

Making assessment transparent: the use of grade criteria combined with peer marking for assessment tasks in practical exercises in biochemistry

Terry Bartlett, Raul Sutton, Matthew Bellamy,
Daron Fincham, Chris Perry
School of Applied Sciences

Background and rationale

The overall aim of this project was to develop a series of grade criteria and detailed marking schemes to engage students in peer marking exercises in Biochemistry practicals. The robustness of this form of assessment was tested by comparison of student and staff marked assignments. A statistical comparison of student performance on this and another module was also used to determine whether peer marking was successful in enhancing student performance.

The objectives were:

- To develop grade descriptors for all assessment criteria used in assessed practical classes.
- To develop a series of marking schemes (in the form of assessment grids containing grade point related criteria) for practical reports.
- To provide students with these schemes after completion of reports for peer marking.
- To compare marks from these peer assessment exercises with a series of staff generated marks in order to test the rigor with which students assess each other's work.
- To monitor changes in performance by comparison of grades for all biochemistry practicals in this cohort with grades obtained for a previous cohort, and to relate these changes to grades obtained in a similar control module.

The innovation

In the first year, practical exercises in Biochemistry (module code BC1001) involve the completion of four assessment tasks. Currently, these are assessed by academic staff using defined criteria. Students are provided with threshold grade descriptors for borderline pass and first class equivalent for each of the assessment criteria that are used in the task. Feedback is currently provided using a form indicating where marks for the work have been allocated. There is some anecdotal evidence that students do not gain maximum benefit from assessments marked in this way, and are not reflecting upon (and thus improving) their own performance as a result. In addition, assessment of practical exercises in this way represents a large investment in staff time because of the extensive marking of students prepared reports.

The format we developed for marking assessments on these modules was highly innovative, with students being asked to peer assess their work. Peer assessment can be highly beneficial to student learning outcomes for the following reasons (Race, 2001):

- Deepening student learning
- Allowing access to the assessment culture
- Developing life-long learning skills in the students
- Increasing feedback to students

To achieve these aims, detailed assessment criteria were developed and marking schemes for peer assessment (see appendix 1). In constructing these criteria considerable care was taken to attempt to define a marking scheme that a student could be reasonably expected to make a judgement about. Thus, for example, judgements based upon a student's detailed analysis of biochemical theory were avoided in favour of questions with clear answers and unambiguous responses. The students were furnished with the criteria as a part of their assessment pack for the practical exercises, and after completion of assignments detailed marking schemes were provided for peer marking. This process was carried out for the first two practicals in the module. The final two were assessed using minimal grade criteria and staff marking as before.

The purpose of this project was to assess whether peer marking was an effective means of providing meaningful objective assessment and feedback for the students. The results from this study will be compared with a cohort from a previous year, and results from another core module will be used to verify the effectiveness of peer marking on improving student learning outcomes.

The results

Preliminary analysis of 2001 cohort data confirmed anecdotal evidence that students do not benefit by preparing reports based on minimal criteria with traditional methods of assessment and feedback (see figure 1).

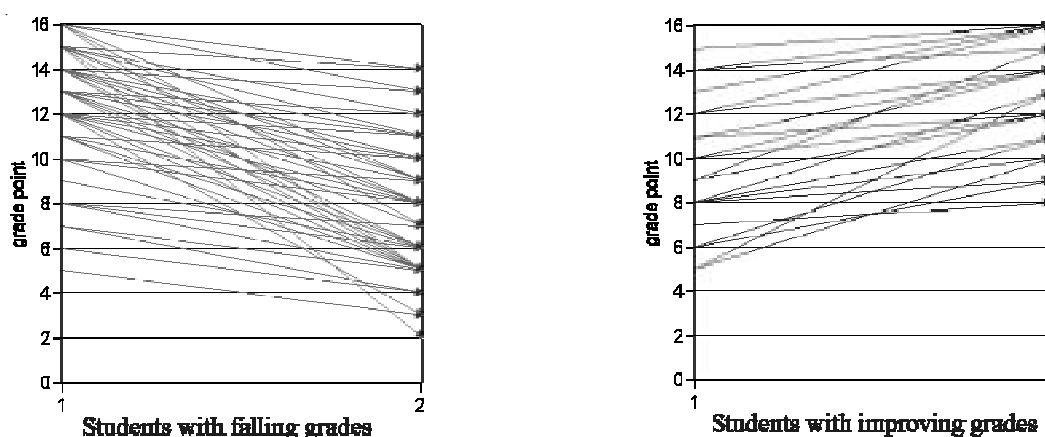


Figure 1. Comparison of student performance between practicals 1 and 2 in biochemistry.

