The effects of group quizzes on performance and motivation to learn: Two experiments in cooperative learning

B. Douglas Clinton a,*, James M. Kohlmeyer III b,1

a Department of Accountancy, College of Business, BH 345R, Northern Illinois University, DeKalb, IL 60115-2854, United States
b College of Business, East Carolina University, East Fifth Street, Greenville, NC 27858-4353, United States

Abstract

This study investigated the effect of group quizzes on accounting students’ performance and motivation to learn. While cooperative learning in accounting education has been studied in recent years, the effects on student performance have been mixed [Lancaster, K.A.S., Strand, C.A., 1999. Using the team-learning model in a managerial accounting class: an experiment in cooperative learning. Issues in Accounting Education 16(4), 549–568; Ravenscroft, S., Buckless, F., Hassall, T., 1999. Cooperative learning – A literature guide. Accounting Education 8(2) (1999) 163–176]. Thus, two experiments were conducted (one using an experimental design and the other using a quasi-experimental design) that examined student performance and motivation to learn. The first experiment used a quasi-experimental design to compare the performance and motivation to learn of students who took a series of group quizzes versus students in comparable classes in a prior semester that did not take group quizzes. Using a series of group quizzes in a mixed factorial design, the second experiment examined the performance of both: (1) long-term groups versus ad hoc groups and (2) self-selected groups versus instructor assigned groups. Findings revealed no performance differences across conditions in either the first or second experiments. However, student subjects in the first experiment
(using group quizzes) reported a significantly greater motivation to learn and perception of learning than those in the second experiment (not using group quizzes).

Keywords: Cooperative learning; Performance; Small groups; Motivation

1. Introduction

Hite (1996) noted that the response to the Accounting Education Change Commission’s (AECC) call for change has mostly been directed toward curriculum reform in content rather than methods of teaching or delivery modes. Although content remains a dominant theme in accounting education literature, recent recommendations have also included careful consideration of pedagogy and, specifically, delivery method (e.g., Albrecht & Sack, 2000).

Besides complying with these recommendations by focusing on pedagogy, this study is motivated by more educators using group quizzes and/or group exams as a form of cooperative learning (Hite, 1996). Accordingly, this study examines the effect of group quizzes on accounting students’ performance and motivation to learn. Using experimental and quasi-experimental designs supported by cooperative learning theory, we present findings showing that, although no performance effects were noted (i.e., either in favor or against cooperative learning), students evidence a greater motivation to learn. Results also document student perceptions indicating that they believed that a greater amount of learning had taken place when cooperative learning techniques were employed. These findings were consistent across all experimental conditions and time frames. These conditions included group assignment method (i.e., whether group members self-selected or were assigned by the instructor to their groups) and time frame (i.e., whether group members remained in the same group for the entire semester or were assigned to a different group for each quiz). As with several other studies in collective decision making (Ciccotello, D’Amico, & Grant, 1997; Hite, 1996; Ravenscroft, Buckless, McCombs, & Zuckerman, 1995), our results, on average, evidenced consistently higher scores on quizzes taken as a group than individually.

The current study contributes to the accounting education literature in several ways. First, the effects of assignment method and time frame are examined for the first time in a cooperative learning accounting education context. Second, the current study provides evidence that students exhibit a greater motivation to learn in the cooperative learning context as compared to a non-cooperative learning context. Finally, as also evidenced by Hite (1996), this study provides evidence that, despite concerns about the negative effects on students’ evaluation of the instructor (Cottell & Millis, 1994; Michaelsen, 1992; Michaelsen, Watson, & Shrader, 1985), cooperative learning may favorably increase students’ ratings of an instructor.

The next section of this article presents a literature review and includes formal hypotheses and theoretical support for hypothesized relationships. The research
method used in the study is then described followed by the results section. Finally, conclusions, limitations, and a discussion of implications are presented.

2. Theory and hypotheses

2.1. Introduction to cooperative learning

Generally, cooperative learning (CL) has been defined as pedagogies that involve the use of groups whose members share interdependent goals and are assessed on individual outcomes (Ravenscroft, Buckless, & Zuckerman, 1997). Cooperative learning is not to be confused with traditional group work (Hite, 1996), which does not necessarily depend on interdependence and individual accountability. Two potential disadvantages of traditional group work, “hitchhikers” and “workhorses,” are discouraged by two essential characteristics of cooperative learning, positive interdependence and individual accountability (Cottell & Millis, 1994). To foster positive interdependence, an instructor must structure assignments so that each group member is dependent on the other members to complete a task successfully. To achieve individual accountability each member must be held accountable for mastering the material (Peek, Winking, & Peek, 1995).

The interactive discussion in the group allows for individuals to express ideas in their own words. By explaining and defending their ideas and listening to those of others in the group, individuals are afforded an opportunity to understand the perspectives of other group members, restructure their own ideas, and reconcile conflict to reach consensus (King, 1992). Other beneficial outcomes include serving in the role of both teacher and student to other group members and being forced to take ownership over ideas, thereby encouraging the retention of information.

