Developmental and Evaluative Contextual Usage of Peer Assessment of Research Presentations in a Graduate Tax Accounting Course

Roger Lirely
Professor of Accounting
The University of Texas at Tyler
3900 University Blvd
Tyler, Texas 75799
USA

Kerri M. Keech
Assistant Professor of Marketing
The University of Texas at Tyler
3900 University Blvd
Tyler, Texas 75799
USA

Courtney Vanhook
Gould Killian CPA Group, P.A.
100 Coxe Avenue
Asheville, North Carolina 28801
USA

Philip Little
Professor of Accounting
Coastal Carolina University
P.O. Box 261954
Conway, SC 29528
USA

Abstract

Prior research has found mixed results on the utility of peer assessment as either a proxy for instructor grading or a means of providing ungraded feedback to peers. This study examined the correlation between peer-assessed and instructor-assessed research presentations in a graduate federal income taxation class over several semesters. Using identical criteria articulated in a grading rubric, both peer and instructor assessments were used to determine grades on the research presentations and to provide feedback to the presenters. The study found that although student-assigned grades were significantly higher than instructor-assigned grades both overall and across criteria, the student- and instructor-assigned grades tended to move in the same direction across students and criteria. These results suggest that peer assessment may not be suitable as a proxy for instructor grading but may be useful in providing ungraded feedback.

Keywords: peer assessment, higher education, verbal presentations, taxation

1. Background

Instructors often have students assess peer work. Sadler and Good (2006) noted that having peers assess others' work enabled teachers to enrich the learning processes by allowing them to employ more open-ended and sophisticated assessments with minimal encroachment upon a critical resource--the teacher's time. When teachers are pressed for time, they may compromise "…in the use of assessments that teachers feel are less than optimal but easier to grade" (Sadler and Good, 2006, p. 2). An additional efficiency associated with employing peer assessments is that students may receive more extensive and expedient feedback from peers than from a teacher (Gibbs, 1999).Efficiency is not the only benefit of employing peer assessment. The process of peer assessment may increase learning and understanding and provide students with desirable workplace skills.

Kerr and Park (1995) asserted that if students go on to fulfill leadership positions, they will need to know how to assess employees' performance as well as their own performance. Cassidy (2006) argued that peer assessment increased employability because it offers students the opportunity to learn a valuable job-related skill--how to assess peers. The developmental and evaluative benefits to students included giving and receiving constructive feedback, which are critical skills important for future employment (Henning and Marty, 2008). Another benefit of peer assessment is its potential for increased learning of course-related material due to the student having to spend more time mastering the material in order to complete peer assessment (van Hattum-Janssen & Lourenço, 2006; Walker & Warhurst, 2000). van Hattum-Janssen and Lourence (2006) offered another reason for increased mastery of subject content in that as students became more responsible for their own education, they also became more motivated learners. Increased learning may also be attributable to the desire of students to perform well because they know that a peer will evaluate their work (Pecham 1978). Liu and Carless (2006) argued that the process of expressing and articulating to others in a peer assessment process enables students to understand at higher levels of complexity.

Much of the research of peer assessment has focused on the reliability of peer grades. Kerr and Park (1995) found that students gave higher grades than the instructor. They noted, however, that students were able to distinguish good work from bad work and suggested that the use of clear grading guidelines and practice through repeated peer assessments may improve students' abilities to assess the work of others. Fahr, Cannella, and Bedeian (1991) found that the leniency of student grading of peers is context sensitive. Peer raters tended to grade more leniently across both peers and dimensions of an assignment when the ratings were used for evaluative purposes than when used for developmental purposes. van Hattum-Janssen and Lourence (2006) suggested that students should be familiarized with the criteria that are being used for grading in order for peer assessment to be accurate and effective. Similarly, Sadler and Good (2006) found that when students had training and a rubric to utilize, the correlation between student grading and instructor grading was very high. Walker and Warhurst (2000) found that students had a better understanding of their responsibilities as an assessor when they are involved in the process of developing the grading criteria. Whether or not students participate in their development, grading criteria requiring a student to possess knowledge of course material should include more specific guidance than qualitative criteria, since it is assumed that the student has some level of comprehension about those criteria already (van Hattum-Janssen & Lourenço, 2006).