The left hand panel shows students with a drop in grade point between practical 1 and 2 ($n=99$). The right hand panel shows students with increasing grades ($n=37$). For purposes of clarity, grades for students which remained the same ($n=22$) have been omitted.

As can be seen, grades for approximately two-thirds of students (62.7%) fall. Grades for students which improve (23.4%) show a smaller rise than the decrease observed when grades fall. The overall drop in grades was $4.27 (\pm 2.67)$ in grade point, as compared with a mean increase of $2.57 (\pm 1.64)$ amongst students whose grades improve. This showed the drop in grades was a statistically significant difference at the 5% level.

Preliminary analysis of the grades for peer marked assessments would appear to show little if any improvement (see figure 2). Grades for 88 students drop (65.7%), whilst only 30 students showed an improvement in grade (22.4%), with 16 (11.9%) showing no change. Once again, average drop in grade was higher (3.01 ± 2.06) compared to average improvement (2.83 ± 2.13) but the difference in this case is not statistically significant.

Whilst these preliminary findings do not appear to be encouraging, it must be borne in mind that potential benefits could be more long term. It was always the intention of this research to monitor the overall performance of the student cohorts, including grades for

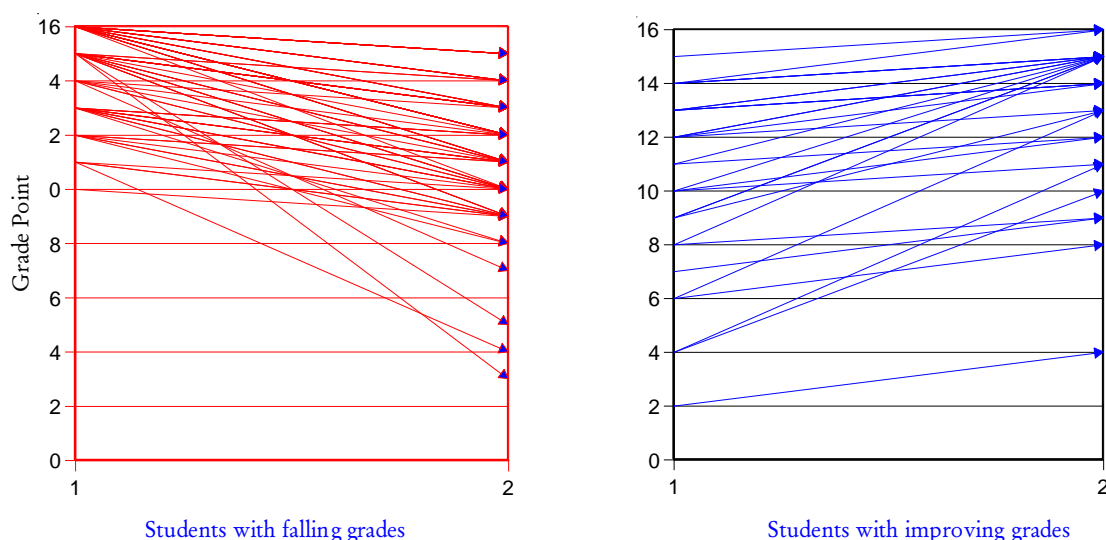


Figure 2. Comparison of student performance between practicals 1 and 2 when peer marked.

The left hand panel shows students with a drop in grade point between practical 1 and 2 ($n=88$). The right hand panel shows students with increasing grades ($n=30$). For purposes of clarity, grades for students which remained the same ($n=16$) have been omitted.

staff marked practicals, in order to determine the effectiveness of peer marking within the whole module.

The design of the experimental approach means that an attempt has been made to compare similar cohorts of students and hypothesise whether there are statistically significant differences observed in student achievement between cohorts who had experienced peer-marking of practicals and those who had tutor-marked practicals. Thus, there had to be tests to show that the student population was equivalent as well as a control for whether there was a significant difference between peer and tutor assessment of practicals.