Extensive theory, research, and practice support the use of CL in all levels of education. Johnson, Johnson, and Stanne (2000) reviewed the CL literature and found over 900 research studies that, in general, validated the effectiveness of cooperative over competitive and individualistic efforts. Within accounting education, the results are not as clear. Research specific to accounting indicates that CL has the potential to increase student satisfaction. Furthermore, the learning of technical accounting material is at least as good under various implementations of CL as under traditional methods (Ravenscroft et al., 1997; Lancaster & Strand, 2001). However, most accounting professors are likely not trained in cooperative learning. Moreover, given the results of studies that have examined CL in accounting, instructors may believe that the potential incremental gains do not justify implementing CL methods.

2.2. Performance

Ravenscroft, Buckless, and Hassall (1999) and Lancaster and Strand (2001) provide a comprehensive literature review of cooperative learning research in accounting. Ravenscroft et al. (1995) found that students in an experimental treatment group, graded on both individual and team effort, performed substantially better
Hite (1996) investigated the effectiveness of group midterm exam “re-takes” following individual exams. Students in a junior-level tax class who participated in group exams scored higher on the final exam as compared to students who took only individual exams. Ravenscroft et al. (1997) report results of seven studies, varying team learning and group grade incentives. They found little or no improvement on exam scores when students worked in teams or when graded using group incentives. Ciccotello et al. (1997) investigated the effects of students preparing for examinations using team problem-solving workshops. They found that students using such workshops improved examination scores as compared to those students who worked individually. Lancaster and Strand (2001) found that cooperative learning did not improve student performance or reveal any significant student attitudes toward this alternative learning format. Thus, the evidence to date is mixed regarding the performance benefits of cooperative learning as compared to alternative methods.

As noted by Hite (1996), the design of her study suffers from several potential weaknesses. First, performance may have resulted from additional time spent discussing exam content for the group condition that did not take place in the control condition. Second, the increased time taking exams in general has been shown to increase retention (Nungester & Duchastel, 1982). Moreover, the fact that each group contained a student with a high GPA may have increased performance for all students if the better students dominated the group discussion. Fourth, the quasi-experimental design may show differences that cannot be verified with an equivalent control group. These issues have all been either removed or mitigated in the current study.

While accounting studies have reported mixed results regarding student performance from CL, the weight of the research in non-accounting areas appears to support an increase in student performance. Therefore, we hypothesize that individual performance will be better for students that have taken quizzes as a group than for those who merely worked individually. Consequently, the following hypothesis is presented (alternative form):

H1. Performance on the final exam will be higher for students that took group quizzes than for those that did not take group quizzes.

2.3. Student attitudes

A few studies have examined the effect of cooperative learning on student attitudes, producing mixed results. Caldwell, Weishar, and Glezen (1996) examined attitudes toward accounting for students enrolled in either financial or managerial accounting. Their analysis found that students in the cooperative learning activity for the financial principles course were more likely to maintain a positive perception

---

2 Although a quasi-experimental design was also used in the first experiment of our study, the second experiment made comparisons of results between students in the same classes.
of accounting than students in the lecture format. Students in the managerial principle classes exhibited no significant difference in positive or negative perception of accounting across the treatments. Lindquist and Abraham (1996) found that cooperative learning activities improved student attitudes and perceived achievement in a senior-level financial reporting course. Hite (1996), while analyzing student evaluations of teaching, reported that group exam participants had more positive attitudes toward the course and the teacher than students from a prior class who took only individual exams. Lancaster and Strand (2001) found that cooperative learning did not generate any significant student attitudes toward the cooperative learning format. Lastly, Alavi (1994) reported findings that indicated that GDSS-supported (group decision support system) collaborative learning leads to higher levels of perceived skill development, self-reported learning, and evaluation of classroom experience in comparison with non-GDSS-supported collaborative learning.

Our study also compares results of student evaluations of teaching for the two groups. Based on favorable student attitudes in past empirical studies and theoretical support (Alavi, 1994; Michaelsen, 1992; Sherman, 1986), we tested the following hypothesis (alternative form):

H2. Students taking group quizzes will provide higher student evaluation of teaching scores than students not taking group quizzes.

2.4. Assignment method

In designing group work, instructors must decide whether to assign students to specific groups or whether to allow students to self-select into groups. This study examines assignment at two levels – self-selection or random assignment. The vast majority of cooperative learning studies have had the instructor assign students to groups based on criteria such as GPA, exam scores, major, and learning styles (Hite, 1996; Lancaster & Strand, 2001; Ravenscroft et al., 1995). We know of no accounting education study that has examined the differences in performance of self-selected groups as compared to instructor-assigned groups. Williams (1981) provides evidence that self-selected groups will outperform random assignment groups because of the developed cohesion in the self-selected groups. A cohesive group of self-chosen friends provides an incentive to exert high effort as compared to low cohesion in randomly selected group members. Thus, we test the following hypothesis (alternative from):

H3. On average, group quiz performance will be greater for groups that self-select their members than for individuals that are assigned to their groups by the instructor.

2.5. Time frame

Another issue common to classroom management involving groups is whether students should work with the same group members for the duration of the semester (i.e., long-term groups) or whether they should be reassigned to different groups.
given independent tasks (i.e., ad hoc or short-term groups). Few studies have compared the performance of long-standing groups and ad hoc groups. The majority of studies utilizing ad hoc groups have found that groups do not outperform the group’s best member (Watson, Michaelsen, & Sharp, 1991). However, Michaelsen, Watson, and Black (1989) and Watson et al. (1991) obtained results for the performance of long-standing groups indicating that best members rarely outperformed the group and became less important to the group’s success as the group members gained experience working together. Thus, these studies provide evidence that long-standing groups may outperform short-term groups (i.e., ad hoc groups).

