



Research Article

No Need to Fear Peer Evaluation: an Alternate Model for Expert and Novice Evaluations in Business Studies

Ross Humby

Higher Colleges of Technology – Dubai, UAE

ross.humby@hct.ac.ae

Received date: 11 May 2013; Accepted date: 16 October 2013; Published date: 21 November 2013

Copyright © 2013. Ross Humby. Distributed under Creative Commons CC-BY 3.0

Abstract

This research was conducted to determine if there were consistent differentials, or patterns in the peer grading that could be utilized to further understand the potential impact on student grades with the use of summative peer evaluations in this presentation style of interactive assessment activity for first year business students. The data was reviewed in aggregate, and then on individual, booth by booth comparison. The evaluation grades from the peer or 'novice' evaluators were within the same range as those provided by the faculty or 'expert' evaluators. The research also introduced a variant on peer evaluation as the peer evaluators were senior year students three years ahead of their 'peers'. The available information on peer assessment studies targeting year one students is limited (Nulty, 2011) and this case study provides further evidence of the use of peer evaluation in first year of post-secondary.

Keywords: peer evaluation, business assessment, higher education

Introduction: Current Issues Forum

The Current Issues Forum (CIF) is an annual event at the DWC where teams of first year business students present their findings on topics relevant to the global business community. Each year the forum has a central theme and all research is focused on that theme. There are typically 3 to 5 students in each team and between 100 and 200 first year, first semester students participating in any given year. Duncan Nulty

(2011) has noted that the specific study of year one students in the peer assessment literature is very limited, and year or level may be an important criterion in the use of peer assessment. The students spend up to eight weeks researching the topics, preparing presentations and display materials for the booth. The CIF is scheduled past the mid-point in the semester in a trade show style format on campus. Each student team presents their findings in a multi-media style trade show display including a computer

based presentations. Both team and individual contributions are assessed based on a standard grading form. Only the expert grades were utilized in the determination of the student's final grade on the project assessment. The research was part of an effort to determine where and how peer evaluations might contribute to formal grading this activity in the future.

Peer Evaluation

Peer Evaluation has many different names, forms, definitions, purposes and applications in higher education literature. Some of the alternative appellations include 'peer review' (Madden, 2000), 'peer evaluation' (Greguras et al., 2001), 'peer appraisal' (Roberts, 2002), and 'peer support review' (Bingham and Ottewill, 2001). The process has been applied to the evaluation of faculty by other faculty, and to the evaluation of students by other students. One working definition as applied to students was provided by Pare and Joordens (2008, p.527) "peer assessment, sometimes called peer evaluation or peer review, is a process where peers review each other's work, usually along with, or in place of, an expert marker."

The benefits and limitations of peer review are well documented when this form of evaluation is applied to the student learning environment in higher education (Kremer and McGuinness, 1998; Falchikov and Goldfinch, 2000; Bloxham and West, 2004; Pare and Joordens, 2008; Liow, 2008; Cestone et al., 2008, Sondergard and Mulder, 2012). These benefits and limitations are not the focus of the research, but they provide context for the use of peer or novice evaluation in the learning process along with expert evaluation.

The definition of peer evaluation as provided by Pare and Joordens (2008, p.527) is the definition chosen for the purposes of this study, with one important variation. Much of the literature cited uses the term 'peer' in the student context to be a member of the same class, or course, where they are evaluating 'peers' involved in the same topics and

process (Cestone et al., 2008). In this study, the student peer evaluators (novice evaluators) were senior business students, evaluating the work of first year business students. This use of peer evaluation is somewhat different than that used in most peer review studies. It was organized in this format to reduce some of the potential limitations associated with peer review (Sondergaard and Mulder, 2102) and to provide evaluation opportunities for senior year students in the human resource discipline. This study also utilizes the terms 'expert' grader or evaluator (faculty and administrators) and 'novice' grader or evaluator (student peers) as described in Bloxham and West (2004).

