

1-1-2009

Using Peer Grading and Proofreading to Ratchet Student Expectations in Preparing Accounting Cases

Laura R. Ingraham
San Jose State University, laura.ingraham@sjsu.edu

B. A. Chaney
University of Montana

Follow this and additional works at: https://scholarworks.sjsu.edu/acc_fin_pub



Part of the [Accounting Commons](#), and the [Finance and Financial Management Commons](#)

Recommended Citation

Laura R. Ingraham and B. A. Chaney. "Using Peer Grading and Proofreading to Ratchet Student Expectations in Preparing Accounting Cases" *American Journal of Business Education* (2009): 39-48.

This Article is brought to you for free and open access by the Accounting and Finance at SJSU ScholarWorks. It has been accepted for inclusion in Faculty Publications by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Using Peer Grading And Proofreading To Ratchet Student Expectations In Preparing Accounting Cases

Barbara A. Chaney, University of Montana, USA
Laura R. Ingraham, San Jose State University, USA

ABSTRACT

Accounting educators struggle with ways to incorporate the development of critical thinking and communication skills into the curriculum. Case analysis is one tool for developing these skills. We examine whether students' case analysis scores improve as a result of participation in peer grading and peer review. We find that students improve their ability to perform case analyses after both evaluating and being evaluated by student peers. Students initially experience an Expectation Ratcheting learning effect after evaluating the case of a peer. Subsequently, students experience an Enhanced Feedback learning effect from the comments and suggestions made by the peers who evaluated (proofread) their cases.

Keywords: peer evaluation, peer grading, proofreading, student expectations, feedback

INTRODUCTION

Accounting educators search for pedagogical tools that will foster students' critical thinking and communication skills because they understand the importance of these skills. The American Institute of Certified Public Accountants (AICPA) spearheaded the CPA Vision Project to identify the core competencies required of accounting professionals. Its report asserts that future accountants must be able to "link data, knowledge, and insight together," "interpret and provide a broader context using financial and non-financial information," and "give and exchange information within meaningful context and with appropriate delivery and interpersonal skills." (AICPA, 2000, p. 11)

The Accounting Education Change Commission (AECC), after years of study and the publication of a detailed research report, issued a Position Statement on the objectives of education for accountants. It recommends that accounting programs consider content, process, and attitude. Content should be more than rote memorization, process should require analysis and interpretation of accounting and other information, and attitude should encourage the ability to solve unstructured problems. In addition, the AECC stresses that the "overriding objective of accounting programs should be to teach students to learn on their own. . . . Students must be active participants in the learning process, not passive recipients of information." (AECC, 1990, p. 309)

Case analysis is a useful pedagogy for promoting critical thinking and communication skills (Campbell and Lewis, 1991). It can be made even more useful by adding peer reviewing and grading to the pedagogy. Active student participation creates the potential for rich feedback. Intuitively, one expects that any feedback received is a resource students use to enhance their learning. However, Scofield and Combes (1993) suggest that students do not fully utilize this resource when the feedback comes as red ink from the professor. In fact, research suggests that students might value the opinions of their peers more than the opinions of the instructor. "A student or student group receiving feedback from members of the class might give it more credence than the feedback they receive from their instructor. At the very least, student feedback reinforces instructor feedback" (Hirsch and Gabriel, 1995, p. 264).

We examine whether peer reviewing and grading improves accounting students' scores on financial statement case analyses in an introductory financial accounting course. We find that students improve their ability to perform case analyses after both evaluating and being evaluated by student peers. Students initially experience an Expectation Ratcheting learning effect after evaluating the case of a peer. Subsequently, students experience an Enhanced Feedback learning effect from the comments and suggestions made by the peers who evaluated their cases. However, we caution instructors to select assignments appropriate for use with peer evaluation and carefully manage the peer evaluation process.

DEVELOPMENT OF HYPOTHESES

The term "peer evaluation" as used here describes a pedagogical technique that is a combination of peer grading and peer review. Peer grading is ex post evaluation, involving students in reading and grading the efforts of their peers. Peer review is ex ante evaluation, in which students proofread and make comments on their peers' work before final submission to the instructor. (Kerr, Park, and Domazlicky, 1995) We use peer evaluation to identify a pedagogy where student evaluators both grade their peers' work *and* provide feedback. When both characteristics are incorporated in an iterative, peer evaluation pedagogy, students learn from both giving and receiving feedback.

