Assessment for learning: using minor assessment to promote major learning

Keith Willey
Faculty of Engineering & IT, University of Technology, Sydney, Australia, Keith.Willey@uts.edu.au

Anne Gardner
Faculty of Engineering & IT, University of Technology, Sydney, Australia, Anne.Gardner@uts.edu.au

The authors have previously reported the effectiveness of using self and peer assessment to improve learning outcomes in groupwork by providing opportunities to practise, assess and provide feedback on students' attribute development. Combining this research and that reported in the literature regarding learning-oriented assessment we theorised that self and peer assessment would be an ideal tool to develop and efficiently facilitate activities specifically designed to be student centred and promote learning.

In this paper we report the effectiveness of a self and peer assessment activity specifically designed to promote collaborative peer learning, require students to take responsibility for their learning and improve their judgement, while at the same time only imposing a small assessment load on academics.

Keywords: self and peer assessment, learning-oriented assessment, collaborative learning, peer learning, SPARK^{PLUS}

Introduction

In addition to providing fairer assessment of group work, self and peer assessment is reported as assisting students to develop important professional skills including reflection and critical thinking (Mello, 1993; Somervell, 1993). Michaelsen, Knight and Fink (2004) discuss the use of self and peer assessment to promote peer learning, while Hanrahan and Isaacs (2001) report that it contributes to students' development of critical thinking skills and motivates students to submit better initial submissions knowing they would be reviewed by their peers. Willey and Freeman (2006a, 2006b) report using self and peer assessment to produce formative learning-oriented feedback to complete the learning cycle and encourage the ongoing development of skills. Furthermore, Boud and Falchikov (2007) discuss its use for developing students' skills for lifelong learning. More recently the authors have reported the effectiveness of using self and peer assessment to improve learning outcomes by providing opportunities to practise, assess and provide feedback on students' graduate attribute development (Willey & Gardner, 2008a).

In the last decade momentum has grown for assessment to change from 'assessment of learning' to 'assessment for learning' (Torrance, 2007). Learning-oriented assessment embeds learning in assessment, reconfiguring its design to emphasise the function of learning (Keppell & Carless, 2006; Keppell, Au, Ma, & Chan, 2006). Its three main elements are: assessment tasks that also focus on learning, involving students in the assessment process to develop their graduate attributes including judgement, and feed-forward to improve subsequent contributions and learning (Carless, 2007, Black & Wiliam, 1998).

While the provision of detailed feedback and assistance by instructors typically leads to higher quality student submissions, care needs to be taken that the challenge is not removed from the learning process nor that the quality and/or validity of the assessment is not reduced (Torrance 2007). The authors have noticed a tendency for students to become "incremental learners" whereby they seek ongoing feedback and advice from instructors to improve their submission. In these instances there is a danger that students are not exercising their judgement, but simply implementing what the instructor has told them to do. Their focus being on securing a better grade rather than learning or even understanding the feedback provided. They believe that their best chance of securing a higher grade is to give the instructor exactly what they want,

without question. This belief is not unfounded and will continue to be a strategy used by students if supported by the chosen assessment practices. Rather than building the skills required for independent learning such practices may actually produce "…students who are more dependent on their tutors and assessors (Torrance 2007, p. 282). This is in contrast to peer learning which encourages students to take responsibility for their own learning (Keppell et al., 2006).

Reflecting on this research we thought that self and peer assessment would be an ideal tool to develop learning oriented assessments that would also address the above issues. In particular we were interested in making students more responsible for their own learning by requiring them to provide their own feedback and contribute to their own self assessment and to the assessment of their peers.

Specifically our research aim was to test whether a carefully designed minor assessment task, conducted outside normal class time would be effective in motivating significant learning.

In this paper we report testing this theory by integrating self and peer assessment into an activity specifically designed to promote collaborative peer learning, require students to take responsibility for their learning and improve their judgement. In designing this activity we chose to use the online tool SPARKPLUS (Willey & Gardner, 2008b) to assess and provide feedback on individual student submissions.

SPARK^{PLUS}

SPARK^{PLUS} is a tool that traditionally has been used to assess a student's contributions to a team project. The tool was recently expanded to also facilitate self and peer assessment of individual work and benchmarking exercises to develop students' judgement (Willey & Gardner, 2008a & 2008b). As this paper only discusses the individual submission assessment (ISA) mode we will only describe this aspect of the tool's operation.

