From Pilot to Practice: Embedding peer and self-assessment into an undergraduate module in computer science

Keith Primrose, School of Computer Science and Electronic Engineering, and Kate Dunton, Learning & Development

Introduction

A few years ago, Keith observed that first-year students on his computing science module were tending to underestimate the work involved in completing their individual projects – in this case, the production of a website – and were typically engaging with it far too late in the term. Remembering a talk given by the educationalist Graham Gibbs, and the literature on student assessment that he had read whilst completing a PGCE, he began to wonder if peer and self-assessment might be a useful way forward. The result was a TALIF-funded initiative to see if the introduction of a formative self- and peer-assessment exercise into the module could improve students' engagement with, and planning of, their projects. This article will discuss that pilot, its evaluation, and the modifications made in subsequent teaching cycles, reflecting on what has been learned from the process.

The core assumptions of self- and peer-assessment are student-centred, that is to say, they work towards 'active engagement in learning and learner responsibility for the management of learning', (Nicol and MacFarlane Dick, 2005, p.200). The rationale for giving students an active role in the assessment and feedback process for this particular initiative, therefore, was not the production of grades, per se, but the skills that students might develop for judging their own performance and reflecting on and regulating their learning processes. Peer-assessment, in this context, was seen an extension of self-assessment. The process of evaluating the work of fellow students might strengthen the individual's sense of what makes a piece of work strong and how it might be improved; skills that could then be applied to more rigorous and sophisticated judgements about their own work.

Sadler (1989) writes that for students to be self-regulating, they must:

a) Possess a concept of the *standard* (or goal, or reference level);
b) Compare the *actual* (or current) *level of performance* with the standard;
c) Engage in appropriate *action* which leads to some closure of the gap.

(p.121. Sadler's italics)

Later studies (Nicol and Macfarlane Dick, 2005) also emphasise the motivational and behavioural aspects of self-regulated learning. Students who do well are typically able take control of, and adapt,

their study strategies and motivational beliefs in relation to the demands of the course. Finding some overlap between their personal, internal goals, and the external goals set by the teacher is an important part of this. The social context also plays an important role, since complex judgements about performance are honed through comparisons to external feedback from peers and teachers.

Bearing the research in mind, it was clear that to create the right conditions for effective peer- and self-assessment it would be important for students to "understand that their learning is being deepened by involvement in their own and other's assessment in order to be persuaded that they are benefitting from the process," (Race, 2001, p.11). The criteria for good performance would also need to be clear and students would need opportunities to assess themselves and peers against these criteria. Finally, it would be necessary to provide tutor feedback to students against which they could orient their self-evaluations, and more generally to provide a climate of discussion around the learning process.

Implementation

In 2010, the CE154 Web Development module had a cohort of 90 students. 40% of the mark was given to an individual project which required students to produce their own website. Although the students would not self- and peer-assess until later in the term, it was decided to give students the grading criteria for the project in the second week so that this could start to inform their approach from an early stage. This grid gave descriptors for each element of the task across a range of marking bands, e.g. 53% and above, 64% and above, (appendix 1). As can be seen, these do not follow the usual classification system. The intention was to discourage students from thinking that coverage of certain elements would automatically gain them a 2:1 or a first, which might cause dissatisfaction when they received their final mark from the tutor. (As it turns out, this was probably unnecessary, see the discussion section below).

The students were told that 10% of the 40% project mark would be given for a self- and peer-assessment activity later in the term. In week six, the students were asked to use the criteria grid to self-assess their progress on the project as it stood at that point. This self-assessment was then submitted to the module tutor electronically. Each student also posted their work to a specific server using their student registration number to ensure anonymity. From this server their work could be accessed by other students. Although the other students could see the website they could not see the source code that had generated it, thus avoiding plagiarism. At this point, the students were asked to grade one other student's submission using the same criteria form that they had used to self-assess, and to submit that assessment back to the module tutor. The module leader returned to each student the peer assessment for their work. Students then had a week before the final

submission date to improve their work if they thought it appropriate. The work was marked by the module staff using the same criteria grid and the results were returned to the students.

Evaluation

The impact of the project was evaluated in two ways. First, the marks profile for the 2010-11 cohort was compared with the two previous years to see if engagement with peer- and self-assessment had impacted on student achievement in the summative assessment. Secondly, three focus groups were conducted with a total of eighteen students in order to explore whether engagement in the project had altered the students' perceptions of the assessment and feedback process and impacted positively on their learning behaviours and motivations. Student participants for the focus groups were selected and then invited individually by Keith to ensure that they represented the full spectrum of marks. It was decided that the focus groups would be led by colleagues from Learning and Development on the basis that the students might find it easier to share their thoughts openly with a third party. At this stage, the authors met to discuss how the information from the focus groups might inform future design and planning of the module and from this were able to design a number of questions and activities that would be repeated across the focus groups.

