Routledge
Taylor & Francis Group

Student involvement in assessment: a project designed to assess class participation fairly and reliably

Diane Dancer* and Patty Kamvounias
University of Sydney, Australia

This article describes a project designed to ensure that class participation in a large introductory commercial law course is assessed fairly and reliably. The subjectivity often associated with this type of assessment is minimized by involving students in the specification of clear criteria and the assessment process as they were asked to assess themselves and their peers. Formative feedback is given mid-way through the semester so that students have the opportunity to take remedial measures where necessary and teachers can reinforce positive behaviour. The data are analyzed to determine the relationships between the assessments given by the students, their peers and tutors. There is evidence of some gender bias in Week 13 that is not consistent with the final marks. Finally, the combining of the individual assessment of the criteria is shown to be a good estimate of the final class participation mark given by the tutors and the students.

Introduction

Class participation is a common requirement of many university courses. Teachers may include class participation in their courses as an important teaching strategy because students actively involved in small group discussions are more likely to understand course material than if it were presented to them while they were simply sitting in a classroom (Ramsden, 1992). However, class participation may be required without necessarily including it in the course assessment. If one of the objectives of a course is to facilitate the development of certain skills, then those skills should be taught and should be assessed because course objectives and assessment tasks should always be linked (Rowntree, 1987). Kamvounias (1996) explores the problems and issues involved in assessing class participation.

Nightingale *et al.* (1996) shift the focus from assessing knowledge to assessing learning and Heywood (1989) states the importance of assessment, learning, teaching

---

*Corresponding author. Econometrics & Business Statistics, Faculty of Economics and Business, University of Sydney, Sydney NSW 2006, Australia. Email: d.dancer@econ.usyd.edu.au

and evaluation all being linked to objectives. Assessing class participation encourages students to develop their oral communication skills and to demonstrate other skills such as interacting and cooperating with their peers and with their tutor. Le Brun and Johnstone (1994) suggest that students' attitudes and values can be affected by class participation as well as increasing their motivation and emphasizing the individual's responsibility for learning.

Biggs (1999) advocates a theory of constructive alignment that supports the view that education is about conceptual changes not merely the acquisition of knowledge. This theory implies that meaning does not come from direct transmission of knowledge but is created by the student's learning activities. One consequence of this theory is that, if students are engaged in active learning, then the less-able students are more likely to succeed. Cardozo (1994) illustrates how the group can become engaged in learning by, for instance, students reading aloud their written work, other students listening carefully and, as a consequence, lively discussion ensuing.

The discussion of class participation and its assessment is broadened by the work of Vandrick (2000). She notes that class participation requires students to speak in class by asking and answering questions, making comments and participating in discussions. Further, she notes that students who do not participate in this way may be considered to be 'lazy, unprepared, passive and/or uninvolved students, and are generally penalized when class grades are assigned' (Vandrick, 2000, p. 2). Vandrick (2000) suggests that some of the reasons for non-participation are shyness, classroom dynamics (domineering students), cultural (in some cultures, speaking is seen as challenging the professor) and the language barrier (students feel insecure with their grasp of the language).

In the work by Hayes and Hayes (1973), the variables that previously had been taken into account by the lecturer when 'subjectively' allocating marks for class performance are formalized and used by the lecturer to create a rating form. The study concludes that the student and lecturer ratings on the same item tend to correlate highly indicating that both interpret the items in the same way and that students are able to estimate accurately their own performance on each of the variables (Hayes, 1973, p. 327).

Lyons (1989, p. 37) argues that student participation in the development of the criteria for assessment should result in greater acceptance of the assessments made by the tutors and that developing the criteria is itself a learning experience because it requires students to face the question of what constitutes 'good' or 'bad' performance. Orsmond *et al.* (2000) undertake a study on the use of student derived marking criteria. Some of the conclusions drawn from this study are that students may be less able to discriminate between criteria given to them and those constructed by them and that construction of the criteria does not increase agreement between student and tutor.

There are a number of studies that note that feedback is associated with more effective learning (see Bangert-Drowns *et al.*, 1991). However, Teekell (1990) finds that structured feedback from the teacher is highly beneficial for weak students but detrimental for strong students. Dweck and Bush (1976) and Henry (1979) conclude that males and females react differently to feedback from adults and their peers as well as to positive and negative feedback. Butler (2004) advocates formative rather than normative feedback because students need to reflect on their performance and to take responsibility for their own actions.