The ISA mode facilitates an instructor creating an assessment whereby students assess their own and their group peer's submission, activity or task against a number of specifically chosen criteria. Instructors can choose from a number of predefined rating scales or create their own, however typically for ISA a Standard Assessment scale is used (for example Unsatisfactory (Z), Pass(P), Credit (C), Distinction (D) and High Distinction (HD) (Figure 1). The flexibility for instructors to choose their own criteria and rating scales means the ISA mode can be used to mark virtually any activity, for example, individual reports and oral presentations.

Students are required to logon and enter their assessments of the work by moving the sliders (orange bars) against a number of criteria, first for themselves and then for each of their team peers. To assist students to make comparative judgements when marking their team peers their self assessment of their submission is displayed for each criterion by the upper (blue) triangle (Figure 1). Furthermore students are encouraged to provide written feedback to their peers regarding their assessment of their peer's submission.

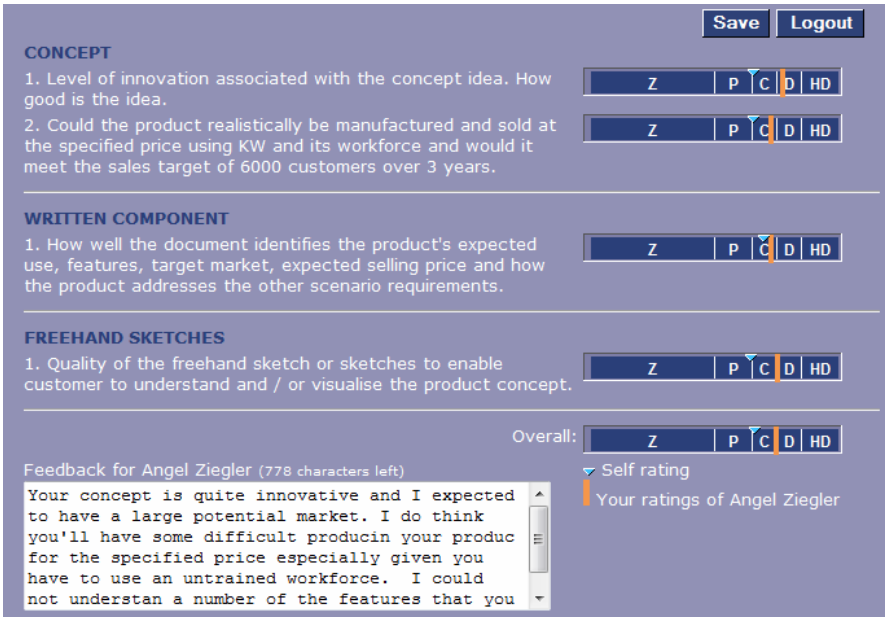

Figure 1. Students enter their assessments for their peers by moving the sliders to their chosen rating against each criterion. To assist students to make comparative judgements when marking their team peers (orange bars) their self assessment of their own submission for each criterion is displayed by the upper (blue) triangle