The comparison of the marks profile did not reveal notable improvements, which may or may not be significant given the relatively small size of the cohort, the fact that only three years of data was available, not to mention the many other contributory factors that may be involved. The focus groups, however, did yield a number of interesting insights and will be discussed in more detail below.

The findings of the focus groups

Keith and Kate wanted to discover whether this pilot initiative had, indeed, improved student engagement and learning in the ways suggested by the literature (Boud and Falchikov, 1989, Boud, 1995, Race, 1998, Nicol and Macfarlane Dick, 2005, Orsmond, 2004). Had it, for example, developed the students' sense of ownership over their own work, encouraging them to judge their performance against their own goals not just those of the teacher? Had it allowed them into the assessment culture, deepening their understanding of the criteria? Had it helped them to anticipate the final grade given by the tutor and encouraged them to construct their own understanding of the feedback? Had it impacted on their learning behaviours, supporting them to identify strategies to close the gap between actual and desired performance? And finally, returning to the starting point of the project, had it helped them to judge more realistically the time and effort that they needed to devote to the task?

Interestingly, when asked about the self-assessment element, most of the students did not see this as specific to the set task of self-assessing in week six, but as an ongoing process through the whole project that began as soon as they had received the criteria grid and been told they would be self-assessing against it. They said that knowing that the self-assessment task would contribute to 10% of the marks for the project made them get involved with the grid earlier, describing it as an important 'carrot', but were emphatic that it was the grid itself that motivated them to continually self-assess and enhance their project against the criteria.

The use of the criteria grid for all three stages of assessment – self, peer and tutor – had also helped them to relate external feedback to their own judgements. Interestingly, there was a general feeling across the groups that the mark from their peers was more reliable than their own.

> Seeing what other people think of your work might make you understand – I didn't do that, I should do it.

They used peer assessment to go back and check certain things, particularly where the peer assessment was lower, working out why it was lower, and where they had made mistakes. All the groups agreed that the tutor was the final arbiter. When the tutor's mark was higher than their self-assessment, for example, they said that it helped them to see how far they had actually progressed. When it was lower it helped them to think why it was lower, and what could be changed. The criteria helped them to understand their mistakes and work out how and why marks had been given (or not given).

> It was a bit annoying... but it makes you think… OK, why has this person said this? What do I need to change on my page to change their mind?

Again, all groups described this as part of an ongoing process of correcting and refining their work against the grid.

A key point that emerged from the discussions was the clear link between levels of motivation and the clarity of the criteria. The students said they were more likely to focus on achieving the criteria where they knew exactly what was expected. Where it was ambiguous they were more likely to leave that element and work on something else:

> Where it's ambiguous if you're feeling unmotivated you'll do the bare minimum… Oh it might mean this so I'll just do that as it seems easiest.

They were, conversely, more motivated where the criteria was clear 'because you know what you've got to work for'.

Across all the groups, the students reported difficulty deciding between marking bands when relative or qualitative words were used, e.g. 'most' 'mostly' 'very good' 'basic' 'consistent', 'limited' etc. It was agreed that where possible it would be helpful to have more specific indicators, e.g. 'the menu is consistent across all pages, but other elements may vary', etc. It was noted here and elsewhere in the discussion that the opportunity to look at examples was crucial to working with the criteria effectively, as well as the opportunity to ask the tutor questions. The students appreciated the clear explanations and feedback they had received from the tutor and the real-life examples of good practice that they had been shown.

So how did the use of the assessment grid impact on learning strategies? Specific strategies varied between students. Some started at the lower end and worked upwards:

> The X nearest to nought was what I needed to work on....if I've got something down here I haven't even started yet, I've got to get this up.

Others looked at what needed to be done to achieve the highest marks then worked backwards to what they thought they could achieve. All the students noted, however, that the grid helped them identify specific goals and monitor progress and this seemed to be the key to motivating engagement with the task:

> As you're going through your assignment you don't really realise how much of it you're getting done. But when it came to submitting the claim form it was like, OK, so I've done this, I still need to do this. You know, it was really helpful for measuring yourself.

As one student noted, the process gave: 'Little bits of positive feedback. Little bits of achievement.'

Another student said it was like having a 'status bar' through the project, and the visual aspect of the grid was noted by a number of the students:

> As we went through the module I was like, yeah, OK, I can do this now, I can do this, I can do this. It was quite nice to have such a visual encouragement of how well you're progressing.

The degree to which the students' understanding of the criteria evolved throughout the project was, not surprisingly, dependent on their familiarity with the technical aspects of the task. Some did not know what the terms meant initially and had to look them up. Students also noted that the peer assessment had prompted them to reassess their understanding of the criteria:

> You see how you've understood [the criteria], and how they've done it, and if there are any differences, you know about it. What does good formatting mean? When you see the others' work you will have an idea what is good and what is not.