This article seeks to address the findings of previous research and to codify the method of assessment of class participation. The aims of this study are fourfold:

1. To involve students and tutors in the development of criteria for the assessment of class participation.
2. In addition to tutor assessment, students would assess themselves and their peers.
3. Formative feedback would be given mid-way through the semester to the students.
4. The data was to be analyzed for gender bias and to model the relationship between the criteria marks and the overall class participation mark.

Method

This project on assessing class participation was developed for *Commercial Transactions*—a one semester first year introductory commercial law course offered to students undertaking non-law degrees. The assessment was designed to give students formative feedback on their class participation midway through the semester and normative feedback at the end of the course. For assessment to be effective, teachers must give feedback to students on their work so that students have the opportunity to take remedial measures where necessary and teachers can reinforce positive behaviour. The past practice of issuing class participation marks at the end of the course did nothing to improve the skills of the students if they discovered too late that their performance was not satisfactory. Clarke (1985) suggested weekly written feedback but this is not practicable for large undergraduate classes.

Development of criteria

The course outline stated that: 'Although attendance at tutorials is compulsory, marks are not awarded for mere attendance—participation in the discussion in all tutorials is the basis for the class participation mark'. The absence of clear criteria in previous years may have resulted in this mark simply assessing a student's record of attendance at tutorials and not really assessing 'participation' at all.

Students were to be involved in the development of the criteria to help them learn about teacher expectations for this aspect of assessment. It is a commonly held view in the literature on peer assessment that students should have some input into the criteria (Searby & Ewers, 1997). Tutors and students were then to assess class participation according to the set criteria. Students would not be awarded their self assessed mark or their peer mark unless it corresponded with their teacher's mark.

The criteria were developed in three steps.

Step 1

During the first tutorial in Week 2 of the semester, tutors explained why class participation was an important component of the course and held a general discussion about class participation with all their students. All students were asked to write down

their ideas for at least three criteria for assessing class participation, which they submitted anonymously to their tutors.

Step 2

After reviewing all the student and tutor comments and suggested criteria, a draft list of five comprehensive criteria for assessment of class participation was drawn up and distributed to the tutors for further comment. Note that there were over 40 suggested criteria which were collated into the five groups.

Step 3

The final list of criteria was circulated to tutors and students during the tutorial in Week 4. Assessment of class participation started once the students knew the criteria for assessment in Week 4. See Dancer and Kamvounias (2004) for more details of the implementation of the project.

The final criteria for assessment of class participation were as follows:

1. Preparation: the extent of your reading, analyzing and understanding of the material, demonstrated by contribution to discussion.
2. Contribution to discussion: the extent to which you volunteered answers, asked relevant questions, expressed your own opinion and analyzed contributions of others.
3. Group skills: the extent to which you allowed others to contribute, avoided class domination, shared ideas with others, assisted others, provided positive feedback to others and exhibited tolerance and respect for others.
4. Communication skills: the quality of your expression, clarity, conciseness, use of appropriate vocabulary, confidence.
5. Attendance: includes punctuality.

Assessment

The assessment form listed each criterion together with a five-point rating scale—very good, good, average, fair or poor. The rating was to be indicated by a simple tick in the appropriate column on the form. No particular weighting was given to each criterion but 'Attendance' was listed last to indicate that it was the least important of the five criteria. Attendance was also rated on a five-point scale as it included punctuality and not just the number of attendances.

In Week 7, students were asked to rate themselves on each of the five criteria, to add comments on their performance in tutorials and to give themselves a mark out of ten for their overall performance in tutorials. The tutor commented briefly (but gave no mark) about the student's performance to date and about the student's assessment of their own performance. The forms were returned to students. Tutors also rated each student on each of the five criteria as well as awarding a mark out of ten. This mark was not revealed to the students as the assessment was of a formative nature.

Butler (2004) advocates 'comments only' marking as it encourages students to reflect on their performance and take responsibility for their own progress.

In Week 13, students reassessed themselves as before but also assessed each of their peers in the group using the same criteria. Students were not being asked to determine the grade for their peers; peer marks were merely to be used to check the reliability of the tutors' marks. After the final tutorial, tutors assessed each student to determine their final grade by taking into account the listed criteria and their own interim assessment of the student's performance. Students were notified of the tutor's grade at the end of Week 14.

Results and analysis

There were 296 students who completed the unit, *Commercial Transactions*. The following data were collected for each student: Week 7—mark by self and tutor; and Week 13—mark by self, tutor and peers. The peer mark for each student was the average of all the peer marks. Students were only included in the analysis if there were data for all five criteria for Weeks 7 and 13, class participation marks and all relevant assessment marks for the course. Only 246 students were included in the final analysis. The average result for the 246 students was significantly higher than those for the excluded group for each set of available data, implying that there may be sample selection bias if the results are used to make inferences about the whole population.

