

Challenges of developing engineering students' writing through peer assessment

Teresa McConlogue (t.mcconlogue@qmul.ac.uk), Jens Mueller (j.mueller@qmul.ac.uk), Julia Shelton (j.shelton@qmul.ac.uk)

Queen Mary University of London, Thinking Writing and School of Engineering and Material Sciences, Mile End Road, London E1 4NS, UK.

Abstract: *Developing engineering students' writing should be a key concern but finding space in a content-full curriculum for explicit teaching of writing is problematic. Moreover, providing detailed feedback on writing to large classes is time consuming. To address this problem, peer assessment (PA) was trialled as a means of giving students practice in, and feedback on their writing. The benefits of PA have been extolled as a means of promoting active learning, facilitating students' understanding of marking processes and helping students to make judgements about quality (Falchikov, 2003; Orsmond, 2004; Bloxham and Boyd, 2007). This paper describes our experiences of trialling and embedding PA of written coursework over a two year period (2008-2010) in a School of Engineering and Material Sciences. PA was initially trialled with a medium sized group (56 students), and more recently, with a large first year cohort (280 students). We describe the design and implementation of these trials which were evaluated through observations, focus groups, questionnaires and interviews with tutors and students. Findings from the evaluation data confirmed our assumption that students would develop a better understanding of tutor expectations. Benchmarking their own work against that of their peers, helped students develop an awareness of quality in report writing. An unanticipated benefit of PA, however, was that students reported checking engineering content and conventions in order to give better feedback. Hence, through 'teaching to learn' (Topping, 2005), peer assessors began to develop both their writing and their engineering knowledge.*

Introduction

Developing students' writing should be a key concern in all disciplines as proficient student writers achieve higher grades (Northedge, 2003). Students need practice in writing so that they can develop their own academic identity and become more experienced at using academic discourse. In a field such as engineering, within a technical report, students need to engage in a variety of types of writing e.g. describing the procedure of the experiment, presenting a full set of results, analysing these results and discussing them in the context of established findings.

The Thinking Writing team collaborate with staff and students at Queen Mary University of London (QMUL), and other institutions, to facilitate the development of the effective teaching and assessment of writing. Since 2002, tutors from the School of Engineering and Material Sciences (SEMS) and writing specialists from the Thinking Writing team have worked together to develop writing opportunities for students. However, giving detailed feedback on writing is problematic because of large student numbers. Our response to this problem was to implement peer assessment (PA). Falchikov (2001: 2) defines PA as the grading of student work by 'their peers using relevant assessment criteria' and 'giving feedback to peers'. PA is claimed to promote active learning, facilitate students' understanding of marking processes and help students make judgements about the quality of their work by 'benchmarking' their work against others (Falchikov, 2003; Orsmond, 2004; Cho *et al.*, 2006; Bloxham and Boyd, 2007). Additionally, PA, it is claimed, enhances students' learning since students are involved in reading and critically reflecting on their peers' work, making judgements about the quality of other students' work, justifying those judgements by writing feedback comments and receiving and acting upon peer feedback (Liu and Careless, 2006). PA implementations have been reported in a range of disciplines, for example Cho *et al.* (2006) report on PA in psychology, history,

education and information science; in the UK, Orsmond (2004) describes case studies of PA in the biosciences.

PA of 'trivial problems', that is problems that require 'closed responses' (Schacter *et al.*, 1999), for example multiple choice questions, is relatively straightforward however, PA of 'ill-structured problems' (Voss and Post, 1989), problems which do not have a single correct answer, typically the kinds of problems that are dealt with in longer written coursework, is more problematic.