Shepperd (1993) suggested that internal incentives for achieving good collective performance exist when individuals identify with, or feel a sense of pride in or duty toward, their group. These qualities facilitate group cohesiveness and can even evoke high-effort contributions from group members. Unfortunately, these qualities often develop slowly over time (if at all) and thus are unlikely to occur among strangers in an ad hoc group. Other studies have provided evidence that ad hoc groups’ lack of social attraction and collective responsibility among members can lead to shirking or social loafing (e.g., Geen, 1991; George, 1995; Shepperd & Taylor, 1999). That is, some group members will perform below their true capabilities in an effort to conserve energy and ‘hide’ within the group. Thus, we expect that performance for long-standing groups will improve over time as group members become more familiar with each other and are able to form more accurate expectations for behavior. Hence, we would expect that an ad hoc or short-term group of this nature would, on average, display lower performance than long-standing groups due to a lack of social attraction and collective responsibility, which arise from fairly low intrinsic feelings of personal accountability toward the other group members. Therefore, the following hypothesis is tested (alternative form):

H4. On average, group quiz performance will be better for long-term groups than for short-term groups.

3. Method

3.1. Subjects

The participants for the first experiment were 146 undergraduate accounting students enrolled in four sections of cost accounting at a large state university located in the midwestern United States. For the second experiment, 76 students in two sections of cost accounting participated in an in-class laboratory experiment as partial fulfillment of a class requirement. Average age of the subjects was 23, and 42 percent of the subjects were male. The classes were composed of juniors and seniors.

3 The terms *ad hoc* and *short-term* are intended to be used interchangeably. Both terms have been used in the literature to describe groups that are short-term in nature.
3.2. Experimental design – experiment 1

The quasi-experimental design structure for the first experiment is presented in Fig. 1. In our study, we use group quizzes as the cooperative incentive structure to promote student-to-student cooperation. Individual quizzes and exams are used to promote individual accountability. The primary method for hypothesis testing is to compare performance and student evaluation responses for one semester’s students versus the other semester’s students. The spring participants are the treatment condition involving the taking of quizzes as groups, while the prior semester participants did not take quizzes as a group. The dependent variable is the students’ scores on the final examination. The student responses used by Hite (1996) are: “(1) ‘Overall, I would rate this instructor as outstanding,’ (2) ‘This instructor was fair and impartial in dealing with students,’ and (3) ‘The subject matter of this course has been mentally challenging.’” This study examined additional evaluation items, including: (1) “The instructor helped me to learn the course content,” (2) “The instructor was able to motivate me to learn the subject matter,” (3) “The instructor increased my enthusiasm for the subject matter,” and (4) “The instructor helped me to improve my problem solving ability.”

The experimenters attempted to ensure that the control and treatment groups (i.e., between the two semesters) were as equivalent as possible. For example, the final exams were identical, the instructional materials were identical, delivery of the instructional materials was identical (i.e., PPT lecture one day, assignment coverage the next), delivery of the instructional materials was by the same instructor, the time of day for the two classes was the same for both semesters, and the same rooms and technology were used. Demographic data (see Table 1) was quite similar for age and GPA. Gender make-up was different between the terms, but testing revealed no significant gender differences that would be expected to influence study results.

3.3. Experimental design – experiment 2

The 76 students used as the treatment group (Spring) for the (quasi-) experiment 1 were the same students used for all conditions of Experiment 2. Students were given instructions concerning the quizzes and our study (Appendix A). Ten quizzes were administered during the semester. Each quiz contained five questions, covered one PowerPoint lecture presentation, and was given immediately following the lecture (i.e., same day). Quizzes were given to the students to complete individually and then as a group. After the instructor collected each quiz completed individually, the students were given a clean copy of the same quiz to complete in a small group (i.e., composed of 3 or 4 students). During this time, the students were told to openly discuss what they believed to be the best answers for the questions on the quiz in an effort to reach a consensus on quiz answers. Then they turned in the quiz as a group, putting all group members’ names on the quiz.

The grade recorded for each individual was equal to the average of the individual and group quiz scores. This was designed to encourage each student to maximize their performance both individually and as a contributing group member. Quiz
Fig. 1. Quasi-experimental design model – experiment 1 (administration of the treatments for the control group (Cell B – fall semester) occurred prior to the experimental group (Cell A – spring semester)).