A key question from the beginning of the pilot was whether self- and peer-assessment might encourage students to start their projects earlier and spend more time on task. The students were clear across all three groups that this had been the case, for a number of reasons. Firstly, having the criteria in week two, then the self-assessment in week six followed by the peer assessment, kept the task in focus throughout the module:

> Other assignments were easier to forget about because they lacked the structure that this had.

One student noted that they treated the peer-assessment like a deadline, and this got agreement from the group:

> It did spread the time out well, the peer assessment. It meant we had to start a few weeks before the deadline. So it meant that, rather than starting a couple of days before it was due, it was like we started a few weeks earlier and got the basic website done.

Students also noted that the grid helped them to overcome procrastination at the start.

> Once you've got the basics done, once you've started it, it's not too hard to go for the higher marks. It's like starting the assignment, once you've started it's easier.

Time on task was clearly also linked to motivation, which was enhanced when the students were clear about what they needed to do:

> Before the grid we didn't have clear goals for our assignment, we just did it, and most of us didn't start at all because we didn't have the goals. After we had the goals we were like well if I want to do at least forty, at least a pass, I should get onto it.

The students noted that the grid helped them to realise how much work they needed to do. Without it, they tended to underestimate the time/effort required:

> Before I was given this, [the tutor] said something, and I was like, yeah, I'm actually going to get above 75%, it seems easy. But then when I look [at the grid], I was like, No, I'm not going to get this. Then I was like, yeah, I should just start by passing.

And from another student:

> I know from my experience that I'm looking at an assignment and I'm like, Oh it's easy and I'll be able to do it in the last week, or last three days, and when I get there I'm like, Oh no, what am I going to do.

Another important factor for keeping students on task was that it the grid helped them to develop strategies for when they hit a problem, encouraging them to keep going when they might have given up:

> When I was getting frustrated about a particular point, [I would think] OK, I'll give up on this one for now, and try a different one. Because you had the list of what you needed to do, it was quite useful, you could just leave something and work on something else, whereas if you didn't have [the list], you'd keep stressing on it, and eventually you'd give up on it and come back to [the whole assignment] later.

The students' initial levels of confidence and motivation on seeing the grid depended on their prior experience. Those that already had experience of web design were reassured by the grid. Others admitted that in the beginning they felt a little panicky when they saw the grid as they were not familiar with the terms and did not know how to do the things required. They were surprised by the volume and standard of work indicated by the criteria and the amount of research required. This made them realise how much work they needed to do. Later on in the project, however, they reported that the grid had helped them to feel more confident and clear about what they were doing. They were able to cover the basics then build up from there.

Another issue explored in the focus groups was whether the role of self- and peer-assessment had helped the students to anticipate and/or better understand the tutor's evaluation. Interestingly, the students expressed increased satisfaction with their mark regardless of whether it had been high or low, saying that the self- and peer-assessment process had helped them to predict what mark they could expect more confidently.

> In other modules, when we got a lower mark we didn't know why, but when we got our marks for this we knew what we didn't do. We were expecting that mark, exactly – well, plus or minus two.'

> After doing this, I think this is the fairest mark I've got. I really expected it. And the feedback is really helpful.

> In other modules we did a lot of work and were confident we were going to get a really good mark, but we didn't, and we didn't know why, and we still don't know why.

Significantly, even when the tutor gave a lower mark than the peer/self-assessment, and even where the mark was 'low' according to the students own perception, they still felt more satisfied with the mark:

> Yes, it's about being forewarned. You resolve in your mind that you're going to get this mark, and you know why. It's less of a shock.

The students in one group were explicit about the fact that this had positively impacted on their module evaluations:

> Yes, yes. It was the best thing. Because now we have to complete like those forms where we say that the marking criteria was fair.
>
> Yeah, and a lot of my lecturers have got 'totally disagree'.
>
> But with Keith, it's totally agree.
>
> Totally agree. Because he did make it clear. And it was really good. Feedback was very good because of this.

Importantly, students were also able to see the benefits of self- and peer-assessment for their professional lives beyond university. One group spontaneously noted that the ability to judge others' work objectively and to respond to feedback from peers was an important transferrable skill.

> Knowing how to present your own abilities and what you've done is really important. Because you can objectively say I haven't done this, I have done this, I have to do this. I like to think employers want honesty and professionalism.

Another benefit that the students mentioned unprompted was the usefulness of the assessment grid for promoting and supporting meaningful discussion with tutors and peers. One group commented that having the criteria helped them to discuss the project with their fellow students – they did not 'waste time discussing stuff that wasn't relevant'. It also helped them to ask specific questions of peers and the tutor. They could also relate the criteria to the lectures, forging stronger links between what they were learning in class and the assessment system. Asked if they would like to self- and peer-assess again, students in all three groups responded that yes, they definitely would if they were given the criteria. They could not understand why this was not standard practice in all their modules.