Much of the research literature on PA has focused on comparing peer-tutor marking (Liu and Careless, 2006) with the assumption that if peer marking agreed with tutor marking (seen as the 'gold standard') then it was reliable. However, the problem is that tutor marking may not be a gold standard, as recent studies on the reliability of tutor marking have shown (for example Baird *et al.*, 2004; Read *et al.*, 2005; Shay, 2005; Orr, 2006). Assessment criteria and descriptors, widely used to make explicit how assignments should be judged, are interpreted in different ways by different markers. Sadler's (2008) influential critique of explicit assessment criteria demonstrates that seemingly clear terms such as 'structure' encompass much larger understandings, drawn from tutors' tacit knowledge. Similarly Lea and Street (1998) argue that understandings of terms such as 'structure' and 'argument' vary depending on tutors' disciplines or sub-disciplines.

In this paper we want to move away from discussions on reliability of peer assessor grades, and focus instead on what students learned from PA. In introducing PA, we felt we could develop the work that SEMS tutors and Thinking Writing had been doing to unpack tutor assessment practices and develop student writing. We felt that the learning in this PA exercise would come about through students reading and commenting on their peers' work. Cho *et al.* (2007) argue that peer review improves learning both for the reviewee and the reviewer as communicating feedback involves explaining how students can improve and constructing the explanation helps the reviewers' to improve their own writing. Falchikov (2003) also comments that PA helps students improve their own writing and argues that it is important for students to give feedback and justify their assessment judgements. Moreover, engaging students in assessment allows them 'to develop their own internalised conceptions of standards and ... to supervise their own learning' (ASKe, 2008).

Implementing PA

In summer 2008, Thinking Writing staff and Julia Shelton (of SEMS) began planning the PA project. Early discussions with Paul Orsmond (consultant for project), and advice in the literature, (Falchikov, 2003; Orsmond, 2004; Bloxham and Boyd, 2007) stressed the importance of a preparation session where students would have an opportunity to do some rehearsal marking, using student reports from last year. PA was piloted in two course modules at levels 5 and 7 in 2008-09. In 2009-10, PA was extended and implemented in level 4 (level 4 = first year of undergraduate study in the UK, level 5 = second year etc.), as student feedback from the first implementation suggested that PA should be carried out at the beginning of their university studies

Implementation in a level 5 module

PA was implemented in a module taken by 60 students, including students studying for medical engineering MEng and BEng, medical students on a BSc programme, an MSc in Bioengineering and a number of students from USA engineering programmes. Students were assigned to a group which designed and undertook an experiment to test polyethylene samples; they were asked to write up this experiment individually.

After reports were submitted, students participated in a rehearsal marking session. Three selected reports from the previous year, were emailed to students who had to rank and write comments and bring to class. Some students did not fully engage in this preparation, therefore some time was provided in class for group interaction to discuss the comments, the issues that they found and to decide upon a grade for each report with justification. Lively discussions occurred and full student participation was observed. This prompted further discussion of assessment criteria and the tutor responded to student queries including technical clarification as to how the Young's modulus may be determined correctly and how to present raw results. A survey of marks from each group was taken and the variability in the marks was discussed. Sample feedback comments for one report were provided at the end of the session; these offered students a template of the range of issues and language that could be used.

Peer assessors were then allocated their marking. In order to ensure that peer assessors saw a range of reports, the tutor rapidly scanned through the submitted reports and divided them into categories

e.g. extremely competent reports, reports that were too long, reports that were clearly problematic etc. If reports were randomly assigned, students might not see the full range, so it was important to categorise them and allocate a report from each category. The reports were then anonymised as both students and staff felt that anonymising reports was essential to allay concern over marking of friends' work. Merry and Orsmond (2004) argue that anonymising reports is necessary to ensure that relationships between friends are not adversely affected by PA. Each peer assessor was allocated 4 reports to mark, ensuring that each report received multiple ratings, thus helping to eliminate bias (Orsmond 2004). Marks and feedback were then submitted to the tutor who calculated the mean mark and returned feedback comments and marks to students. Students marks were reduced by up to 10% if they did not write good feedback comments; all students participated in PA.

Implementation in a level 7 module

The level 7 group was a smaller group (26) whose task was to research an allocated topic and write a report appropriate to a line manager in industry. The assessment criteria were developed in class before the students started their research and were co-constructed between students and tutor. The tutor organised a rehearsal marking session, where the assessment criteria could be further discussed, and followed the procedures described above for allocating reports and collating and distributing marks. The level 7 group of students were highly motivated and just about to graduate. Most of the group wrote extensive feedback comments on their allocated reports.