Table 1
Subject characteristics and questionnaire data

<table>
<thead>
<tr>
<th>Number of respondents</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>76 (52%)</td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>70 (48%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of participants (percentage) by gender</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>21</td>
<td>41</td>
</tr>
<tr>
<td>Female</td>
<td>55</td>
<td>29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GPA (not self-reported)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.20</td>
<td>3.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Initial and semester-end questionnaires (5-point scale, 1 = highest rating; 5 = lowest)*</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairness of the group process</td>
<td>1.71</td>
<td>2.07</td>
</tr>
<tr>
<td>Fairness of the overall grade</td>
<td>1.92</td>
<td>2.21</td>
</tr>
<tr>
<td>Satisfaction with the group arrangement</td>
<td>1.93</td>
<td>2.11</td>
</tr>
<tr>
<td>Satisfaction with the overall grade</td>
<td>2.03</td>
<td>2.51</td>
</tr>
<tr>
<td>Familiarity with other class members</td>
<td>1.18</td>
<td>1.63</td>
</tr>
<tr>
<td>Expectation of a learning benefit</td>
<td>2.08</td>
<td>2.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post-quiz questionnaire (average of all 10 quizzes)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree of preparation for quiz with group members</td>
<td>4.66</td>
<td></td>
</tr>
<tr>
<td>Degree of individual preparation for quiz</td>
<td>3.85</td>
<td></td>
</tr>
<tr>
<td>(5-point scale, 1 = highest rating; 5 = lowest)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree to which the quiz was difficult</td>
<td>3.53</td>
<td></td>
</tr>
<tr>
<td>Degree to which the quiz was easy</td>
<td>3.94</td>
<td></td>
</tr>
<tr>
<td>(5-point scale, 1 = lowest rating; 5 = highest)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Although all means of the data presented is reflected of raw data, the wording of the item Familiarity
grades were worth 10% of the course grade (i.e., each quiz was worth one percent of the course grade).

An Initial Questionnaire (Appendix B) was given at the start of the semester to gather demographic data and perceptions regarding the fairness of the group process and to provide information to establish a benchmark on the familiarity of students with each other. The benchmark of familiarity was used to speculate regarding the salience of the prospect of self-selection. That is, if the subjects were unfamiliar with their classmates, they might not have really cared whether or not they were afforded the luxury of choosing their own group members. This instrument was also used to assess satisfaction expectations with the group process and to determine whether the subjects perceived that they would be treated fairly or whether they perceived a learning benefit from being involved in the group process.

A Semester-End Questionnaire (also reflected in Appendix B) was also administered to all participants. This questionnaire was similar to the Initial Questionnaire, except that it did not include the demographic questions asked at the beginning of the semester. This was useful to determine any changes between perceptions from the beginning of the term and group project versus perceptions at the end of the term.

A Post-Quiz Questionnaire (completed individually – Appendix C) was administered following every quiz (i.e., ten times). This questionnaire asked how each group member contributed to the group quiz results, how much they studied or otherwise prepared for each quiz, etc. The objective of this instrument was to gather information regarding the equivalence of the quizzes in general and potential covariates (i.e., degree of quiz difficulty or degree of student preparation) that might affect results.

The experimental design structure for the second experiment is presented in Fig. 2. This experiment examined the effects on performance of two factors – time frame and assignment. About half of the students were randomly chosen and assigned by the instructor to a group. The rest of the subjects were allowed to self-select their own group members. For the self-select condition, we took no part in determining how student groups were formed other than to tell the students that they needed to form groups of three. Generally, this allows the students to use any of a variety of ad hoc mechanisms to form groups presumably based on desired or actual friendships, who they believe might be hard working, intelligent, contributing members, or other criteria.

Within each of the two spring sections of the course, all subjects were members of a group (i.e., either self-selected or randomly assigned) that was maintained for the entire semester. The purpose of this was to determine if there is a difference in performance between the self-selecting groups and the assigned groups. This grouping was maintained for half of the 10 quizzes (i.e., all odd numbered quizzes – 1, 3, 5, 7, and 9). These five quizzes were administered over the entire semester (i.e., these groups were together for 16 weeks and not just for half of the semester), which provided the measurement of the long-term time-frame condition. This testing of long-term groups versus ad hoc groups is not dependent on the number of times a group meets but rather on the duration of time the members were in the same group.
The ad hoc groups met one time while the long-term groups met five times over a 16 week period.

The instructor also explained that on the other half of the quizzes (i.e., all even numbered quizzes – 2, 4, 6, 8, and 10), all individuals would be in a group requiring new and different members. All individuals, regardless of whether they had self-selected their group or had been randomly assigned by the instructor on odd-numbered quizzes also participated in these ad hoc groups. The objective of this was to determine the relative impact on performance of self-selection and assignment when compared to temporary group arrangements. The purpose of using a counterbalanced assignment was to achieve a prolonged time frame for evaluation (16 weeks was the longest time frame possible in a semester class venue), and to minimize possible interaction effects due to demand issues or timing of specific semester events (e.g., increasing performance by learning how to do better over time or poor performance at the end of the term due to being busier).

From a pedagogical perspective, the instructor explained to the subjects that there were tradeoffs to both assignment approaches. For example, it was explained that self-selection allows group members to choose to work with people that they believe...
will best serve the purposes of the group. However, assignment by the instructor allows a potential benefit of experiencing the growth and learning inherent in working with unfamiliar people and adjusting for mutual benefit.

3.4. Response variable definitions – both experiments

On the first quasi-experiment, the scores achieved on the final examination were used to measure performance. The affective variables were defined with respect to particular items as worded on the student evaluation of teaching. Group quiz scores were used to measure Performance in the second experiment.

3.5. Covariates – both experiments

The covariates used for both experiments were Grade Point Average (GPA) and Gender. The GPA covariate relates to the possibility of student subjects performing better or worse due to general trends for academic performance. Given the differences in gender makeup between the fall and spring sections, examination of gender as a covariate was considered especially important for the first experiment. GPA was objectively obtained from University records.