A well-constructed peer evaluation process allows students to receive significant feedback that is relevant and intriguing to them, without placing an undue grading burden on the instructor. By considering the peers' feedback, students should learn to focus and perfect their critical thinking and communication skills on subsequent assignments. By providing feedback, students should learn as a result of being exposed to the different styles, insights, and intellectual processes of their peers. This should sensitize students to the need for effort, critical thinking, and good communication skills in their subsequent assignments. Finally, if students feel more accountable to their peers than to the instructor, they may learn more from an assignment that will be peer graded because they will put more effort into the assignment, ex ante.

The essence of our empirical observation of learning enhancement is changes in students' performance on financial statement case analyses. Because student performance varies from student to student, class section to class section, and as a result of other learning effects besides peer evaluation, we used four treatment groups randomly spread over multiple sections of an introductory accounting course. By using a between and within subjects research design, we attempt to isolate the cause of changes in performance according to specific learning effects:

- (1) Accountability awareness (AA) from knowing ahead of time that *peers* will evaluate an assignment,
- (2) Enhanced feedback (EF) from the explicit feedback received on a previous assignment that was peer evaluated (i.e., proofread),
- (3) Expectation ratcheting (ER) from raising self-expectations after having previously evaluated a peer's assignment and becoming attuned to the level of thought and effort required to perform adequately.

We present the following hypotheses about the specific learning effects created by student peer evaluation:

- H_A:** Students' scores on case analysis assignments will improve if the students are aware that their assignments will be peer evaluated (*Accountability Awareness*).
- H_B:** Students' scores on case analysis assignments will improve if the students have previously received significant feedback from the peer evaluation process (*Enhanced Feedback*).
- H_C:** Students' scores on case analysis assignments will improve if the students have previously peer evaluated other students' assignments (*Expectation Ratcheting*).

RESEARCH DESIGN

Approximately 460 students from 12 sections of an introductory financial accounting course at a large, public, southeastern U.S. university were given five simple financial statement analysis case assignments over the course of the semester. The cases are based on the annual report of a publicly traded company provided in an appendix to the textbook used in the course. The points for these assignments accumulated to 15 percent of a student's grade.

Experimental Cases

The cases are based on actual cases provided in the Libby, Libby, and Short textbook used for the course. Minor modifications were made to the textbook assignments to make the five cases as similar as possible in terms of difficulty. All five cases consisted of five questions. Two or three of the questions required the simple calculation of a financial ratio or the extraction of an accounting number from the financial statements. Usually two questions required analytical thinking and at least one question posed a communication challenge. Students were not made aware of the fact that the cases were based on textbook assignments. Instead, the cases were typed and handed out to the students.

The first four cases were assigned after chapters 5, 6, 7, and 8 of the Libby et al textbook. The fifth case was not assigned until after chapter 13 in order to isolate persistence effects. The time lag would omit any new learning effects yet allow for the persistence of improvements in students' learning. Students were asked to complete questionnaires at the end of the semester. Responses provided demographic information to use as covariates. In addition, open-ended comments provided insight into the results.

Treatment Groups

The twelve class sections were assigned to one of four groups. A "Control" group prepared all case assignments as routine homework assignments without any experimental treatment. Three treatment groups prepared the first case (after chapter 5) without any experimental manipulation as an additional control. All control case analyses were graded by a grading team consisting of seniors and master's students who had completed the introductory accounting course with a grade of "A."

The three treatment groups received unique experimental treatments in order to isolate the various hypothesized learning effects. An "Evaluators" treatment group evaluated the cases of peers, but did not receive peer evaluations. Evaluators were given explicit guidance as to how to conduct the peer evaluation both in terms of how to grade the case and how to provide explicit feedback. For peer review to be successful, it is important to establish a set of criteria for students to use (Hirsch and Gabriel, 1995). To motivate the Evaluators, the grading team graded the peer evaluations based on the quality and quantity of feedback the Evaluators provided. In addition, the grading team graded the peer evaluator on how well the case was graded. This characteristic of the experiment was intended to minimize the halo effect (grade inflation) often found when peer grading is utilized.