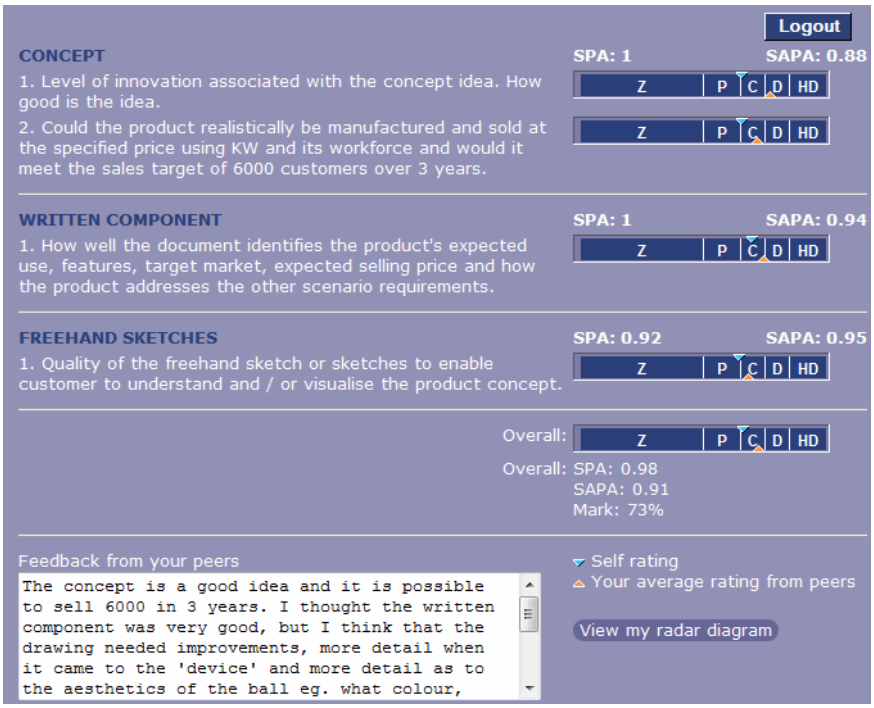

Figure 2. Student's result screen: Once published students may logon and see their results. The upper (blue) triangle shows a student's self rating against each criterion while the bottom (orange) triangle shows the average rating received from their peers.

A student's score is automatically calculated by combining the rating submitted by the student (self-assessment) and each of their team peers (peer assessment).

Once the exercise is complete and the results are published students may logon to receive their score/grade for the exercise. In addition, students are provided with the feedback in regard to their submission against each criterion. The upper blue triangle shows the student's self rating of their own submission while the lower orange triangle shows the average rating for each criterion submitted by their team peers. In addition a feedback box displays written comments from their team peers (Figure 2). Further feedback is provided by rating factors and radar diagrams which summarises a student's results. These diagrams are easily uploaded to their e-portfolio.

Method

The self and peer assessment of individual submission activity was implemented in the subject Design Fundamentals in the Spring semester of 2008. Design Fundamentals is a second year subject taken by all engineering students at the University of Technology, Sydney (UTS). The subject's typical cohort is approximately 300 students. Tutorial classes are limited to 32 students with project teams consisting of 4 students.

The subject's primary aims are to:

- Develop students' understanding of the engineering design process
- Provide students with the skills to develop a small engineering project from initial concept to the production of a prototype.
- Continue the development of students' generic professional skills including teamwork, critical evaluation, feedback and communication commenced in earlier subjects.

The activity consisted of a series of distinct processes:

1. Students were required to use SPARK^{PLUS} to assess their own submission and submissions from seven of their peers. This assessable part of the overall task was completed individually by students outside of class.

2. In the following tutorial the group of eight students debated the merits of each individual submission (discussing their individual strengths and weaknesses) and collectively placed them in order from best to worst awarding a mark for each.

3. Students then received the results from SPARK^{PLUS} (as shown in Table 1) and were asked to reflect on any differences between the results produced from their individual assessments (SPARK^{PLUS}) and those produced collectively in their peer group.

4. The tutor marked the best report from each group (as identified by the students in Stage 1) using it as an exemplar to discuss the assignment with the group. The marks for the other reports were determined using the weighting produced by SPARK^{PLUS}. Hence tutors only marked one in eight individual submissions.

The order of the activities within this task meant that students were required to engage in some individual thinking about the assessment criteria for the project concept before meeting in the tutorial to discuss the submissions with their peers. As a result most students came to class prepared, allowing discussions to quickly focus on areas where there was a difference of opinion. While not directly assessable, we specifically designed steps 2 and 3 to involve collaborative reflection with the expectation that facilitating students to explore differences in their opinions and understanding would make a major contribution to their learning.

The motivation to actively participate in the activity was on two levels. Groups were required to select one of the project concepts to work on for the rest of the semester, so choosing the concept that was the 'best' fit with the overall project constraints would both simplify their subsequent tasks and potentially provide them with the highest grade. More immediately, it was in the group's interest to correctly identify the 'best' concept to maximise their mark as the tutor only marks the concept identified as being the best by the group. The mark awarded to this concept caps those allocated to the remaining submissions which are calculated in proportion to the ratings calculated by SPARK^{PLUS}.