4. Results

Subject demographic characteristics and descriptive means for the specific questionnaire responses are presented in Table 1. Questionnaires were administered to the spring sections only (i.e., those individuals participating in the group quizzes – experiment 1). Responses to the Initial and Semester-End Questionnaires indicated that, on average, subjects agreed that the group process and grade outcomes were fair. Subjects also indicated that they were generally satisfied with the group arrangement and overall grade and had a reasonably high expectation that they would/did receive a learning benefit for their participation. In evaluating the salience of self-selection, subjects indicated a low degree of familiarity with other student colleagues in their section. This provides a substantial limitation on the potential effectiveness of the self-selection manipulation for the first experiment. Responses to all items on the two questionnaires tended to decrease slightly from the initial administration to the semester-end. Although most differences were insignificant, this decrease may be reflective of the disappointment of their earlier expectations regarding the group process in terms of the fairness, satisfaction, and achievement of a learning benefit.

The Post-Quiz Questionnaire responses are also presented in Table 1. These questionnaires were administered after each of the 10 quizzes, so the results presented in the table are the cumulative averages of all of the quizzes. Scores for individual group quizzes did not appear to indicate a longitudinal trend over the 10 quizzes for any of the variables examined. The results appear to indicate that, on average,
students mostly did not study individually for the quizzes. Moreover, they studied even less with group members outside of class. The subjects also indicated the quizzes, on average, were neither too difficult nor too easy. Thus, they appear task salient as a basis for experimental results.

4.1. Hypothesis tests – first experiment

Hypothesis 1 was examined by using the quasi-experimental approach similar to Hite (1996). The expectation offered by H1 was that performance on the final exam would be higher for students in the spring semester who were required to take group quizzes than for those students in the fall semester who were not required to take group quizzes.

For Hypothesis 1, the dependent variable of final exam score (FINAL) was tested using a univariate ANCOVA with GPA and Gender as covariates. The independent variable was the group condition (GRPCOND) depicting a two-level factor for the fall sections that did not take the group quizzes (FINAL mean = 80.60; standard deviation = 14.03) and the spring sections that did take the group quizzes (FINAL mean = 80.49; standard deviation = 9.93). The univariate tests of the factor GRPCOND showed a non-significant effect ($F = 0.240$, $p < 0.625$). The non-significant $F$ statistic does not support the experiment’s performance hypothesis (H1) and, given the power of the test to find a potential difference, indicates that there are likely no differences between the measures of performance on the final exam for the classes. Specifically, the power of the test (0.99) indicates only a 0.01 beta risk of not finding a moderate effect if it exists, assuming alpha = 0.05. These results are presented in Table 2.

Hypothesis 2 relates to students’ evaluation of teaching. The only statistically significant difference in means ($t = 2.522; p = 0.015$) was noted for the item, “Overall

Table 2

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of freedom</th>
<th>Mean square</th>
<th>$F$</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANCOVA tests of between-subjects effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>3</td>
<td>496.417</td>
<td>3.613</td>
<td>0.015</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>5551.809</td>
<td>40.411</td>
<td>0.000</td>
</tr>
<tr>
<td>GPAa</td>
<td>1</td>
<td>1465.281</td>
<td>10.666</td>
<td>0.001</td>
</tr>
<tr>
<td>GRPCONDb</td>
<td>1</td>
<td>32.933</td>
<td>0.240</td>
<td>0.625c</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>8.872</td>
<td>0.065</td>
<td>0.800</td>
</tr>
<tr>
<td>Error</td>
<td>143</td>
<td>137.382</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>147</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a GPA = Student GPA from administrative records.

b GPACOND = Two-level factor for the fall sections that did not take the group quizzes versus the spring sections that did take the group quizzes.

c Power $= 1 - \beta$ for this test is greater than 0.99 for $\alpha = 0.05$ and a medium effect size $= 0.25$. 
the instructor’s performance in teaching this class was... (excellent, good, satisfactory, poor, or unacceptable).” Other evaluation items not included in the Hite study that were of interest also resulted in significant differences. The means for these four items are shown in Table 3. Thus, we consider our second hypothesis to be successfully supported.

4.2. Hypothesis tests – second experiment

The third hypothesis predicted that, on average, group quiz performance would be greater for groups that self-selected their members than for individuals that were assigned to their groups by the instructor. As shown by the experimental design in Fig. 2, H3 was tested by a comparison of means of group quiz performance in cells A (self-selected) and B (randomly assigned). Since subjects in these two cells maintained the same long-term group status for the entire semester, the assignment