In both level 5 and level 7 implementations, moderation procedures were established. Where a mark provided by a peer assessor was lower than 10% of the mean (as it was felt these would be the marks students would be most uncomfortable with), the report was read by the tutor. Either the mean mark was retained, if it was felt representative of an 'accurate' grade for the report, or the mark was discounted if it was regarded as incorrect and biasing the mean mark too heavily away from a 'fair' mark. Students were allowed to appeal with respect to their mark.

Drawing on student evaluations and tutor reflections from 2008-2009, we began to plan ways of further developing PA in SEMS. We felt that students needed to be introduced to PA in level 4, as feedback from level 5 and 7 students suggested that PA would give them a better understanding of tutor expectations at the start of the programme and would better prepare them for PA in levels 5 and 7.

Implementation in a level 4 module

In 2009-2010 in the level 4 induction (in week 1) students carried out a peer marking exercise, marking and commenting on two sample laboratory reports. Marks and comments were then discussed as a way of clarifying the assessment criteria and explaining the structure and content of a technical report.

As a follow up, in the middle of the first semester, a PA exercise involving a short (300 - 500 word) writing task was devised. The aim of the task was to help students write part of the background section of a lab report which would be submitted later in the course. They were asked to explain the principle of Bernoulli's theorem and then explain how it was applied during their laboratory experiment. PA constituted 10% of the overall assessment for this module; 5% for the student's writing task and a further 5% for the for the quality of their feedback comments.

The level 4 PA followed the procedures established in 2008-2009 (as discussed above) but as this module involved a large cohort of 280 students, it was important to manage the process electronically. Students uploaded their writing tasks through a web-interface which enforced the maximum text length. With a manual quick scan, 50 tasks judged to be well-written and 50 problematic tasks were then selected. Each student was randomly assigned 6 tasks to mark, with one well-written and one problematic task included in the 6 so that students were exposed to a range of quality. Students could only download and view those tasks that had been assigned to them. Once students had marked their tasks, they uploaded their marks and feedback through the web-interface.

As this was the first time we had implemented PA at level 4, the quality of each peer assessor marks and feedback was carefully checked. Each individual mark for a task was compared with the average mark for that task. Where a peer assessor's mark differed by more than 2 marks (out of 10) from the average, the marks and feedback from that assessor were eliminated. However, if more than one

assessor gave such 'outlying' marks, they were retained. In cases where assessors detected plagiarism, they indicated this in their comments and awarded zero marks. The tutor then reviewed all indicated cases of plagiarism and, if confirmed, only the marks and feedback of assessors who had detected the plagiarism were retained. Students who had plagiarised were contacted by the module organiser with an explanation and appropriate guidance. Students were given all the retained feedback, but only the average mark in each assessment category. Overall, nearly 10% of marks (138 of 1430 marks) were excluded.

Evaluating PA

We recognise that it is difficult to evaluate any educational intervention as, unlike experimental studies, the researcher does not have total control over the 'object' of research. It is impossible to isolate students' PA experiences from all the other experiences and interactions that they are learning from e.g. writing in other modules, interactions with other students and their own reading. However, in order to get some idea of students' and tutors' experiences of PA we administered questionnaires (adapted from Ferguson et al 2008) in level 5 and level 7 implementations and these were followed up with focus groups in 2008-2009. In 2009-2010, students in the large level 4 implementation were asked to respond to 3 questions: a) Was PA helpful? b) What should be improved? and c) Any other suggestions? Students uploaded their comments along with their marks and feedback, so the response rate was good (234 respondents). Some issues to improve emerged from these evaluations, but this paper focuses on what students said they learned from PA.