After the activity students were asked to complete a questionnaire to assess the effectiveness of the activity to both promote collaborative peer learning and motivate students to achieve the content learning outcomes. The survey questions were a mixture of free response and Likert format (4 point). While all students undertaking the project were required to participate in the assessment exercise, in accordance with our ethics approval, participation in the survey was voluntary. The survey was conducted in tutorial classes resulting in 209 (eligible cohort 256) students responding (82% response rate).

Results and analysis

Results from the student responses to the survey are plotted in Figures 3, 4 and 5.

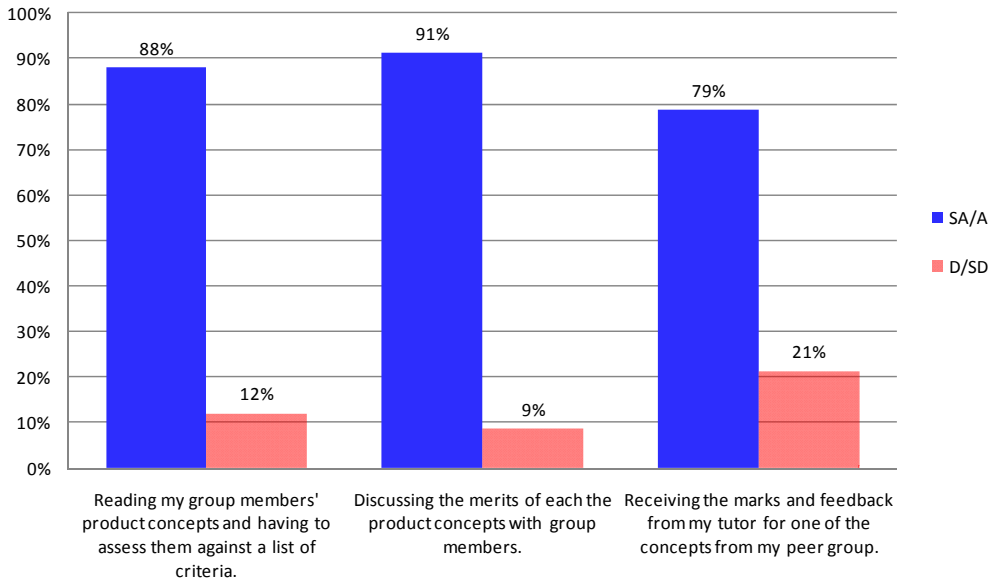

Figure 3. Results from student survey of Self and Peer Assessment Marking of Individual Project Concepts in response to the question "My ability to choose a product concept and write a concept document to meet a list of requirements increased as a result of:"

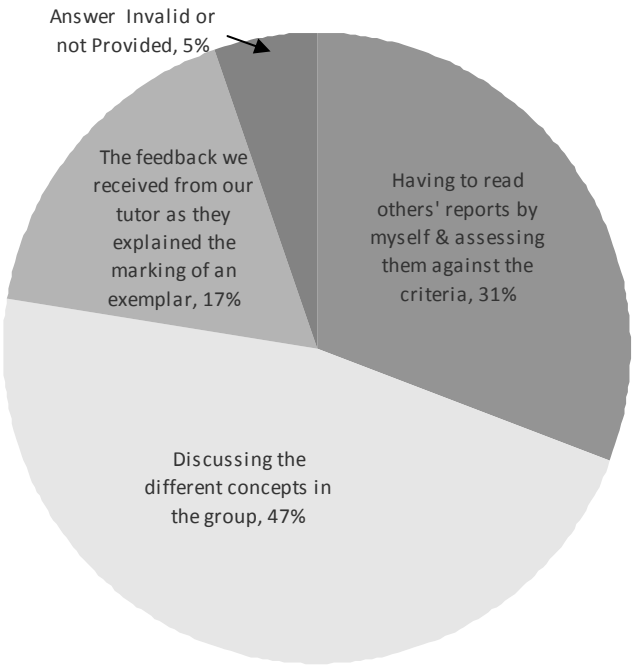

Figure 4. Results from student survey in response to the question: "Which part of the whole process improved your understanding / ability the most?"