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Experiment 1: Test of Hypothesis 2 comparison of teaching evaluation scores for group vs. non-group conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student evaluation of teaching item</td>
<td>Fall non-group means</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>1. The instructor challenged me to achieve a high level of performance</td>
<td>3.91</td>
</tr>
<tr>
<td>2. The instructor was fair and consistent in applying course policies</td>
<td>4.37</td>
</tr>
<tr>
<td>3. Overall the instructor’s performance in teaching this class was:</td>
<td>4.20</td>
</tr>
<tr>
<td>4. The instructor helped me to learn the course content</td>
<td>3.80</td>
</tr>
<tr>
<td>5. The instructor was able to motivate me to learn the subject matter</td>
<td>3.50</td>
</tr>
<tr>
<td>6. The instructor increased my enthusiasm for the subject matter</td>
<td>3.35</td>
</tr>
<tr>
<td>7. The instructor helped me to improve my problem solving ability</td>
<td>3.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean difference</th>
<th>Standard deviation</th>
<th>Standard error mean</th>
<th>t</th>
<th>Degrees of freedom</th>
<th>Significance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.22</td>
<td>1.383</td>
<td>0.188</td>
<td>1.181</td>
<td>53</td>
<td>0.243</td>
<td>0.2144</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>0.947</td>
<td>0.129</td>
<td>1.294</td>
<td>53</td>
<td>0.201</td>
<td>0.2502</td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>0.971</td>
<td>0.132</td>
<td>2.522</td>
<td>53</td>
<td>0.015</td>
<td>0.6940</td>
</tr>
<tr>
<td>4</td>
<td>0.44</td>
<td>1.383</td>
<td>0.188</td>
<td>2.362</td>
<td>53</td>
<td>0.022</td>
<td>0.6402</td>
</tr>
<tr>
<td>5</td>
<td>0.48</td>
<td>1.514</td>
<td>0.206</td>
<td>2.337</td>
<td>53</td>
<td>0.023</td>
<td>0.6306</td>
</tr>
<tr>
<td>6</td>
<td>0.67</td>
<td>1.441</td>
<td>0.196</td>
<td>3.401</td>
<td>53</td>
<td>0.001</td>
<td>0.9071</td>
</tr>
<tr>
<td>7</td>
<td>0.52</td>
<td>1.328</td>
<td>0.181</td>
<td>2.869</td>
<td>53</td>
<td>0.006</td>
<td>0.8033</td>
</tr>
</tbody>
</table>

a All items (except 3) above were on a fully-anchored, 5-point scale where 1 = strongly disagree; 5 = strongly agree. Item three used a fully-anchored, 5-point scale where, 1 = unacceptable; 5 = excellent.
variable was treated as a fixed, between-subjects factor with \( N = 40 \) subjects in the self-selected (SS) condition (cell A) and \( N = 36 \) subjects in the randomly assigned (RA) condition (cell B). To control for the possibility that results were driven by academic skill level or other demographics, GPA and Gender were included in an ANCOVA model as covariates. Results are presented in Table 4. Hypothesis 3 was not supported, and in fact, the predicted directional effect of the means (i.e., SS = 3.976; RA = 4.039) was not achieved. The power of the test (>0.87) does, however, increase our confidence in the null hypothesis.

Although we did not formally hypothesize a relationship in this regard, we also compared the final exam grades for students as related to the assignment condition, finding performance results similar to those of the group quizzes. Specifically, mean scores on the final were 82.11 for the random assignment condition (\( N = 36; SD = 9.34 \)) and 79.17 for the self-selected condition (\( N = 41; SD = 10.30 \)). The differences were not statistically significant.

Hypothesis 4 reflected the expectation that average quiz performance would be greater for long-term groups (i.e., groups that maintained the same membership for the entire semester) than for short-term groups (i.e., groups whose membership was randomly assigned for all quizzes). As depicted by the experimental design shown in Fig. 2, H4 was tested by a comparison of means of group quiz performance in cell C (within-subjects). Since all subjects (\( N = 76 \)) took all 10 quizzes, but were in long-term groups for the odd and short-term groups for the even numbered quizzes, testing of the time-frame variable was conducted by treating the manipulation as a random, within-subjects factor. As with H3, to control for the possibility that results were driven by academic skill level or other demographics, GPA and Gender were included in an ANCOVA model as covariates. Results are presented in Table 5.

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2: Test of Hypothesis 3 comparison of quiz performance for self-selected vs. randomly assigned groups</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assignment – self-selected vs. randomly assigned group quiz performance</th>
<th>Means</th>
<th>Standard deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-selected quiz group average</td>
<td>3.976</td>
<td>0.365</td>
<td>40</td>
</tr>
<tr>
<td>Randomly assigned quiz group average</td>
<td>4.039</td>
<td>0.291</td>
<td>36</td>
</tr>
<tr>
<td>Total</td>
<td>4.005</td>
<td>0.332</td>
<td>76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of freedom</th>
<th>Mean square</th>
<th>( F )</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANCOVA tests of between-subjects effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>3</td>
<td>0.172</td>
<td>1.602</td>
<td>0.196</td>
</tr>
<tr>
<td>GPA(^a)</td>
<td>1</td>
<td>0.060</td>
<td>0.562</td>
<td>0.456</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>0.337</td>
<td>3.136</td>
<td>0.081</td>
</tr>
<tr>
<td>SSRA(^b)</td>
<td>1</td>
<td>0.104</td>
<td>0.971</td>
<td>0.328(^c)</td>
</tr>
<tr>
<td>Error</td>
<td>72</td>
<td>0.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) GPA = Student GPA from administrative records.  
\(^b\) SSRA = Assignment (two-level factor: self-selected into group and randomly assigned to group).
As with the third hypothesis, H4 was also not confirmed, and the predicted directional effect of the means (i.e., LT = 3.900; ST = 4.114) was not achieved. Again, however, the power (>0.87) adds confidence that performance is not a function of the time frame of group membership, since the risk of a moderate effect not being discovered is less than 0.13.