Methods

The survey population is those Emirati students in the first year of the Applied Science Degree in Business at the HCT-Dubai Women's Campus during the 2010-2011 winter semester. Faculty, administrators (described as 'expert graders'), and a class of senior business students (described as 'novice graders') provided assessment feedback based on a structured grading sheet. The senior students were part of the Human Resource Management program specialization and have had some background learning in assessment and evaluation in the HR context. There were written instructions for completing the grading sheet which also contained the evaluation criteria. These were the same instructions and forms as received by faculty evaluators. There were four separate evaluation categories; Booth Display, Presentation, Current Issues Web Site, and the Individual Student Performance grades. The grading sheets were collected and tabulated using an Excel spreadsheet. There were a total of 44 student project booths in the evaluation process in 2011-12. A total population of 114 surveys was received with two surveys from 'guests' excluded, leaving 112 surveys in the data base. A breakdown by respondent group appears in table 1.

Table 1: Count of Assessments by Assessor Group

Replies	Assessor Group
23	Students
89	All Faculty and Administrators
12	Administrators
39	Faculty Female
38	Faculty Male

The mean and standard deviation were calculated for total grades allocated by assessor group (novices and experts) with the simple Excel AVERAGE and STDEV functions. In these calculations zero or null responses were excluded from the calculation of the means. A second level of comparison was then conducted on a booth by booth assessment, including a subsequent differential analysis. In these comparisons, only those booths that were completely assessed (having other than null or zero responses in all categories on the response sheet) with at least one expert and one novice response sheet included. This comparison involved the assessment of 13 booths. Again comparisons were made between the assessed scores by assessor category and finally by individual assessor.

Limitations

The data was collected only after the fact rather than having been designed and planned as part of a research methodology. As such there was no hypothesis being tested, rather this was an exploratory look into peer (re-defined as students not in the same class) review versus expert (faculty and administrators) evaluation in a presentation assessment for students in business.

The students performing the peer evaluation were part of a single, upper year class from a human resource course in performance management. They were not intimately familiar with the assigned project, or the preparation of the year one students. Many of the novice assessors chose to complete the evaluations with partners so the result recorded is actually a blended peer assessment. Further, there were only 23 novice evaluations collected.

Expert evaluators were drawn from a cross section of the campus including both faculty and administrators. There may have been faculty and administrative evaluators who were also not intimately familiar with the assigned project, or the preparation of the year one students.

Results

Comparison of Evaluations Aggregate Assessor Categories

The first set of comparisons examined the differences in grades assigned in each of the four assessment categories by novice, compared to expert (faculty and administrators) evaluation. The result of the data compilation is displayed in table 2.

Table 2: Scores by Project Segment for All Assessments by Aggregate Group

Novice Graders	Expert Graders	Evaluated Segment
21.4	21.6	Display (maximum 25)
20.3	21.6	Presentation (maximum25)
8.3	8.6	CIF Web site (maximum 10)
4.3	4.3	Individual Student Average (maximum 5)
54.3	56.2	TOTAL MEAN SCORE (maximum 60)
10.0	7.0	Standard Deviation (total scores)
23	89	n =

In Table 2, data shows that novices gave lower assessed values (total mean score) than the expert group of assessors. There was more variability or spread (standard deviation) in the total scores across the student group. These findings are consistent with some findings comparing peer and expert grading and different than others (Lawson, 2011; Liow2008).

Aggregate categories can sometimes hide significant but offsetting differences. As the

research was conducted to determine if there were consistent differences in peer evaluations, there needed to be a further analysis of the data by assessor group including a splitting out of the expert evaluators into faculty and administrator groups. Finally a third level of analysis was conducted, moving from the aggregate to the individual assessed project booth for each of the assessor categories.

Table 3: Detailed Assessment Results by Assessor Group

Novices	Administrators Only	All Faculty	ALL Faculty and Admin	Evaluated Segment
21.4	19.9	21.9	21.6	Display (maximum 25)
20.3	21.5	22.0	21.6	Presentation (maximum 25)
8.3	7.4	8.8	8.6	CIF Web site (maximum 10)
4.3	4.1	4.4	4.3	Individual Student Average (maximum5)
54.3	52.8	57.1	56.2	TOTAL MEAN SCORE (maximum 60)
10.0	7.7	7.2	7.0	Standard Deviation (total scores)
23	12	77	89	n =
83.6%	81.3%	87.8%	86.5%	Total Grade (%)
B+	B+	A-	A-	Letter Grade (DWC scale)

When the data is split out by assessor group there is some consistency in evaluation,

though there are additional findings to note on a percentage and letter grade basis.