Our assumption that students would benchmark their work against that of their peers was confirmed as many students reported that PA was valuable because it gave them the opportunity to read peers' reports and make comparisons:

[Key: Q5/Q7 – open comments from questionnaire administered to level 5 and level 7 students respectively; FG5/FG7 – focus group quotes from level 5 and level 7 students respectively; C4 - online comments submitted by level 4 students].

S: How other people approached the same problem that you did. . . and how they're all better or not so good. [FG5]

S: It was interesting to find what I do well and what I lack and thus learn to do. It was interesting to have the opportunity to give my opinion on others' work.[Q5]

S: It was helpful to have an opportunity to go through the work of peers in detail and see what level they are at in comparison to myself.[C4]

S: To begin with I was not sure [of] the point of marking other people's coursework. I believed it was just an unnecessary task. However I have completely changed my mind. What I found most beneficial, was that I could compare my writing with those that I believe[d] were better, in turn allowing me to find mistakes in my own writing and subsequently improve my writing. [C4]

As discussed above, we wanted students to see a range of work and assumed they would learn from seeing variation in report writing; seeing how other students approached the task differently, helping them to develop judgements about quality in student writing (Sadler in press). Students reported identifying errors in peers' reports and this helped develop their understandings of what a good report looked like. One student commented that she began to:

S: ...understand, by reading strong reports, how things should be done, and weaker reports, how they never ever should be done.[FG 7]

Reading peers' reports seemed to help students identify their own errors and understand how they could improve:

S: I learnt by reading other people's report[s] how to improve my report writing skills.[Q5]

S: It helped me realise what was wrong about my own report and how to correctly structure referencing and calculations.[Q5]

S: It is really helpful to see the mistakes that people do almost systematically. We also see the importance of reading the final essay before handing it in. [C4]

S: I had an insight into [how] other students were approaching the essay [which] gave me a much broader thought to what I could have also added to my essay. [C4]

Alongside this sense of developing judgements about quality in report writing, students also reported a growing understanding of tutor expectations. In SEMS students are often given guidance on expectations through written task instructions and assessment criteria, supplemented by tutor's oral explanations of the task in lectures. Nevertheless, students reported a greater understanding of tutor expectations as a result of PA:

S: It help[ed] me to understand what lecturers expect from a good report.[Q7]

S: It made [...] clearer the requirements of a report.[Q7]

S: Helped me understand how academics might analyse my work an[d] thus trying to improve future course work remembering this or trying to think as an academic.[Q7]

S: Most of all, looking at the essay from the brain of the corrector is crucial [C4]

S: It also taught me how essays are marked and also the main general rules of essays, such as the referencing, grammar, the amount of words etc. This exercise has been helpful. [C4]

S: In addition, I found [out] how the mark scheme was structured. [It] was clear and told you exactly what you have to be looking for, so marking was made easy. [C4]

It seems that reading and commenting on a range of reports, using guidelines and assessment criteria, helped some students clarify their understandings of tutor expectations. One level 7 student was particularly enthusiastic about her PA experience and felt she could use the understanding she had gained in other coursework:

S: Well, I have learned how to assess my own work, [...] reading my work as if I was another person, another point of view, completely different, and impartial. And I was so enthusiastic with these new abilities that I have had in doing peer assessment, that I started applying it to all the course work.[FG7]

[...] it's nice that I am influenced by this activity, and now I wish my colleagues to read my work before delivering it, so I don't see it [as] harmful I see it as very helpful.[FG7]

For this student, PA seems to have brought about a profound change in the way she learns.

A further and unanticipated benefit of PA occurred when students approached the task of giving feedback to others. In an effort to give 'fair' and 'correct' feedback some students checked on details such as correct referencing:

S: it did help me, because, with referencing, the referencing websites, I didn't know how to do it properly, and so I actually looked it up, so that I could make sure that I was being fair, the way I was commenting on the people's [work] so, I did actually learn stuff along the way, as well, which was good, so, it did make me look at things that maybe I'd neglected before... [FG7]

S: and you go and read up how it should be done before you mark, and you learn it yourself. And the referencing and figures are the major reasons for deduction by markers. [FG 7]

S: Also what I found useful was that this task forced me to do some more reading on the Bernoulli principle, and make sure what they were writing was correct and to the point. [C4]

Many students said reading and writing feedback on peers' reports helped them recognise that they'd been referencing incorrectly. Even level 7 students (a mixture of MEng and MSc students on this course) commented on uncertainty about referencing 'rules', despite the fact that they had received guidance on referencing from tutors and library staff. It seems that having to engage actively with referencing conventions through assessing their peers' work, and the need to give accurate advice, motivated students to seek out information on referencing conventions.

