have occurred as a result of students' expectations. That is, students quite likely never encounter a course in which a project in itself is not graded. Purposeful effort on a project that is not evaluated may be viewed as wasteful effort by students, decreasing the expected value of taking such a course and completing a group project.

As indicated by Table 3, the test for multivariate effects of the evaluator variable was not significant, which was contrary to our hypothesis that students would prefer group grading systems in which they provided some input into their grades. The lack of a significant preference for such a procedure may be explained by the fact that the instructor in this study evaluated students' group work frequently (i.e., every 2 weeks). Thus, the beneficial effects (e.g., voice in grades, increased expected utility) of students serving as evaluators may have been negated. It is plausible that the present study provided a scenario in which students believed instructors would view their performance frequently enough to provide accurate evaluations (again, at least with regard to the individual-effort and hybrid evaluation conditions). The accuracy of evaluations may also explain why an effect was not found for the project-quality condition. That is, given that the evaluation of a group’s product is an objective endeavor, students may feel confident that the instructor will provide their group with a fair grade when evaluating their group project’s quality. Thus, the preference for input into grades may be irrelevant in such a situation, given that it is not unreasonable for students to assume that an instructor will grade their performance objectively and accurately (through their group project’s quality).

### Future Research and Limitations

Future research should examine the effects of other individual differences on a variety of students' reactions to group grading systems. For example, collectivism and individualism may be related to students' reactions to group grading systems (Chen, Meindl, & Hui, 1998), so that students with strong individualistic tendencies may prefer individually targeted procedures, whereas students with strong collectivist tendencies may prefer group-targeted procedures. The effect of group grading systems on other dependent variables (e.g., actual expenditure of effort or satisfaction in a course) would be another interesting avenue for future study. This work would build on that of Deutsch (1947), who examined a variety of dependent variables (including grading-system preferences, satisfaction, group cohesion, and performance). It would be interesting, for example, to examine whether students' preferences for group grading systems align with the system(s) they are exposed to (as Deutsch found). Additionally, future work should examine different types of hybrid grading procedures, given that current work in other fields, such as organizational reward practices, examines different versions of complex reward systems (e.g., differing amounts of individual and group incentives; Lawler, 1990).

This study has certain limitations. The data collected were solely self-report. The collection of additional data, such as actual enrollment figures for courses using group grading procedures, would provide stronger evidence of the effects of group grading systems. However, behavioral intentions are a useful predictor of future behavior (e.g., Ajzen & Fishbein, 1977), so the data collected in this study do have value. A more important concern arises from the possibility that we may have confounded two variables within the evaluation-target conditions. That is, evaluations can be group or individually based and also assess effort (i.e., during the semester) or an end product (i.e., a written report). The two conditions we chose for this study were those that we found to be much more common in the literature (and anecdotally from our own experience); however, a more complete theoretical understanding of the effects of group evaluation systems might manipulate these as separate factors (i.e., is a quality of work vs.
quantity of effort effect driving students’ preferences, or is an equality vs. equity distinction driving the effect?).

Summary and Implications

In summary, the present study found that the evaluation targets of a group grading system and, to a lesser extent, the percentage of grades based on group work were related to students’ reactions to group grading procedures. As the use of groups continues to heighten in classrooms, with instructors turning to grading of group work to determine some portion of students’ course grades, research such as the present study is increasingly important.

Students’ a priori desire for some form of individually based group grading procedure is clear. Interestingly, students with low GPAs also preferred a higher portion of their course grades being dependent on group work. Educators who use group grading procedures in their classrooms should find these results potentially useful. For example, some form of individually based grading system may result in higher satisfaction levels in a course, as well as attraction of high-performing students.

References