Administrators evaluated the work with the lowest assessed values of any group. Novice evaluations were the next lowest. In terms of DWC letter grade, faculty assessor groups evaluated the work as 'A-' grade. Administrator's and Novice evaluator's mean grade was 'B+'.

Finding #1: In aggregate, novice or peer evaluations in this sample fell within the same range as the expert evaluators (faculty and administrators). This would mean there is little variation in potential impact on student grades if using this peer evaluation strategy.

The first sets of data were aggregate data using the results for all evaluations across the CIF assessments. But students were

assessed based on the work of each group's project booth. The research needed to compare the results at this level.

In this next section the data is compared with the different evaluations on the same booths. This level of analysis was conducted on a booth by booth basis comparing the assessment scores assigned by experts (faculty/administrators) and those assigned by peers. In these comparisons, only those booths that were completely and fully assessed (having other than null or zero responses in all categories on the response sheet) with at least one expert and one novice response sheet were included. This comparison involved the assessment of 13 project booths from the original population of 44.

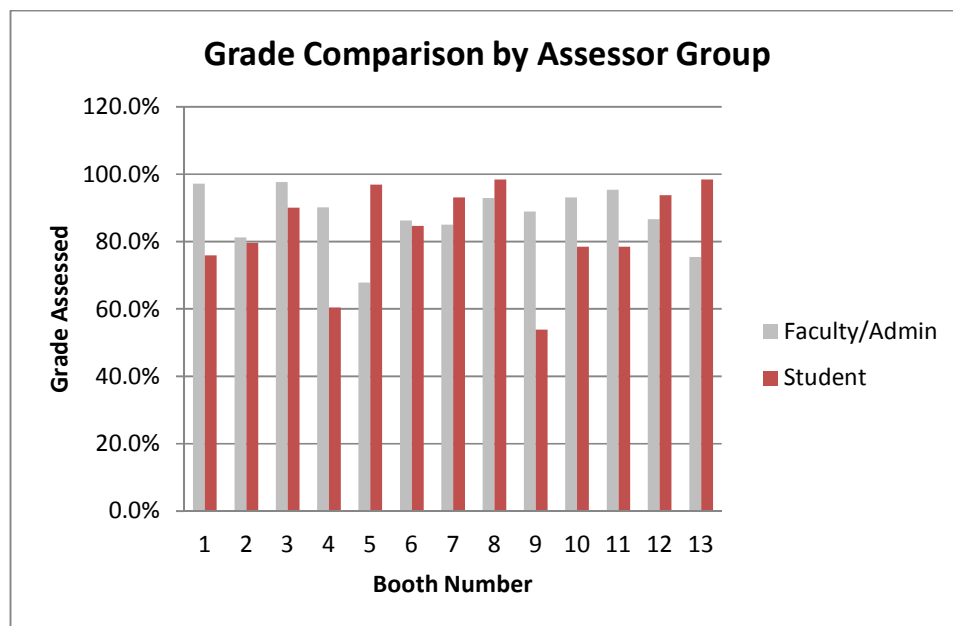


Figure 1: Grade Comparison by Assessor Group

In this instance there is far more variability, and potential impact on grades. Figure 1 represents the comparison of the differences in the total assessed score by experts compared to the total assessed score by novices. Figure 2 compares the value of the differences in the mean scores between

expert and novice evaluators. In the case of more than one evaluation in the data set for a booth, the mean of student peer assessments were deducted from the mean of expert assessed scores for that booth.