The "Evaluatees" treatment group did not perform peer evaluations but its cases were peer-evaluated by another group, thus it received the feedback of its peers. By assigning individual "Evaluators" and "Evaluatees" groups we are able to isolate the Expectation Ratcheting learning effect in the former and the Enhanced Feedback learning effect in the latter. We also used a "Full Treatment" group that both performed evaluations and was evaluated in order to observe whether the Expectation Ratcheting and Enhanced Feedback learning effects occur at the same time, consecutively, or are cumulative.

Participants

We began our experiment with approximately 460 students enrolled in twelve normal sections of introductory financial accounting. Before analyzing the results, however, we eliminated three sections. Two were eliminated because the instructor was unable to provide us with the pre-experiment exam scores that we used as an important covariate. The results presented are not substantively different from results from an analysis including those two eliminated sections but eliminating the covariate. We also eliminated one section from the analysis that was taught by an instructor who focuses on financial statement analysis as a primary pedagogical tool. This section was an outlier due to the students' extraordinary preparation for the case assignments. The students in this section performed significantly better on all the case assignments than students in other sections. "Section within a Group" was an important covariate to include in our model, even after dropping this outlier section.

To complete our repeated measures analysis, we eliminated all students who did not perform all of the assignments. In addition, we eliminated 12 students who scored less than 4 out of 10 on a case. Such an extremely

low score indicates a lack of effort on the student’s part and creates an extreme improvement on the subsequent case. Our final set of observations for evaluation was 175 students. We also conducted sensitivity tests on a reduced set of observations that includes demographic information, such as gender and GPA, that is not available for all 175 students. The results are not substantively different from those presented. Thus, they are not reported.

RESULTS

We provide descriptive statistics for our sample of 175 students in Table 1. The case scores are measured out of 10 possible points. The case score means and standard deviations for the control and treatment groups are presented in Panel A. The covariate-adjusted means of the improvements in scores are presented in Panel B. Adjustments are made to the case score improvement means because of the unequal number of observations in the four groups. The four improvements relate to the five cases. For example, Improvement 1 is the difference between Case 2 and Case 1.

The control group performed better on the control case (Case 1, Chapter 5) than all treatment groups. However, the treatment groups experienced better improvements than the control group on all cases except the last one. All groups performed worse on the second case (Chapter 6). This is evidence, in hindsight and upon review of student comments, of our failure to adequately control the level of difficulty between the first and second case. However, the effect of unequal difficulty levels is mitigated because we analyze improvements in scores rather than raw case scores, and we use the first case score as a covariate in all analyses.

The results of hypothesis testing are difficult to interpret because of the complicated research design. Therefore, it is useful to consider the cells of Panel B of Table 1 to appreciate the between subjects and within subjects nature of the design and to consider our expectations of when learning effects would occur. At the time of Improvement 1 (at the completion of Case 2), the Evaluatees and the Full Treatment students were expected to have an Accountability Awareness effect from knowing their peers would evaluate their cases (H_A). The improvements of those students are expected to be greater than those of the Control and Evaluators.

Table 1
Descriptive Statistics For 175 Students
By Treatment Group

| | Control Group n=34 | Evaluatees Only n=39 | Evaluators Only n=32 | Full Treatment n=70 |
|---|-------------------------------|---------------------------------|---------------------------------|--------------------------------|
| Panel A: Mean Case Scores (Std Deviations) Out of 10 Points | | | | |
| Case 1 (Chp 5) | 9 (1.28) | 8.79 (1.15) | 8.84 (1.02) | 8.93 (1.03) |
| Case 2 (Chp 6) | 6.85 (1.62) | 6.69 (1.51) | 6.66 (1.60) | 6.59 (1.44) |
| Case 3 (Chp 7) | 8.15 (1.28) | 8.08 (1.51) | 8.84 (1.55) | 7.91 (1.32) |
| Case 4 (Chp 8) | 7.47 (1.26) | 8.28 (1.59) | 8.31 (1.33) | 8.44 (1.36) |
| Case 5 (Chp 13) | 8.35 (2.10) | 7.85 (1.95) | 8.00 (2.42) | 8.04 (1.88) |
| Panel B: Covariate-Adjusted Mean Improvements in Case Scores | | | | |
| Improvement 1 | -1.65 | -2.39 | -2.18 | -2.29 |
| Improvement 2 | + .80 | +1.36 | +2.06 | +1.31 |
| Improvement 3 | -.57 | + .48 | -.52 | + .56 |
| Improvement 4 | +1.11 | -.88 | -.10 | -.43 |