Cho and Wilson (2006) concluded that peer assessment was reliable only when four to six student-given grades were averaged. Another study found that peer assessment was more reliable than self-assessment and that peer assessment improved with practice (Papinczak, Young, Graves and Haynes, 2007). Study findings indicate that the more students are exposed to peer assessment, the more capable and effective a student becomes at assessment (Kerr & Park, 1995; Liu & Carless, 2006).Other studies have examined student opinions about the use of peer assessments in both developmental and evaluative contexts. Cassidy (2006) found that students had a more positive outlook on peer assessment when it was used for *formative* feedback (developmental context) rather than for *summative* assessment (evaluative context). Overall, students expressed a positive attitude toward peer assessment but were more comfortable with being assessed than with assessing. Davis et. al (2007), who employed the terms continuous improvement and accountability rather than formative and summative, stated that student frustration occurred due to the clashing of these two purposes of peer assessment. The authors sought to address students' concerns over the subjectivity inherent to multiple assessors by developing rubrics for peer assessments in future courses. Cho, Schunn, and Wilson (2006) concluded that even though student grading was reliable, students still did not trust the validity of peer grading.

Liu and Carless (2006) distinguish peer assessment from peer feedback. They define *peer feedback* as a process of communicating "rich detailed comments" related to performance standards, and *peer assessment* as grading, whether or not feedback is provided (Liu and Carless, 2006, p. 280). In their view, feedback is a precursor to assessment and has greater potential for enhancing learning than assessment. They argued that prior research has too closely focused on peer assessment by over-emphasizing the reliability of peer measures, generally measured as the correlation of peer assessments with those of the instructor. They contend that prior research has proved the reliability of peer assessors and that the focus of research should shift to evaluating the merits of peer assessment in the learning processes.Liu and Carless (2006) based their conclusions in part on the work of Falchikov and Goldfinch (2000), who conducted a meta-analysis of forty-eight prior empirical studies of the reliability of peer assessments.

Falchikov and Goldfinch (2000) found high correlations between peer assessments and those of instructors, suggesting that peer assessments have high reliability. Reliability was high across various peer assessment contexts, including global versus dimensional assessments, subject area assessments, nature of task involved, class level, number of peer assessors and student involvement in identifying dimensions of assessment. Of particular interest to the current study are four findings. One, Falchikov and Goldfinch (2000) found no differences in reliability between introductory (year 1), intermediate (year 2) and advanced courses (year 3 and above). Two, business and management peers tend to over-mark peers relative to their instructors. Three, they found lower correlations when the number of peer assessors was large (more than twenty) than when the number of assessors was lower. Finally, they found greater reliability when peers were asked to provide a global assessment without specific criteria than when peers were asked to assess multiple dimensions or criteria.

2. *Purpose of the Study*

The Core Competency Framework of the American Institute of Certified Public Accountants (AICPA) stresses the importance of good communication skills in both the functional (technical) and personal (individual attributes and values) competencies desired of those entering public accounting (AICPA, 2005).

> Accounting professionals are called upon to communicate financial and non-financial information so that it is understood by individuals with diverse capabilities and interests. Individuals entering the accounting profession should have the skills necessary to give and exchange information within a meaningful context and with appropriate delivery. They should have the ability to listen, *deliver powerful presentations* (emphasis added) and produce examples of effective business writing (AICPA, 2003).

Lin, Grace, Krishnan and Gilsdorf (2010) found significant differences between the perceptions of accounting students about the importance of communication skills and those of accounting professionals. Accounting professionals value writing and speaking skills much more than accounting students. One way to close the gap between accounting students' and professionals' perceptions about the importance of communication skills might be to engage accounting students in providing peer feedback on communication skills. This idea would be consistent with Liu and Carless' (2006) argument that students who are engaged in peer feedback develop the self-awareness of their own strengths and weaknesses. It is also consistent with the findings of Kerr and Park (1995) and Cassidy (2006) that assessing one's peers provides students with meaningful work-place skills.