The regime used was to compare the mean scores of the practical work of the student population in a companion module to BC1001. The module chosen to act as a control was the semester 2 iteration of a similar module (Cell Biology and Genetics, module code AB1102). The mean practical grades for AB1102 in 2001/2 (when BC1001 was tutor-marked) and 2002/3 (when BC1001 was peer-marked) were tested for differences using a t-test. The mean grade for 2001/2 iteration of AB1102, the control module (96 observations) and 2002/3 (53 observations) was found to be 10.49 and 10.96 respectively. The P value (43%) for the two tailed test shows that the two data sets are not significantly different at the 5% level. This means that in the control module, the two populations are the same.

Analysis of the grades obtained in for BC1001 shows that the cohort mean grades for 2001/2 (143 observations) and 2002/3 (134 observations) were found to be 11.42 and 13.36, respectively. The P value calculated for these data sets (1.97%) was found to be significant at the 5% level, suggesting that the 2002/3 cohort performance was significantly better than the 2001/2 cohort and that the differences have not arisen by chance variability.

It should be noted that the numbers of students used for this statistical analysis differ from the preliminary analysis. This is because students who did not take the semester 2 iteration of AB1102 were excluded from the study, in order to make a comparison of performance as equivalent as possible. However, even taking this into consideration, the results show that provision of detailed assessment criteria and peer marking significantly improves student performance.

Conclusions

There are considerable advantages which can result from implementation of peer assessment. In this study, it appears that student learning can be enhanced by involving them in the assessment process. An added bonus of this study was that staff time spent marking practical assignments was reduced from approximately 25 to 6 hours per practical exercise, thus representing a significant time saving benefit to staff involved in teaching on modules with large student cohorts.

Whilst peer marking may be of benefit, it must be stressed that the process must be carefully monitored to cope with interpretation of marking schemes by weaker students. In the early stages of development of detailed criteria and mark schemes, considerable staff input needs to be invested to make sure that all documents are detailed and unambiguous, and that students of all ability are able to implement them. Monitoring to ensure that both positive and negative bias is not involved is also necessary, as is a means of ensuring that all students within the group engage in the peer marking process. If all of these criteria are dealt with, there is no reason why most practical exercises could not be marked in this way.

One note of caution to sound is that, should changes to assessment methods (*ie*, to include peer marking) be made, it is important that students are informed about all of the potential benefits, both within the module and in the longer term. Development of peer assessment skills is an important aspect of life-long learning which students can take with them to the workplace after graduation.

References

Race, P. (2001) A briefing on self peer and group assessment. *LTSN Generic Centre assessment series N^o 9 (2001)* (www.ltsn.ac.uk/genericcentre)

Appendix 1.

Example assessment criteria for biochemistry practicals (abridged)

Writing up the practical	
The report for this practical should contain two sections: Results and Conclusions . The details of what should be included are given below.	
Name, Date, Title of practical	Marks available 4
Results - Table 1	Marks available 16
Draw a table showing protein amount (mg/test volume) and absorbance data collected from the Biuret assay. Include the data for the standards, control and two unknowns (X & Y).	Marks are awarded for: A descriptive title/legend for the table Accurate column headings, including units if appropriate Including all the data for each column Good presentation
Results - Calculation 1	Marks available 5
Calculate the average absorbance for each of the unknowns (X & Y).	Marks are awarded for: Showing the calculation and getting the correct result
Results - Graph 1	Marks available 25
1) Use the data in Table 1 to draw a calibration curve for the Biuret assay, showing amount/test vol against absorbance. 2) After plotting the data, draw a (straight) line of best fit on the graph and use it to convert the average absorbances of samples X and Y into their amounts/test vol.	Marks are awarded for: A descriptive title/legend for the graph Putting the right data on each axis Accurate axis labels, including units if appropriate Choosing a sensible scale for the axes Accurately plotting the axes Correctly plotting the data points Correctly drawing the line of best fit Correctly indicating how the X/Y amounts/test vol are read off the graph Good presentation
Results - Table 2, Calculation 2, Graph 2	Marks available 46
Repeat what was done for Table 1, Calculation 1 and Graph 1 but using data from the Folin assay.	
Results - Table 3	Marks available 16
Draw a table showing the data for X/Y found from Graphs 1 and 2. Remember to convert the X/20 and Y/20 values for the Folin assay back to X & Y first. The values for the assay results should have the same units .	Marks are awarded for: A descriptive title/legend for the table Accurate column headings, including units if appropriate Including all the data for each column Good presentation
Conclusions	Marks available 8
Comment on the similarity (or otherwise) of the results with the two protein assays. Try and explain the reason(s) for any differences between the two assay results based on how they work.	