Discussion

For many SEMS students being able to benchmark their writing against their peers' was invaluable and helped clarify the standard they needed to aim for. Through reading a range of reports (seeing variation in report writing), students could learn about different ways of tackling the task e.g. different

ways of writing an introduction or analysing results. In reviewing research on peer tutoring, Falchikov (2001) claims that peer review is successful in improving students' writing.

As anticipated most students reported a better understanding of tutor expectations and the assessment criteria as a result of PA. Actively engaging with assessment criteria by discussing it with the tutor in the rehearsal marking and then trying to apply it to their own marking, seemed to give students a better understanding of the criteria. In Sadler's (2008) terms, students had the opportunity to develop and expand their tacit understandings of the criteria.

Some students reported that PA gave them a better grasp of what a good report looks like. Sadler (in press) argues that students need experience in making judgements about the quality of written work. They need to see a 'variety of works of different quality' in order to develop their judgements. We ensured that students saw a range of quality, as tutors pre-screened the writing tasks and ensured that peer assessors were allocated tasks representing a variety of quality. Through making and justifying assessment judgements, some students seem to have developed their concept of quality.

It seems that having the responsibility to comment on peers' work and offer guidance, prompted students to learn. An unanticipated result of PA was students reporting that they were looking up information, especially referencing conventions, in order to give better feedback. Through 'teaching' their peers, students seemed to have enhanced their own understandings. In order to advise their peers, they had to check their own understandings and then construct an explanation, and doing this clarified their own knowledge. This may have been the most valuable aspect of peer assessment for our students and may have been most useful for developing their writing. Topping (2005) theorises 'teaching to learn' arguing that the 'helpers' (in this case the peer assessors) develop their understanding of a concept through constructing an explanation for their peers. The need to communicate (and to check one's understanding of the concept) develops the helpers' knowledge. In addition the need to articulate this knowledge to another challenges and develops the helpers' communication skills.

Future work

The introduction of PA into SEMS modules has required careful planning and preparation. Managing PA so that students see variation, and can therefore start to develop judgements about quality of work, has been challenging. Moreover, managing PA with a large group (280 level 4 students) has necessitated developing an electronic system which was trialled in 2009-2010. Following student and tutor feedback, the system will be modified to address the issue of variability in peer marking drawing on SWoRD's (scaffolded writing and rewriting in the discipline) system of reviewer rating. SWoRD is an online peer review system where peer assessors' marks are automatically compared to the average mark (Cho *et al.*, 2007). Peer assessor marks found to significantly deviate from the mean mark will be weighted and peer assessors' will receive feedback on their own consistency and on the consistency of their assessors. Peer assessors will then be graded on the quality of their feedback comments, and we anticipate this will motivate students to think carefully about their grades. Moreover, we hope it will reassure students that PA is as fair as any other type of marking.

The data from the evaluation of PA has given us an insight into what students felt they learnt. Our assumption that students develop a better understanding of assessment criteria and tutor expectations seems to have been supported. What we did not expect was the enhanced understandings that students reported as a result of 'teaching to learn'. Some students reported improving their understandings of concepts as they looked up information in order to give 'correct' feedback to their peers. We want to find out more about how students learn from PA and have begun to conduct in-depth interviews with a small number of students, tracking them through into their next module to try to understand how their PA experience has developed their writing.

Acknowledgement

In 2008-2009 the peer assessment project received funding from the mini-projects for development and research in learning and teaching scheme administered by the Higher Education Academy Engineering Subject Centre.