If peer feedback is to be useful in addressing the skills necessary for accounting students to be good communicators, it must include feedback on specific criteria that are associated with effective communication skills. This potentially creates a reliability problem if peer assessment is used for evaluative purposes, since peer assessment of specific criteria was found to be less reliable than global assessment (Falchikov and Goldfinch, 2000). The current study addresses this potential reliability problem in two ways. First, the sample included only graduate students. Though Falchikov and Goldfinch (2000) found no differences in reliability between introductory, intermediate and advanced courses, they defined advanced courses as those offered in year 3 and above. The students in this study are in their 5th year of college work at a minimum. Second, many of the students in this study had workplace experience—some in their first few years of employment but also others who had extensive workplace experience both in and outside of the accounting field. If these students had gained experience in evaluating peers and subordinates in the workplace, this may be reflected in higher reliability of peer assessments.

3. *Method*

The study was conducted in a graduate-level federal income tax class. Three semesters of data were collected from peer and instructor assessments of presentations of tax research cases. Research cases were assigned to teams of three or four students with a different student taking the lead for the research case. The cases were assigned such that each student had at least one opportunity to lead and present during the semester. The leader was responsible for presenting the results to the rest of the class and fielding questions following the presentation. Three to four cases were presented each week, with presentations lasting from twenty to thirty minutes each. A total of 57 students resulted in 68 observations. The first semester included data from 25 students, 3 of whom presented twice for a total of 28 observations. The second semester of data included 14 students, 8 of whom presented twice for a total of 22 observations. The third semester included data from 18 students, none of whom presented twice.The instructor provided the peer graders with a rubric that contained the criteria and evaluation-level definitions prior to the assessments. The rubric is presented in Figure 1.

The criteria were selected because the student was already assumed to possess an understanding of their meaning and use. Therefore, a student would not need to have a prior knowledge of federal income taxation during the presentation since only the presenter's skills, and not the material being presented was being assessed. At the end of each presentation, students were given time to complete the rubric with their assessment ratings and to provide written comments. To maintain anonymity, raters were asked not to sign the rubrics nor make any remark that might identify the rater. To assign a grade to the presentations, the categories "Poor", "Fair", "Good", and "Exemplary" were assigned numerical values of 65, 75, 85 and 100, respectively. Occasionally, a rater would mark a criterion as being between two levels. In those cases, grades of 70, 80, or 92 were assigned. The final assessment grade was computed by first averaging the grades assigned by peers and then averaging the peer grade with the instructor's grade.

Figure 1: Presentation Evaluation Rubric

Criteria	Poor	Fair	Good	Exemplary
Organization	Audience cannot understand presentation because there is no sequence of information.	Audience has difficulty following presentation because student jumps around.	Student presents information in logical sequence which audience can follow.	Student presents information in logical, interesting sequence which audience can follow.
Subject Knowledge	Student does not have grasp of information/student cannot answer questions about subject.	Student is uncomfortable with information and is able to answer only rudimentary questions. Some information is confusing, incorrect or flawed.	Student is at ease with expected answers to all questions, but fails to elaborate. Information is clear, appropriate, and correct.	Student demonstrates full knowledge (more than required) by answering all class questions with explanations and elaboration.
Authoritative Support	Student uses only secondary authority or misapplies/misinterprets primary authority.	Student uses some primary authority but fails to summarize in clear, succinct manner.	Student uses some primary authority, applied correctly and summarized clearly.	Student uses ample primary authority, applied correctly and summarized in a clear, concise, easy-to-understand manner.
Design	Screens are either confusing or cluttered or barren and stark.	Fonts are inconsistent, confusing and cannot be read from the back of the room.	Text size reinforces importance of concepts. Fonts are consistent and easy to read.	Text, color, contrast are combined to reinforce each other and add impact.
Verbal Skills	Student mumbles, incorrectly pronounces terms, and speaks too quietly for students in the back of class to hear. Speech has many pauses.	Student's voice is low. Student incorrectly pronounces terms. Audience members have difficult hearing presentation. Speech has several pauses.	Student's voice is clear. Student pronounces most words correctly. Most audience members can hear presentation. Speech has few pauses.	Student uses a clear voice and correct, precise pronunciation of terms so that all audience members can hear presentation. Speech is fluid.
Eye Contact	Student reads all of report with no eye contact.	Student occasionally uses eye contact, but still reads most of report.	Student maintains eye contact most of the time but frequently returns to notes.	Student maintains eye contact with audience, seldom returning to notes.