Similar to findings in Burchfield and Sappington (1999), the students rated themselves higher than their instructor did, the average peer mark was consistent with the instructor mark and there was a high correlation between instructor and peer evaluations. Thus the reliability of tutor's marks was established.

The results also indicate that formative feedback may have had a significant impact on the students' grades as the t-statistic of 6.82 is very large indicating a very significant difference between the Week 7 and Week 13 self-assessed marks. It is not possible to be more precise on the impact of the feedback from the tutors in Week 8 but possibly this difference may have been reduced if the students had received a grade as well as written feedback.

Gender differences

A further consideration is the possibility of a gender bias. Do males receive higher class participation marks from tutors than females? If they do receive a higher mark for class participation, is this reflected in their final mark? The average final mark for males is 55.9 and for females 57.2. This difference however is not significant (p-value = 0.32 for a two-tail test). The results in Table 1 use the overall class participation mark not the individual criterion mark. The results also assume that there is no difference between males and females, when the overall class participation marks are analyzed for gender bias. In each case, males are scoring higher on average than females suggesting that there could be some gender bias. The only two differences that are significant at the 5% level of significance are for Week 7 Self marks and for

Table 1.   Results of testing for gender bias

	Mean—males	Mean—females	Pooled standard deviation	p-value (2 tail test)
Week 7 Self	7.33	6.84	1.24	0.002***
Week 13 Self	7.23	6.76	1.55	0.080*
Week 7 Tutor	7.57	7.31	1.13	0.062*
Week 13 Tutor	7.46	6.98	1.38	0.014**
Week 13 Peer	7.33	7.17	0.79	0.182

Note: *** Coefficient is significant at 0.01 level (2-tailed); ** Coefficient is significant at 0.05 level (2-tailed); * Coefficient is significant at 0.10 level (2-tailed)

Week 13 tutor marks. Note that this outcome of males performing better than females is not demonstrated in their final marks.

One result that is of concern is that the difference in Week 13 of the tutor's mark between males and females is so large. Unfortunately, there is no data to indicate any reasons for this difference. However, these differences may reflect actual ability as the class participation component is testing oral communication. The tutor's assessment suggests that males are communicating more than females. Anecdotal evidence from the tutors suggests that at least some of the female students do not participate orally for various reasons (see Vandrick, 2000). As a consequence these students will receive lower marks even though these students may be competent to answer the questions.

Relationship between the assessment criteria and the overall class participation mark for each student

Is the overall class participation mark related to the five assessment criteria? Students and tutors were given no guidance about how to combine the assessment of the criteria to form the mark out of 10. We believe that there is a strong relationship between the criteria assessments and the overall class participation mark given in both Weeks 7 and 13 by tutors and students. One problem with analyzing the individual assessment for the criteria is that the options were: poor, fair, average, good and very good. These results were coded as 1, 2, 3, 4 and 5. The values of the codes suggest that there is an equal distance between each of the options. Ordinary least squares (OLS) is used to analyze these data, but this method implies that the difference between 'poor' and 'fair' is the same as between 'good' and 'very good' which may not be true. However, since the interest here centres on how well the criteria predict the class participation mark, OLS is a satisfactory method.

The model used is:

$$mark = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 C_3 + \beta_4 C_4 + \beta_5 C_5 + u_i$$

where C_1 = Preparation, C_2 = Contribution to Discussion, C_3 = Group Skills, C_4 = Communication Skills and C_5 = Attendance and u_i is the disturbance term
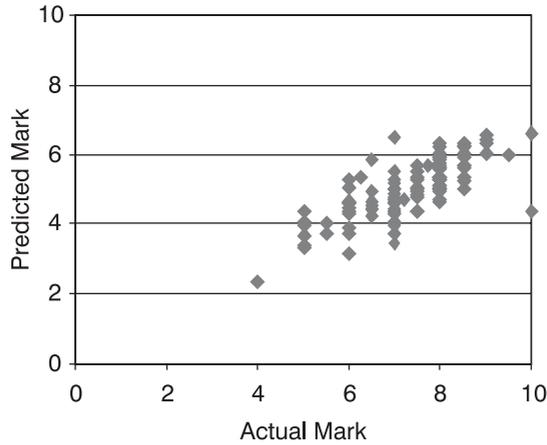
Figure 1.   Actual and predicted self-assessed mark for Week 13—females