At Case 2 we introduced peer evaluation treatments. Therefore, Improvement 2 should have produced observable learning effects. However, the learning effects depended upon whether the students were performing or receiving the evaluation. The Enhanced Feedback learning effect of H_B would be suggested if the Evaluatees’ improvement were greater than the Control group’s. The Expectation Ratcheting learning effect of H_C would be

suggested if the Evaluators' improvement were greater than the Control group's. The Full Treatment group at Improvement 2 is a mixture of learning effects that cannot be observed until subsequent case assignments.

Observing the learning effects at subsequent case assignments provides a rich experimental setting for testing the stated hypotheses and pursuing an understanding of how the learning effects work together and how they persist. Based on the design of the experiment, any learning effects found for Evaluators can be considered Expectation Ratcheting and for Evaluatees can be considered Enhanced Feedback. Interpretation of the Full Treatment group results over the five cases is more difficult. The statistical results from these expectations will be discussed in the following sections.

Results of Repeated Measures Analysis of Variance

The results from our initial repeated measures analysis are provided in Table 2. The test of overall within subjects effects in the repeated measures analysis indicates that treatment group is significant (at less than .01) in explaining case score improvements. The test of overall between subjects effects (an effect between each and every treatment and control group overall throughout the experiment) indicates that treatment group is not significant (p-value of .204) in explaining improvements in case scores.

Table 2

| Repeated Measures Analysis Of Variance Of Improvements In Case Scores For 175 Students | | | | | |
|---|----------------|-----|--------------|---------|-------|
| All Four Treatment Groups Combined For An Overall Test Of Between And Within Subjects Effects On Four Case Score Improvements | | | | | |
| Source of Variation | Sum of Squares | df | Mean Squares | F-value | p<* |
| Panel A: Tests of Between Subjects Effects | | | | | |
| Peer evaluation group | 4.56 | 3 | 1.52 | 1.55 | .204 |
| Section within group | 15.90 | 5 | 3.18 | 3.24 | .001 |
| Case 1 | 34.54 | 1 | 34.54 | 35.16 | <.001 |
| Pre-Exam | .32 | 1 | .32 | .32 | .572 |
| Error | 161.15 | 164 | .98 | | |
| Panel B: Tests of Within Subjects Effects | | | | | |
| Improvement | 43.39 | 3 | 14.46 | 3.69 | .012 |
| Improvement x Group | 130.09 | 9 | 14.45 | 3.69 | <.01 |
| Improvement x Section | 193.88 | 15 | 12.93 | 3.30 | <.001 |
| Improvement x Case 1 | 124.49 | 3 | 41.50 | 10.60 | <.001 |
| Improvement x Pre-Exam | 28.69 | 3 | 9.56 | 2.44 | .064 |
| Error | 1926.12 | 492 | 3.91 | | |
| * All tests are two-tailed | | | | | |

The between subjects test results are likely understated because of the complicated research design. We have three treatment groups, one of which is a combination of the other two treatment effects. A lack of overall results is not necessarily evidence that there are no significant differences among the individual treatment groups or between an individual treatment group and the control group. In order to analyze the effects between and among groups, we conduct separate analyses for each specific improvement. The results from the individual analyses will provide evidence for our specific hypotheses on Accountability Awareness, Enhanced Feedback, and Expectation Ratcheting. We discuss each of our hypotheses in conjunction with the individual analyses of covariance of improvements provided in Table 3.