4. Results

Table 1 contains the mean and standard deviation for both peer- and instructor-assigned grades for each of the six grading criteria. For each criterion, the instructor's grade was lower than that assigned by peers, and with one exception, the standard deviations were larger. This result is consistent with Kerr and Park's (1995) conclusion that peer grades are generally higher than those assigned by instructors and with Falchikov and Goldfinch's (2000) findings that business and management students tend to overrate their peers relative to ratings assigned by instructors. Pearson correlation coefficients range from a low of 0.29 for "Design" to highs of 0.51 for "Verbal Presentation" and 0.56 for "Eye Contact." Although each of the correlations is significant, they are relatively low, a finding which suggests that peer-assigned grades may not be a suitable substitute for instructor-assigned grades.

Table 1: Mean, Standard Deviation and Correlations for Presentation Grading by Criterion

	Peer Grade		Instructor Grade			
Criterion	Mean	S.D.	Mean	S.D.	r	p value (2 tailed)
Organization	92.29	8.00	87.97	9.64	0.38	0.001
Subject Knowledge	91.23	8.23	87.53	8.62	0.37	0.002
Authoritative Support	92.75	8.16	88.35	8.92	0.32	0.007
Design	91.97	8.06	90.77	8.07	0.29	0.017
Verbal Presentation	91.38	8.53	88.89	7.64	0.51	<0.001
Eye Contact	91.61	8.65	88.13	8.77	0.56	<0.001

Figure 2 contains a graphical representation of the means summarized in Table 1. The direction of grades across the criteria is consistent between the instructor and peer with the exception of "Design." This suggests that the interaction effect between grader and criterion may be of interest in determining whether peer grades may be useful for feedback purposes. If this effect is significant, then presenters may not garner peer feedback that adequately mirrors the feedback they might receive from an instructor.
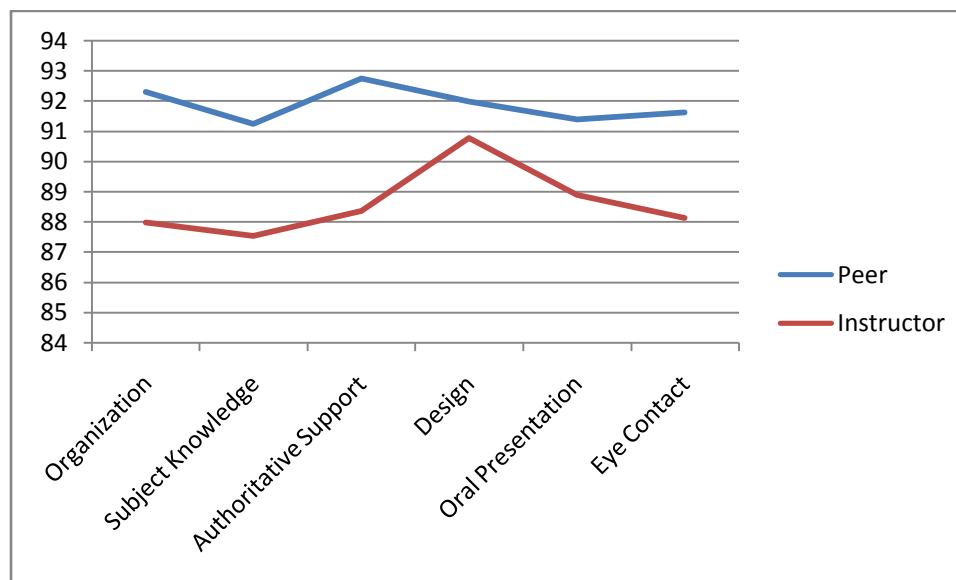
Figure 2: Graphical Representation of Means

Table 2 contains the results of an ANOVA. All of the primary variables were significant. The fact that grader was significant indicates that the instructor gave higher grades than did peers. The interactions of greatest interest were those between student and grader and between criterion and grader. The significant student-grader interaction indicates that the difference between the instructor's grade and that of the peers is greater for some students than for others. This confirms our prior suggestion that peer grades are not a satisfactory substitute for those assigned by an instructor. Though we had no means to test our results against those assigned by students in introductory- or intermediate-level courses, the results are consistent with those of Falchikov and Goldfinch (2000), who found low reliability of grades assigned by students in advanced courses.