Discussion and Subsequent Iterations

Looking at the results of the focus groups, it was evident that the criteria grid had played a key role in supporting the students' self-reflections, prompting them to monitor their learning well before the self-assessment exercise in week six. The peer assessment process was also useful for prompting further reflection on, and clarification of, the criteria and modifying their own self-assessments. The students evidently accepted that the tutors on the course were the final arbiters of the mark. They understood that the self-assessment helped them to evaluate their own work, and motivated them to improve their performance, and saw the final mark as a continuation of this process. At worst, where students had not performed well, they at least knew why they had not performed well and had the opportunity to address that in future assignments if they wished. The authors also note the link between student motivation and the quality of the criteria; the clearer students felt about what was expected of them, the greater their engagement in the task.

As a result of the experience of the pilot exercise and the findings from the focus groups, future iterations and enhancements have focussed on a number of key areas: streamlining the process, which had proved to administratively burdensome; improving the wording of the criteria; developing the supporting structures around the self-assessment process; and retaining the features that had worked well first time around.

In subsequent cycles, it was decided not to include the peer assessment element since this had proved to be the most time-consuming from the tutor's perspective and, whilst useful, had not proved to be an essential part of the process. Retained was the introduction of the criteria grid in week two of the module, and the formative self-assessment exercise in week six. Bearing in mind the students' insistence that the percentage mark had been a valuable 'carrot', an incentive was retained but reduced to 5% of the 40% marks allocation for the project. In this way, the learning benefits were retained and the administrative load reduced.

In terms of improving the wording of the criteria, the focus was on eliminating those ambiguous terms that the students in the focus groups had struggled with – basic, good, very good – and replacing them with clear indicators of what was expected. The key was to strike a balance between providing criteria that were clear enough for students to reflect on whether or not it has been met, but not so prescriptive that it did the thinking for them. For example, a criteria might say, 'takes due account of accessibility and usability', but would not say, 'should include a button that enables to user to change the size of the font'. To put it another way, the criteria needed to provide a clear idea of the standard required for each element and in each marks band, but also make clear to students that it was not a recipe that could be mechanically followed to achieve a certain grade. Work was

also done on the wording to ensure that the distinctions between the marking bands were absolutely clear.

The supporting structures around the self-assessment process were also maintained and developed. In subsequent cycles, students have been introduced to the grading criteria in an early lecture, with the point made that they may not understand all the elements – yet – but they should take it on trust that by the end of the term they will be able to complete every aspect of the assignment. For each technical area, the course web-pages include guidance on further reading and a self-assessment quiz. In this way, once students have identified a goal or a weak area to work on, they have immediate support on how to develop their understanding and competence in this topic.

From ongoing observation and discussion with students it is clear that the practice of sharing and discussing criteria with students early in the module, and engaging them in a self-assessment exercise several weeks before the project deadline, continues to have significant benefits. It promotes improved levels of self-awareness in the students around their own progress. It motivates greater time on task and better planning, and encourages greater ownership over the learning process.

Conclusion

It is only through trial and error, and, just as importantly, reflecting on what did and did not work and why, that students can gain genuine insight into what they need to do in order to learn and progress effectively. This was the message that the authors felt emerged most clearly from this project. The key benefits of using self-assessment were the opportunity it gave students to try things out and reflect on the results against clear external reference points before revising and trying again, and the space it gave them to step back from and consider their own learning processes. Given that the average age of a student coming to university is eighteen it is unlikely that most will arrive with these reflective skills. Sharing clear, well-written criteria with students and engaging them in self-assessment processes is one way of introducing them to a reflective approach as early as possible in the hope that by their third year they might be effective and independent learners. In the opinions of the authors, self-assessment is a powerful tool for significantly changing students' attitudes towards their studies – moving them from 'I'm here – learn me', to 'I'm here to learn'.

References

Boud, D. (1995). Enhancing learning through self-assessment, London, Kogan Page.

Boud, D and Falchikov, N. (1989). Student Self-Assessment in Higher Education: A meta-analysis. *Review of Educational Research,* 59 (4): 395-430.

Nicol, D. & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education,* 31: 199-218.

Orsmond, P. (2011). *Self- and Peer- Assessment; Guidance on Practice in the Biosciences*, 2nd Edition Ed: Stephen Maw, (Leeds, The Higher Education Academy Centre for Biosciences).

Race, P. (2001). A Briefing on Self, Peer and Group Assessment, LTSN Generic Centre Assessment Series No. 9.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18: 119-144.