Results of Individual Analyses of Covariance

We hypothesize in Hypothesis H_A that students who know they will be peer evaluated will expend greater effort and improve their performance on the case analyses. When students were assigned Case 2, those in the Evaluatees and Full Treatment groups were made aware they would be peer evaluated. Therefore, we expect the improvements for the Evaluatees and Full Treatment groups to be significantly greater than the Control and

Evaluators groups. We present the result of the analysis of Improvement 1 (Case 2 score less Case 1 score) in Table 3, Panel A. We find no significant treatment effect between the subjects in the four groups (p-value of .246). In fact, from Panel B of Table 1, we see that Improvement 1 for the Evaluatees and Full Treatment groups is actually more negative than the Control and Evaluators groups'. The results do not provide evidence in support of the Accountability Awareness hypothesis.

Table 3

| Analysis Of Covariance On Individual Models Of Improvement In Case Scores For 175 Students | | | | | |
|--|----------------|-----|--------------|---------|-------------|
| All Four Treatment Groups Combined To Test Treatment Effects On Individual Case Score Improvements | | | | | |
| Source of Variation | Sum of Squares | df | Mean Squares | F-value | p<* |
| Panel A: Improvement 1 (from Case 1 to Case 2) | | | | | |
| Peer evaluation group | 9.04 | 3 | 3.01 | 1.40 | .246 |
| Section within group | 29.49 | 5 | 5.90 | 2.73 | .021 |
| Case 1 | 154.63 | 1 | 154.63 | 71.58 | <.001 |
| Pre-Exam | 4.83 | 1 | 4.83 | 2.24 | .137 |
| Error | 354.30 | 164 | 2.16 | | |
| Model | 182.01 | 10 | 18.20 | 8.42 | <.001 |
| Panel B: Improvement 2 (from Case 2 to Case 3) | | | | | |
| Peer evaluation group | 22.67 | 3 | 7.56 | 2.26 | .083 |
| Section within group | 42.29 | 5 | 8.46 | 2.53 | .031 |
| Case 1 | 1.84 | 1 | 1.84 | .55 | .458 |
| Pre-Exam | 3.16 | 1 | 3.16 | .95 | .332 |
| Error | 547.33 | 164 | 3.34 | | |
| Model | 70.40 | 10 | 7.04 | 2.11 | .026 |
| Panel C: Improvement 3 (from Case 3 to Case 4) | | | | | |
| Peer evaluation group | 42.22 | 3 | 14.07 | 4.72 | .004 |
| Section within group | 30.01 | 5 | 6.00 | 2.01 | .080 |
| Case 1 | 2.27 | 1 | 2.27 | .76 | .384 |
| Pre-Exam | 5.64 | 1 | 5.64 | 1.89 | .171 |
| Error | 489.14 | 164 | 2.98 | | |
| Model | 87.72 | 10 | 8.77 | 2.94 | .002 |
| Panel D: Improvement 4 (from Case 4 to Case 5) | | | | | |
| Peer evaluation group | 60.72 | 3 | 20.24 | 4.77 | .003 |
| Section within group | 108.00 | 5 | 21.60 | 5.09 | <.01 |
| Case 1 | .28 | 1 | .28 | .07 | .797 |
| Pre-Exam | 15.37 | 1 | 15.37 | 3.62 | .059 |
| Error | 696.49 | 164 | 4.25 | | |
| Model | 172.94 | 10 | 17.29 | 4.07 | <.001 |
| * All tests are two-tailed | | | | | |

We hypothesize in H_B that students will improve their performances on cases after receiving enhanced feedback from their peers. Therefore, we expect that the Evaluatees' improvement from Case 2 to Case 3 (Improvement 2) will be greater than the improvement of the Control group, and the adjusted mean improvement presented in Panel B of Table 1 suggests the hypothesized result. The overall analysis of covariance of Improvement 2 for all treatment groups is presented in Panel B of Table 3. There is a marginally significant (p-value of .083) treatment group effect on Improvement 2 when all groups are included. However, when the analysis of covariance includes only the Evaluatees and the Control group to isolate the Enhanced Feedback learning effect, the treatment group effect is not statistically significant (p-value was .25, results not reported). Thus, it appears there is no statistical difference between the Evaluatees and the Control group.