A finding that interaction between criterion and grader was statistically significant would have implied that the difference between peer grades and the instructor's grades was greater for some criterion than others. The fact that it is not significant, therefore, may lend credence to the suggested use of peer evaluation as a feedback mechanism rather than as a grade substitute, especially considering the representations we see in Figure 1.

Table 2: Analysis of Variance for Presentation Grading

Source	Df	Sums of Squares	M^2	F Ratio	P
Student	67	60,225.50	89.89	15.44	<.0001
Criterion	5	1834.85	366.97	6.30	<.0001
Grader	1	4094.74	4094.74	70.31	<.0001
Student*Criterion	335	36816.19	109.90	1.89	<.0001
Student*Grader	67	10084.05	150.51	2.58	<.0001
Criterion*Grader	5	480.68	96.14	1.65	0.1430
Student*Criterion*Grader	334	14181.80	42.46	0.73	0.9999
Model	814	127717.81	156.90	2.69	<.0001
Error	6233	362980.16	58.24		
Total	7047	460697.97			

5. Discussion

We compared grades given by peers to those given by an instructor in a graduate federal income taxation class over three semesters. Our results confirmed prior findings that peer grades are significantly higher than those of the instructor and, therefore, that peer grades are not a good substitute for instructor grades even when students may have more subject knowledge and greater workplace experience. Experience was presumed in this study because the students were older and had more workplace experience in a professional field than may have been the case in prior research. Future research may help determine whether workplace experience in evaluating peers might close the gap between peer and instructor grading. The differences between student-assigned and instructor-assigned grades were generally uni-directional and did not significantly differ. This suggests that peer grades may be a useful feedback tool for either ungraded presentations or where practice presentations precede a graded one.

References

American Institute of Certified Public Accountants (2003). Core competency framework. Retrieved from http://www.aicpa-eca.org/library/ccf/competencies/default.asp

Cassidy, S. (2006). Developing employability skills: peer assessment in higher education. *Education + Training, 48*(7), 508-517.

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891-901.

Davis, N. T., Kumtepe, E. G. & Aydeniz, M. (2007). Fostering continuous improvement and learning through peer assessment: Part of an integral model of assessment. *Educational Assessment*, 12(2), 113-135.

Falchikov, N. & Goldfinch, J. (2000, Fall). Student peer assessment in higher education: A meta analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.

Farh, J. L., Cannella, Jr. A. A., & Bedeian, A. G. (1991, December). The impact of purpose on rating quality and user acceptance. *Group and Organization Studies*, 16(4), 367-386.

Henning, J. M., & Marty, M. C. (2008). A Practical Guide to Implementing Peer Assessment in Athletic Training Education. *Athletic Therapy Today*, 13(3), 30-33.

Kerr, P., & Park, K. (1995, July/Aug). Peer grading of essays in a principles of microeconomics course. *Journal of Education for Business*, 70(6), 357-361.

Liu, N. F. (2005). Hong Kong academics; and students; perceptions of assessment purposes and practices. *Learning Oriented Assessment Project report*, Hong Kong Institute of Education.

Liu, N., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. Teaching in Higher *Education, 11*(3), 279-290.

Papinczak, T., Young, L., Groves, M., & Haynes, M. (2007). An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Medical Teacher*, 29(5), e122-132.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.

van Hattum-Janssen, N., & Lourenço, J. M. (December 2006). Explicitness of criteria in peer assessment processes for first-year engineering students. *European Journal of Engineering Education*, 31(6), 683-691.

Walker, M., & Warhurst, C. (2000). 'In most classes you sit around very quietly at a table and get lectured at …': debates, assessment and student learning. *Teaching in Higher Education*, 5(1), 33-49.