The results in Table 2 are for Week 13 only as the Week 7 analyses gave very similar results. The models for the tutor give good estimates as evident from the high R^2 for both males and females. For example, if a student was given a 'very good' on Criterion 1–3 and 'good' on Criterion 4 and 5, the predicted mark for a male would be 8.69 but for a female 8.73. If the prediction uses the self coefficients then for a male it is 8.53 and, for the female, 8.28. For males, the correlations between the actual mark and the predicted mark are 0.83 for the self-assessed marks and 0.92 for the tutor marks. For females, the correlations are 0.77 and 0.91 respectively. Figures 1 and 2 give a clear picture of the strong relationships between the actual and predicted marks for females. The results are very similar for males. As is always the case with regression analysis (see Dancer & Tremayne, in press), the predicted marks over-estimate the low marks and under-estimate the high marks.

Conclusion

This project resulted in the identification of criteria for assessing class participation using both student and tutor input. The assessment criteria were clearly stated on the

Table 2.   Results by gender for the student's mark and the tutor's mark

FEMALE	Constant	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5	R^2
Self	2.54***	0.13	0.57***	0.08	0.39***	0.12	0.59
Tutor	2.10***	0.29***	0.29***	0.45***	0.29***	0.08	0.83
MALE							
Self	2.22***	0.16	0.53***	0.02	0.53***	0.16*	0.69
Tutor	2.47***	0.36***	0.30***	0.28***	0.24***	0.14*	0.84

Note: *** Coefficient is significant at 0.01 level (2-tailed); * Coefficient is significant at 0.10 level (2-tailed)
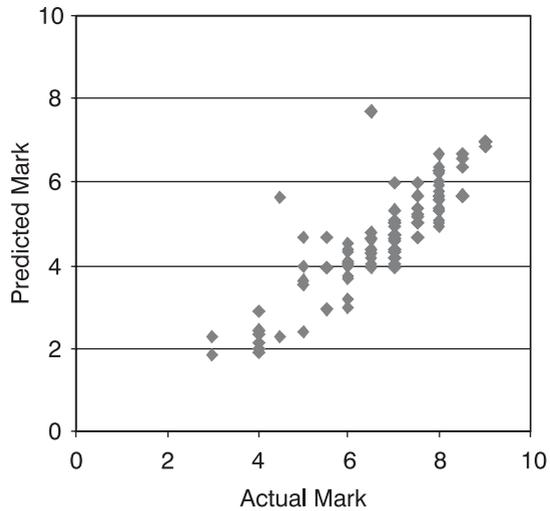
Figure 2.   Actual and predicted tutor mark for Week 13—females

class participation assessment forms so students knew exactly what their tutors expected and how they were being assessed. The assessment criteria and the self assessment forms developed as part of this project have been used in subsequent semesters to assess class participation in *Commercial Transactions*. As the assessment of class participation is, to some extent, the assessment of the generic skills of oral communication and group work, it follows that there should be some basic criteria by which those skills can be evaluated irrespective of the subject matter of the course. These criteria for assessment of class participation are as relevant in a commercial law course as they are in many university courses.

There is evidence of significant gender bias with males receiving higher class partic- ipation marks than females for which there is no justification as indicated by the final grades. There is strong evidence (Table 1) that tutors in Week 13 were awarding higher marks to the male students than the female students. This may be that the male students participated more than the female students. It is also evident that these higher marks for class participation, on the average, do not translate into higher final marks. The results in Table 2 indicate that there is a strong relationship between the criteria assessments and the overall class participation mark given in both Weeks 7 and 13 by tutors and students which implies that both students and tutors were able to reliably combine the criteria.

The evidence from the peer assessment results also suggests that, when explicit criteria are made known, overall the tutors' assessment of class participation is reliable and so there may be no need for peer assessment. However, these results also strongly indicate that the groups of people who may not feel empowered to speak (e.g., females) may need to be addressed (Vandrick, 2000). This issue needs further research to be confident that some groups of students are not disadvantaged by this form of assessment.

Distributing the self-assessment forms to students and requiring them to complete them on two occasions indicated to students that they must take their performance in class seriously. Self-assessment was therefore very worthwhile as was the opportunity for students to receive their tutors' comments during the course. As a result of this project, the self-assessment forms and the mid-semester tutor feedback became standard practice in *Commercial Transactions*.