References

- ASKe (2008) *Assessment Standards Manifesto*. Available from <http://www.brookes.ac.uk/aske/manifesto.html> [accessed 7 April 2010].
- Baird, J., Greatorex, J. and Bell, J.F. (2004) What makes marking reliable? Experiments with UK examinations. *Assessment in Education: Principles, Policy and Practice*, 11 (3), 331-348.
- Bloxham, Sue, and Pete Boyd. (2007) *Developing effective assessment in higher education: A practical guide*. Maidenhead: Open University Press.
- Cho, K., Schunn, C. D. and Kwon, K. (2007). Learning writing by reviewing in science. *Proceedings of Computer-Supported Collaborative Learning*. New Jersey.
- Cho, K., Schunn, C. D., and Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication*, 23(3), 260-294.
- Falchikov, N. (2001) *Learning Together: peer tutoring in higher education*. London and New York: RoutledgeFalmer.
- Falchikov, N. (2003) Involving students in assessment. *Psychology Learning and Teaching*, 32 (2), 102-108.
- Ferguson, G., Shearer, E. and Ruth Grady. (2008) Computer-Assisted and Peer Assessment: A Combined Approach to Assessing First Year Laboratory Practical Classes for Large Numbers of Students. *Bioscience Education E-journal*, 11(4) [available at <http://www.bioscience.heacademy.ac.uk/journal/vol11/beej-11-4.aspx> , accessed 29th May 2009]
- Lea, M. & Street, B. (1998) Student writing in higher education: an academic literacies approach. *Studies in Higher Education*. 23 (2),157-172.
- Liu, Ngar-Fun and Careless, David (2006) Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290.
- Merry, S. and Orsmond, P. (2004) The effect of marking criteria and exemplars on students' learning during peer- and self-assessment of scientific posters. In: Orsmond, P. (ed.) *Self and Peer-Assessment: Guidance in Practice in the Biosciences*. Leeds: Centre for Bioscience, The Higher Education Academy.
- Northedge, A. (2003) Rethinking Teaching in the Context of Diversity. *Teaching in Higher Education*, 8 (1), 17-32.
- Orr, S. (2006) Justify 66 to me! An investigation in to the social practice of agreeing marks in an HE Art and Design department. In: Rust, C. (ed.) *Improving Student Learning Through Assessment*. Oxford : Oxford Centre for Staff and Learning Development, 283-293.
- Orsmond, P. (2004) *Self and Peer-Assessment: Guidance in Practice in the Biosciences*. Leeds: Centre for Bioscience, The Higher Education Academy.
- Read, B., Francis, B. and Robson, J. (2005) Gender, 'bias', assessment and feedback: analyzing the written assessment of undergraduate history essays. *Assessment and Evaluation in Higher Education*, 30 (3), 241-260.
- Sadler, D.R. (2008) Transforming holistic assessment and grading into a vehicle for complex learning. In: Joughin, G. (ed.) *Assessment learning and Judgement in Higher Education*, Netherlands : Springer, 45-63.
- Sadler, D. R. (In press). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*.
- Schacter, J., Herl, H.E., Chung, G.K.W.K., Dennis, R. A., and H. F. O'Neil Jr. (1999) Computer-based performance assessments: a solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behaviour*, 15 (3-4), 403-418.
- Shay, S. (2005) The assessment of complex tasks: a double reading. *Studies in Higher Education*, 30 (6), 663-679.
- Topping, K. J. (2005) Trends in peer Learning. *Educational Psychology*, 25 (6), 631-645.
- Voss, J. F., and Post, T. A. (1988) On the solving of ill-structured problems. In: Chi, M. T., Glaser, R., and Farr, M. J. Hillsdale (eds.) *The nature of expertise*, NJ: Erlbaum, 261-285.

Copyright © 2009 Teresa McConlogue, Jens Mueller and Julia Shelton: The authors assign to the EE2010 organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to the Engineering Subject Centre to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the EE2010 conference proceedings. Any other usage is prohibited without the express permission of the authors.