5. Discussion

This study was designed to provide insight into how cooperative learning could affect students’ performance and attitudes. Specifically, we examined the impact of allowing students to self-select their group members versus having the instructor select the group members. Findings revealed no performance differences between the two types of groups. These findings are consistent with Ravenscroft et al. (1997) who reported results of seven studies, varying team learning and group grade incentives. They found little or no improvement on exam scores when students worked in teams. Lancaster and Strand (2001) also found that cooperative learning did not improve student performance. Our findings suggest self-selection did not produce a beneficial amount of increased cohesiveness that might be expected from allowing students to self-select their own group members. While this was a surprising finding, this result provides evidence for accounting instructors when they are faced with the decision of either choosing group members or allowing students to select their own group members.

This study also investigated actual performance differences between ad hoc and long-term groups. Unexpectedly, there was no difference in performance between the two time frames. Evidently, being in the same group throughout the spring
semester did not benefit the groups’ performance over time. While the general education literature contains many studies that report improved student performance via cooperative learning, similar results in accounting have been elusive (Lancaster & Strand, 2001).

However, we did note that students in the group-condition had significantly different affective reactions than those in the non-group condition. In addition to replicating Hite’s finding that the overall performance of the instructor was rated higher for the group quizzes sections, these students expressed greater motivation to learn and increased enthusiasm. They also expressed a belief that their problem solving abilities had improved and that the instructor helped them to learn the course content more than those in the non-group condition. Although there are cost/benefit tradeoffs with using cooperative learning, we believe this finding is important given that these students did not evidence any significantly different performance results from those in the non-group condition.

Some may argue that the performance results of H1 and affective results of H2 are in conflict with each other. We suggest that this is not a necessary conclusion. While the study evidenced no statistical difference in student performance, this need not significantly correlate with student perceptions of the instructor’s effectiveness. Thus, we suggest that the results of H1 and H2 can easily and beneficially co-exist in the classroom. That is, if an instructor had an expectation of similar performance in both conditions, higher instructor effectiveness ratings would remain beneficial.

Several instructors have argued against the approach of group learning citing that group tasks involve more class time and cause the instructor to sacrifice lecture time or further explanation (Michaelsen & Black, 1994). However, in our study, performance was not degraded. Thus, performance not dropping while student attitudes improved may help alleviate concerns among instructors that cooperative learning may hurt students’ performance or their own teaching evaluation scores. On the other hand, performance did not improve as a result of group learning. Accordingly, we acknowledge the cost/benefit tradeoffs that exist with using a cooperative learning approach.

6. Limitations

There are several limitations to our study. First, while the second experiment was not subject to the same potential weaknesses of the first quasi-experiment, it was only the first study that evidenced significant affective results. Results produced by using quasi-experimental designs are often presented in the education literature (e.g., Hite, 1996; Lancaster & Strand, 2001). However, they remain limited by the potential threat of non-equivalent groups. Nevertheless, we examined relevant covariates in an attempt to mitigate against this possible weakness. Second, the effectiveness of the manipulation of the assignment factor was limited by the response from students that, on average, they did not know many of their
classmates. Thus, the potential attractiveness of self-selection in forming groups was likely diminished. This directly threatens the salience of the self-selection manipulation. Third, while we assume that there are beneficial effects to the cohesiveness and understanding achieved only by long-term groups, we are not sure how long it takes to achieve these effects and how specific these effects are from one context to another. Fourth, both of our experiments suffer from a reduced ability to generalize results that is common to all experimental designs. Fifth, while student attitudes are captured in evaluations of instructors, it is possible that a student may like an instructor but not like accounting. Sixth, the well-known Hawthorne effect may have impacted the students in the treatment condition of the quasi-experiment. That is, since the students knew they were part of an experiment, they may have responded favorably to the increased attention. However, given that the students knew they were being required to take quizzes that are likely not normally required, we were concerned that they might, as a result of having to do more work, provide lower teaching evaluation scores. Thus, we believe the Hawthorne effect was quite unlikely to have occurred. Lastly, because the university aggregates course evaluation responses we could not match individual student perception with performance.

While this study was a first attempt at self-selection versus instructor-assigned group members, another study ought to investigate whether performance does indeed not differ between the two types of assignments. Achieving a salient manipulation of this factor is complicated by the fact that cohesive classes tend to be smaller. Thus, researchers face a challenge in achieving an adequate sample size for testing. Moreover, in addition to the examination of the assignment factor, our study seems to be unique in examining the time frame factor as well. This process should possibly involve longer time frames with multiple, periodic measurements similar to those used here. This achievement also poses a challenge to researchers in the university environment where most classes are limited to 16 weeks at most. While cooperative learning appears to motivate students to learn, more research should investigate the factors that influence why this may be so. Another study might consider other affective variables of interest such as self-esteem, retention of content, and group cohesiveness.