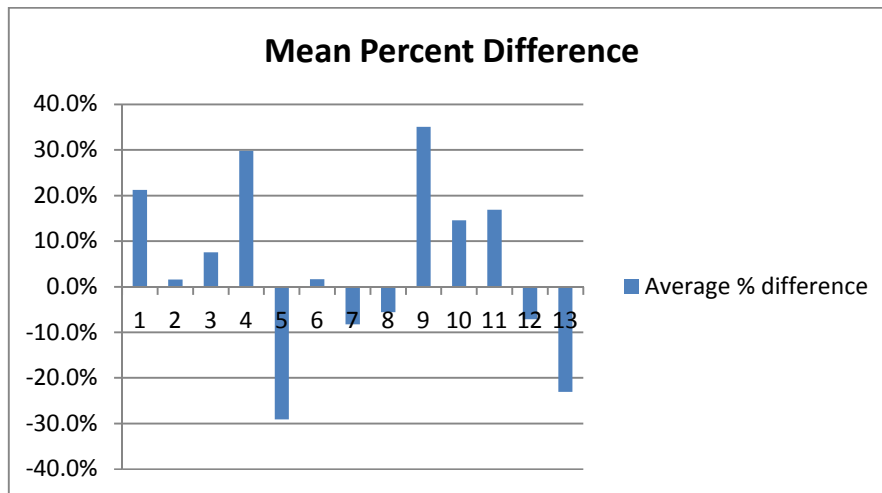


Figure 2: Assessed Grade Percentage Differences between Expert and Novice Graders on the same project booth.

Figure 2 is a graph showing the differences in assessed percentage grades between expert and novice grades. Of the 13 booths in this sample, only 6 were assessed within a plus/minus range of 10% between assessor groups. The highest single grading differential was +35%, and two other assessment differentials were +29.8% and -29.1% respectively.

Finding # 2: There were large value differences that existed on the evaluation of a specific booth between novice and expert assessors in this sample. The cumulative assessment means hid the potential grade

impact at the individual assessment level. This can have significant impact on student grades if this form of peer assessment is utilized.

The existence of very large percentage differences between evaluator groups means there is the potential for wide variances in final marks if peer review is included in the assessment strategy. A final analysis between the individual responses for each of the 13 project booths assessed was conducted to determine how the peer evaluations differed from the expert evaluations and to identify any patterns.

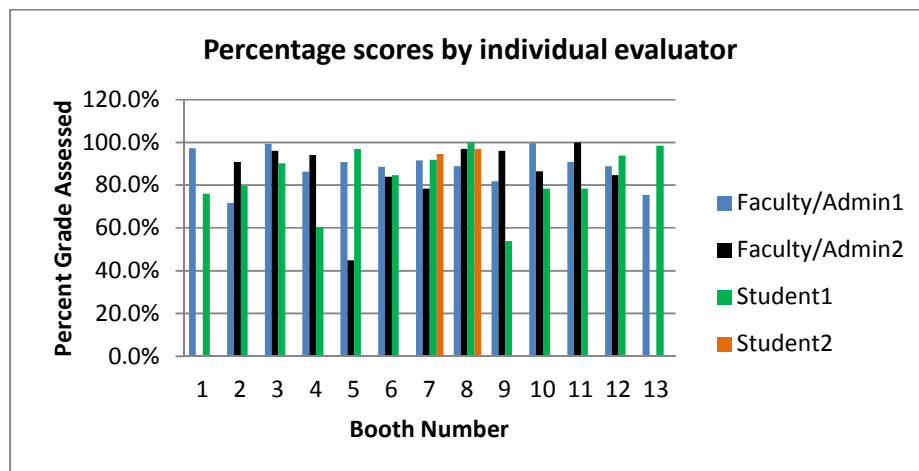


Figure 3: Percentage scores by individual evaluator.

The data in figure 3 shows there is some variability in the grades assessed by the individual assessors. In some cases the gaps between intra group assessors are larger than the inter group differences. The

percentage score differentials from each of the assessors were then calculated. These differences are presented numerically in table4.

Table 4: Intra and Inter Percentage Differentials by Booth, by Individual Assessor

Booth	F1 - S1	F2 -S1	F1- F2	F1-S2	F2- S2	S1- S2
1	21.2%					
2	-7.9%	11.2%	-19.1%			
3	9.2%	5.8%	3.4%			
4	25.9%	33.8%	-7.8%			
5	-6.2%	-52.0%	45.8%			
6	4.0%	-0.6%	4.6%			
7	-0.3%	-13.3%	13.1%	-3.1%	-16.2%	-2.8%
8	-11.1%	-3.1%	-8.0%	-8.0%	0.0%	3.1%
9	28.0%	42.2%	-14.2%			
10	21.2%	8.0%	13.2%			
11	12.3%	21.5%	-9.2%			
12	-5.0%	-9.2%	4.2%			
13	-23.1%					
F1 = Expert Assessor 1						
F2 = Expert Assessor 2						
S1 = Novice Assessor 1						
S2 = Novice Assessor 2						

Using an arbitrary +/- 10% as a reasonable difference in grading, a frequency count of the number of differences larger and less than 10% was calculated as a proxy for intra

group consistency. This analysis included only the 13 booth sub-sample and assessments from table 5.