Some reflections on where the project could be improved are necessary. There was no control on the quality of feedback given to students. Natriello (1987) suggested that the quality of feedback is important. Perhaps more definite instructions to tutors on how to give helpful feedback would be useful to ensure all students receive good quality feedback. Another aspect where instructions were not given to students and tutors was how to combine the criteria to form the overall class participation mark out of ten. However, the analysis suggests that the overall class participation marks were consistent with the criteria. A more difficult problem concerns students not participating fully for different reasons. This is difficult to overcome if the class participation is measuring oral communication. In other courses, some of the suggestions in Vandrick (2000) could be considered.

Acknowledgement

Notes on contributors

Diane Dancer is a lecturer whose areas of research and interest are in economics education research, limited dependent variables, applied econometric modelling, development economics and demography. One area of research is to model the performance, progression and time management of first year students. Econometric techniques are used to predict educational outcomes and the implications for universities with respect to policy issues. Other areas of research involve investigating schooling outcomes and gender issues for Egypt and modelling the impact of corporate women on the glass ceiling.

Patty Kamvounias teaches a variety of commercial law subjects within the School of Business at the University of Sydney. Her main research areas and interests are in competition law, consumer law and legal education.

References

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A. & Morgan, M. T. (1991) The instructional effect of feedback in test-like events, *Review of Educational Research,* 61, 213–238.

Biggs, J. (1999) *Teaching for quality learning at university* (Buckingham, Open University Press).

Burchfield, C. M. & Sappington, J. (1999) Participation in classroom discussions, *Teaching of Psychology,* 26(4), 290–291.

Butler, S. (2004) Question: when is a comment not worth the paper it's written on? Answer: when it is accompanied by a level, grade or mark! *Teaching History,* 115, 37–41.

Clarke, E. G. (1985) Grading seminar performance, *College Teacher,* 33, 129–133.

Cardozo, L. F. (1994) Getting a word in edgeways: does 'not talking' mean 'not learning'? *TESOL Journal,* 4(1), 24–27.

Dancer, D. & Kamvounias, P. (2004) Student involvement in assessment: a project designed to assess class participation fairly and reliably. Report number: ECMT2004-1 (Sydney, Sydney University, School of Economics and Political Science). Available online at: www.econ.usyd.edu.au/content.php?pageid=13579 (accessed 4 August 2004).

Dancer, D. & Tremayne, A. (in press) R-squared and prediction in regression with ordered quantitative response, *Journal of Applied Statistics.*

Dweck, C. S. & Bush, E. S. (1976) Sex differences in learned helplessness: differential debilitation with peer and adult evaluators, *Developmental Psychology,* 12(2), 147–156.

Hayes, S. L. & Hayes, R. A. (1973) Towards objective assessment of class participation, *The Journal of the Society of the Public Teachers of Law,* 12, 323–332.

Henry, S. E. (1979) Sex and locus of control as determinants of children's responses to peer versus adult praise, *Journal of Educational Psychology,* 71(5), 605–612.

Heywood, J. (1989) *Assessment in higher education* (London, John Wiley).

Kamvounias, P. (1996) Assessment of class participation in Commercial Law courses, *Current Commercial Law,* 4(2), 128–139.

Le Brun, M. & Johnstone, R. (1994) The quiet (r)evolution: Improving student learning in law (Sydney, Law Book Company).

Lyons, P. R. (1989) Assessing classroom participation, *College Teaching,* 37(1), 36–38.

Natriello, G. (1987) The impact of evaluation processes on students, *Educational Psychologist,* 22, 155–175.

Nightingale, P., Te Wiata, I., Toohey, S., Ryan, G., Hughes, C. & Magin, D. (1996) *Assessing learning in universities* (Kensington, NSW, University of New South Wales).

Orsmond, P., Merry, S. & Reiling, K. (2000) The use of student derived marking criteria in peer and self-assessment, *Assessment and Evaluation in Higher Education,* 25(1), 23–38.

Ramsden, P. (1992) *Learning to teach in higher education* (London, Routledge).

Rowntree, D. (1987) *Assessing students: how shall we know them?* (London, Harper & Row).

Searby, M. & Ewers, T. (1997) An evaluation of the use of peer assessment in higher education: a case study in the School of Music, Kingston University, *Assessment and Evaluation in Higher Education,* 22(4), 371–383.

Teekell, D. M. (1990) Effectiveness of peer feedback in helping skills training of nonprofessionals, *Masters Abstracts International,* 28, p. 164.

Vandrick, S. (2000) Language, culture, class, gender and class participation, paper presented at *TESOL Annual International Convention,* Vancouver, Canada.