In Hypothesis H_C we hypothesize that students will improve their performances on cases by ratcheting up their self-expectations after evaluating their peers' performances. Because we found no Enhanced Feedback learning effect at Improvement 2, it is possible that the Expectation Ratcheting effect is being obscured by the marginal significance of overall treatment group (treatment group effect with a p-value of .083 in Panel B of Table

3). Thus, we prepared an analysis of covariance of Improvement 2 using only Evaluators and the Control group. We find that treatment group is statistically significant when the model is structured to isolate the Expectation Ratcheting effect (p-value of .01, results not reported). Thus, the difference between the Evaluators and the Control group suggested in Panel B of Table 1 at Improvement 2 is statistically significant. This is consistent with Expectation Ratcheting.

It appears from these results that initial exposure to peer evaluation through Case 3 creates a learning effect from performing the evaluation—Expectation Ratcheting—even though there is no initial learning effect from being evaluated—Enhanced Feedback. This likely explains why there is no statistically significant result when analyzing the covariance of Improvement 2 comparing the Control group and the Full Treatment group. The Full Treatment was confounded with both an “evaluator” and “evaluatee” treatment. If one treatment was effective and the other one not effective, the statistical significance of the treatment would likely be compromised. In addition, students do not appear affected by knowing their peers will be evaluating their cases—Accountability Awareness. We summarize these results in an informal overview in Table 4, which reports learning effects observed according to treatment groups.

Table 4

| Summary Of Hypothesized And Actual Results From Analysis Of Covariance Of Individual Improvements Using Specific Treatment Groups In Comparison To The Control Group For 175 Students | | | | |
|--|-------------------------------|---------------------------------|---------------------------------|--|
| | Control Group n=34 | Evaluatees Only n=39 | Evaluators Only n=32 | Full Treatment n=70 |
| Improvement 1 | | Accountability Awareness | | Accountability Awareness |
| Improvement 2 | | Enhanced Feedback | Expectation Ratcheting* | Enhanced Feedback & Expectation Ratcheting |
| Improvement 3 | | Enhanced Feedback* | Expectation Ratcheting | Enhanced Feedback* & Expectation Ratcheting |
| Improvement 4 | | Enhanced Feedback | Expectation Ratcheting | Enhanced Feedback & Expectation Ratcheting |
| * In Analysis of Covariance of Improvement using specific treatment group with Control group, treatment effect is significant at $\leq .01$ in the hypothesized direction. Other results not reported but similar to results presented in Table 3, Panels A through D. | | | | |

We continue our individual analyses of improvements in case scores in order to determine whether the Expectation Ratcheting learning effect persists and whether the Enhanced Feedback learning effect occurs later. At Improvement 3 we find a statistically significant overall treatment effect with a p-value of .04. The results of the analysis of covariance are presented in Panel C of Table 3. The statistical significance, however, does not indicate the nature of the learning effect(s) because it is an overall effect for all treatment groups. Therefore, we analyze the covariance of Improvement 3 using the Control group matched to Evaluatees to isolate the Enhanced Feedback learning effect and matched to Evaluators to isolate the Expectation Ratcheting learning effect.

We find that Expectation Ratcheting does not persist—the treatment is not statistically significant when comparing the Evaluators and the Control group at Improvement 3. Peer grading attunes students to the effort required to improve their performance, but it is a one-time effect. However, we find that an Enhanced Feedback learning effect finally occurs at Improvement 3 because the treatment effect is statistically significant (p-value of .01) when comparing the Evaluatees and the Control group and when comparing the Full Treatment and Control groups. (See Table 4.) Only then do students avail themselves of the Enhanced Feedback available from their peers to further improve their own performances. It appears that students require a first round of proofreading as sensitization before a learning effect takes place at the second round.

The results of the analysis of covariance of Improvement 4 are presented in Panel D of Table 3. Although the p-value of .003 of the treatment effect suggests one or more learning effects, the nature and direction of

Improvement 4 suggested in Panel B of Table 1 suggest otherwise. All treatment groups performed worse than the Control group! In fact, each group's treatment has a statistically significant effect on the covariance of Improvement 4, but in the opposite direction expected. Therefore, no significant results are presented in Table 4 for Improvement 4.