Table 5: Count of Differentials in Grade between Assessors >10%

Count of Differentials on grading <10%			
	>10%	<10%	n=
Expert to Novice	14	14	28
Expert to Expert	5	6	11
Novice to Novice	0	2	2

Half of the expert to student assessments were within the 10% range, and 6 of the 11 expert to expert grades were within the 10% envelop.

Finding #3: There is as much consistency in the grades between the expert and the novice evaluators as there is between the different expert evaluators in this small sample result. The implications for using peer evaluations in grade calculation mean that there is likely to be little impact on the final grades assigned if this model of peer evaluation is included.

Conclusions

The decision to use peer evaluations as part of the assessment strategy could result in lower student grades in this case study based on the aggregate findings. However, the detailed analysis has yielded the range of variability between novice and expert evaluators to be similar, so in fact there is less likely to be any dramatically different grades if this model of peer evaluation is included.

There is no overall consistent pattern in peer assessment in this small sample size when compared to expert evaluation. The range in aggregate for peer evaluations lies within the range of the two groups of expert graders (faculty and administrators), and there is likely to be as much variability between expert evaluators as between expert and

novice evaluators. However, at the level of each individually assessed project booth, there is considerably more variability in the grades as assessed by peer and expert graders.

There are important indicators that could be the focus of further research to determine when and under what conditions peer evaluation makes the most sense in the assessment strategy. Finally, the use of upper year students to perform summative peer evaluation on first year student work yields result consistent with 'expert' grades. This means that the benefits associated with peer evaluation strategies are retained, and some of the limitations (student collusion on grades, or a tendency to grade friends higher) are greatly reduced.

References

- Cestone, C., Levine, R., & Laine, D. (2008). Peer assessment and evaluation in team based learning. *New Directions for Teaching and Learning*, 116 (Winter), 69-78.
- Bingham, R., & Otterwill, R. (2001). Whatever happened to peer review? Revitalising the contribution of tutors to course evaluation. *Quality Assurance in Education*, 9(1), 32-39.
- Blackmore, J. (2005). A critical evaluation of peer review via teaching observation within higher education. *International Journal of Education Management*, 19(3), 218-232.

- Bloxham, S., & West, A. (2004). Understanding the rules of the game: marking peer assessment as a medium for developing students' conceptions of assessment. *Assessment and Evaluation in Higher Education*, 29(6), 721-733.
- Falchikov, N., Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.
- Goldberg, L., Parham, D., Coufal, K., Maeda, M., Scudder, R., Sechtem, P. (2010). Peer review: The importance of education for best practice. *Journal of Teaching and Learning*, 7(2), 71-84.
- Greguras, G., Robie, C., & Born, M. (2001). Applying the social relations model to self and peer evaluations. *Journal of Management Development*, 20(6), 508-525.
- Kremer, J., & McGuinness, C. (1998). Cutting the cord: student-led discussion groups in higher education. *Education + Training*, 40(2), 44-49.
- Lawson, G. (2011). Students grading students: a peer review system for entrepreneurship, business and marketing presentations. *Global Education*, 2011(1), 16-22.
- Liow, J.-L. (2008). Peer assessment in thesis presentation. *European Journal of Engineering Education*, 33(5-6), 525-537.
- Madden, A. (2000). When did peer review become anonymous? *Aslib Proceedings*, 52(9), 273-276.
- Nulty, D. (2011). Peer and self-assessment in the first year of university. *Assessment and Evaluation in Higher Education*, 36(5), 493-507.
- Pare, D., & Joordens, S. (2007). Peering into large lecture: examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning*, 24, 526-540.
- Roberts, G. (2002). Employee performance appraisal system participation: a technique that works. *Public Personnel Management*, 31(3), 119-125.
- Sondergaard, H., Mulder, R. (2012). Collaborative learning through formative peer review: pedagogy, programs and potential. *Computer Science Education*, 22(4).