Acknowledgments

We are grateful to the Editor, James E. Rebele, for the detailed and thoughtful comments, and we are especially grateful to the Journal of Accounting Education reviewer who provided excellent guidance on power analysis to make our conclusions more meaningful. We are also appreciative of the reviewers at the 2003 AAA Southeast Regional meeting for awarding this paper ‘Best Manuscript for the Teaching and Curriculum Section,’ and also to the reviewers who chose this paper as ‘Best Paper for the Teaching and Experiential Learning Track’ for the 2003 Southeast Decision Sciences Institute meeting.
Appendix A. Instructions regarding quizzes

1. **Ten quizzes will be administered during the semester.** Each quiz will cover one PowerPoint lecture presentation and will be given immediately following the lecture (i.e., same day).

2. **You will be given each quiz to complete individually and then as a group.** After I collect each quiz completed individually, you will be given a copy of the same quiz to complete in a small group. During this time, you should openly discuss what you believe to be the best answers for the questions on the quiz with your group members. The group should strive to reach a consensus on quiz answers, but if that fails, you should work it out as you and the others in your group see fit. You will then turn in the quiz as a group, putting all group member names on the quiz.

3. **The grade I record for each individual will be equal to the average of the individual and group quiz scores.** This is designed to encourage you to perform to the best of your ability both individually and as a contributing group member. Quiz grades will be worth 10% of the course grade.

4. **A post-quiz questionnaire (completed individually) will be administered following every quiz.** This questionnaire will ask how each group member contributed to the group quiz results, how much you studied or otherwise prepared for the quiz, etc. The idea here is to gather information regarding the effectiveness and fairness of the group process.

5. **Level of individual preparation – not to mention attendance – will affect both your grade and the grade of others in your group.** There will almost always be at least one question on each quiz that can only be answered by attending and listening carefully to the lecture. Also, there is an attendance requirement: *Excused* absences of group members during the term will be dealt with on a case-by-case basis. However, the *group* quiz grade will not count and be averaged with your individual grade. Unexcused absences will receive no credit and will hurt the group effort by denying the other group members access to your knowledge.

6. Regarding group assignments, *all* of you will be a member of a group that you will maintain membership in for the entire semester. **About half of you will be randomly chosen and assigned by me to a group.** The rest of you will be asked to self-select your own group members. The purpose of this is to determine if there is a difference in performance between the self-selecting groups and the assigned groups. There are tradeoffs to both approaches. For example, self-selection allows group members to choose to work with people that they believe will well-serve the purposes of the group. However, assignment allows a potential benefit of experiencing the growth and learning inherent in working with unfamiliar people and adjusting for mutual benefit.

7. Also, regarding group assignments, **on every other quiz, I will require all class members to be in a group requiring new and different members.** All individuals, regardless of whether they previously self-selected their group or were assigned will be assigned new members randomly. The objective of this is to determine the relative importance of self-selection and assignment when compared to temporary group arrangements.
Appendix B. Initial and semester-end questionnaire format

1. My gender is
   a. Female
   b. Male

2. My age is _________.

3. I believe that I will be treated fairly in terms of the process of being either allowed to self-select my group members or in being assigned to a group during the semester.
   a. Strongly agree
   b. Agree
   c. Not sure
   d. Disagree
   e. Strongly disagree

4. I believe that I will be treated fairly in terms of the group outcome (i.e., grade impact) during the semester.
   a. Strongly agree
   b. Agree
   c. Not sure
   d. Disagree
   e. Strongly disagree

5. I am currently satisfied with the group arrangement (self-selection or assigned) or group process I am in and/or anticipate during the semester.
   a. Strongly agree
   b. Agree
   c. Not sure
   d. Disagree
   e. Strongly disagree

6. I expect to be satisfied with the group outcome (i.e., grade impact) during the semester.
   a. Strongly agree
   b. Agree
   c. Not sure
   d. Disagree
   e. Strongly disagree

7. How many people in the class are you familiar with regarding their academic performance in general (i.e., not just on quizzes) relative to your own (either real or perceived)?
   a. None or very few class members
   b. Less than half the class
   c. About half the class
   d. More than half the class
   e. All or most class members

8. I am expecting to experience a learning benefit (or some other benefit other than just a grade benefit) from participating in the group experience.
   a. Strongly agree
   b. Agree
   c. Not sure
   d. Disagree
   e. Strongly disagree

---

Questions 1 and 2 presented here were only presented on the Initial questionnaire for the gathering of demographic data. Otherwise, the format and presentation of the two questionnaires was virtually identical.
Appendix C. Post-quiz questionnaire

1. Which of the following most accurately reflects the degree to which you prepared for the quiz by involving other group members to discuss information relevant to the quiz outside of class prior to taking it?
   a. I prepared strongly with other members
d. I prepared some with other members
   b. I prepared more than moderately with other members
e. I did not prepare at all with other members
   c. I prepared moderately with other members

2. Which of the following most accurately reflects the degree to which you individually prepared for the quiz outside of class prior to taking it?
   a. I prepared strongly
   b. I prepared more than moderately
   c. I prepared moderately
d. I prepared some
e. I did not prepare at all

3. I felt like it did not matter if I studied or prepared much before the quiz because it was so difficult (i.e., I would do about as well whether I studied or not).
   a. Strongly agree
d. Disagree
   b. Agree
e. Strongly disagree
   c. Not sure

4. I felt like it did not matter if I had studied or prepared much before the quiz because it was so easy (i.e., I would do about as well whether I studied or not).
   a. Strongly agree
d. Disagree
   b. Agree
e. Strongly disagree
   c. Not sure

References