There is no hypothesized explanation for the Improvement 4 results. One anecdotal explanation we offer relates to the four-week time lag between Case 4 and Case 5. It is possible that the Enhanced Feedback and Expectation Ratcheting learning effects are transitory. Further research is necessary to explore this possibility. Another anecdotal explanation we offer is prompted by student comments on exit surveys. In summary, many students were frustrated with the peer evaluation pedagogy because they felt they were used as "slave labor," that the grading was "too subjective," and mostly that they felt inadequately prepared for financial statement analysis. Our explanation is that when the students got to the last case, they were tired of the pedagogy and unwilling to give it any more effort. It was the last week of the semester and they wanted to prepare for Final Exams.

While the results for Improvement 4 are ambiguous for rejecting or failing to reject our hypotheses, we have learned something useful about using peer evaluation in accounting classrooms. First, an ideal application for peer evaluation will likely have components to it that must be graded subjectively. Because subjective grading is difficult for students, instructors must provide explicit directions and continuous direct support. Second, instructors should promote the benefits of peer evaluation whenever they use the pedagogy to sensitize students to the learning effects. Third, instructors must be committed to the pedagogy, but willing to adapt it if necessary. Fourth, and most importantly, the pedagogy is not a replacement for teaching the underlying material.

CONCLUSIONS

Student peer evaluation requires active participation that develops analytical and communication skills. We envision its usage in an accounting curriculum as twofold. First, students grade other students' assignments based on explicit grading criteria. Second, students provide explicit written feedback (i.e., comments, questions, corrections, and suggestions) on the assignment. The students learn from receiving the feedback, by critically reviewing the work of others and forming opinions on that work, by communicating the feedback, and by developing their own model of the learning process. Thus, students' learning should be enhanced by peer evaluation. We find that students' learning is enhanced both by evaluating and being evaluated by their peers through Expectation Ratcheting and Enhanced Feedback learning effects.

We recommend accounting educators incorporate peer evaluation into their curriculum. It is a useful technique for introducing more case analysis into the classroom curriculum. However, we caution educators to carefully manage the peer evaluation process. Students must receive adequate in-class training to complete the assignments and must understand the usefulness of the pedagogy. Otherwise they may become frustrated. In addition, students may become resistant if they feel that peer evaluation is being overused.

Peer evaluation is not well suited for every course or every assignment. Peer-evaluated assignments must be suitable for *both* peer grading and peer reviewing because the pedagogy includes both components. Peer *grading* is best suited for assignments where students can objectively measure the grade, such as a problem with a quantitative answer. Subjectivity in the grading criteria can cause grader anxiety that may be counterproductive. Peer *reviewing* is most effective in writing assignments where critical thinking and communication are displayed. Financial statement analysis assignments in any financial accounting (introductory, intermediate, or advanced) or financial statement analysis course seem to blend the characteristics of both. The same could be said for case analyses used in any management or cost accounting course. The highly subjective nature of auditing or systems/information technology might make it challenging for an educator to find an appropriate assignment for peer evaluation in those courses. The educator may choose to incorporate the peer *review* characteristics rather than the peer *grading* characteristics of peer evaluation in those courses.

AUTHOR INFORMATION

Barbara Chaney is an associate professor of accounting at the University of Montana in Missoula. Her area of specialization is governmental accounting. Laura Ingraham is professor of accounting at San Jose State University in California. Her area of specialization is information systems. Both professors conduct pedagogical research.

REFERENCES

1. Accounting Education Change Commission, "Position Statement No. One, Objectives of Education for Accountants," *Issues in Accounting Education*, Vol. 5, No. 2, pp. 307-312, 1990.
2. Campbell, J.E. and W.F. Lewis, "Using Cases in Accounting Classes," *Issues in Accounting Education*, Vol. 6, No. 2, pp. 276-283, 1991.
3. Hirsch, M. L. and S. L. Gabriel, "Feedback Strategies: Critique and Evaluation of Oral and Written Assignments," *Journal of Accounting Education*, Vol. 13, No. 3, pp. 259-279, 1995.
4. Kerr, P.M., K.H. Park, and B.R. Domazlicky, "Peer Grading of Essays in a Principles of Microeconomics Course," *Journal of Education for Business*, Vol. 70, No. 6, pp. 357-364, 1995.
5. Scofield, B.W. and L. Combes, "Designing and Managing Meaningful Writing Assignments," *Issues in Accounting Education*, Vol. 8, No. 1, pp. 71-85, 1993.

NOTES