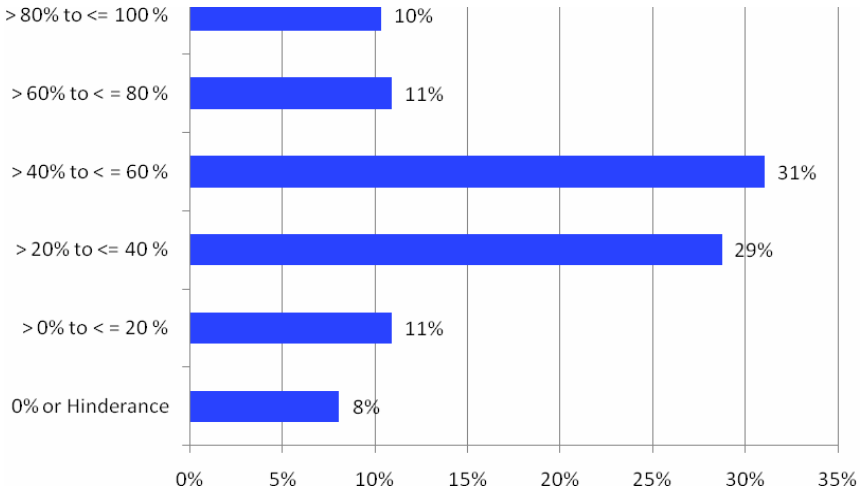

Figure 5. Student's expected percentage improvement in their possible mark for the individual project concept if they had to do it again after completing the activity.

While the results show that the majority of students (ranging from 79% to 91%) felt that all stages of the exercise improved their ability to achieve the prescribed learning outcomes (Figure 3), 47% rated the collaborative group discussions as being the most effective in helping them to learn (Figure 4). This was followed by reading the reports themselves and assessing them against the criteria (32%), with only 17% of students agreed that their tutor's explanation of their marking of the best report was the major contributor to their learning (Figure 4). This result is probably in part a function of the tutor feedback occurring after students had already learnt through completing the individual and collaborative stages of the exercise. Thus the tutor was only required to cover the issues not already addressed by the students. However the fact remains that 83% of students reported their most effective learning occurred without direct academic participation (Figure 4).

The survey also asked students *"Now after receiving feedback if you had to do the individual concept component of the project again, what mark do you think you could achieve?"*. In interpreting these results care needs to be taken not to introduce bias as a result of the following:

1. The lower a student's initial mark the greater the potential for improvement.
2. Academics typically do not assess using a linear scale. For example, the level of improvement required to achieve an increase of 10 marks from 50 to 60 is generally lower than that required to increase a mark from 80 to 90.
3. The higher your mark, the higher the increase required to achieve a designated percentage improvement. For example a 20% improvement in a submission that received 50% is only 10 marks, while a 20% improvement in a submission that received 80% is 16 marks.

After careful consideration we chose to use the percentage of available improvement (Equation 1), explained in the following example, to analyse this data. For instance, a student who received 70/100 for their individual submission has a possible mark increase of 30. If after the exercise the student expected to improve their mark to 85/100 this equates to achieving 50% of the possible improvement available.

Similarly, if a student received 40/100 for their individual submission, they have a possible mark increase of 60. If after the exercise the student expected to improve their mark to 70/100 this equates to achieving 50% of the possible improvement available. Thus we are estimating that the learning required by a student who before the exercise, knew enough to receive 70/100 to increase their mark to 85/100, is equivalent to the learning required by a student who before the exercise knew enough to receive 40/100 to increase their mark to 70 /100.

$$\%\ of\ available\ improvement = (expected\ mark - received\ mark) / (100 - received\ mark) \quad (1)$$

The results of this analysis, reported in Figure 5 found that 31% of students reported they expected their learning from the exercise would enable them to achieve 40% to 60% of the available improvement. Furthermore, 81% of students expected their learning from the exercise would enable them to achieve at least 20% of the available improvement.

Discussion

In designing this activity we considered that students who scored well by themselves may not learn as much from the collaborative components of the exercise as students that received lower scores. To investigate this we cross-referenced the grade students received for their submission with how many of the four stages within the exercise contributed at least 15% to a student's total improvement in their understanding / ability to meet the exercise learning outcomes. The results of this analysis are reported in Figure 6 for the entire cohort and then according to the grade they received for their individual submission (High Distinction HD Distinction D, Credit C, Pass P, Fail Z)

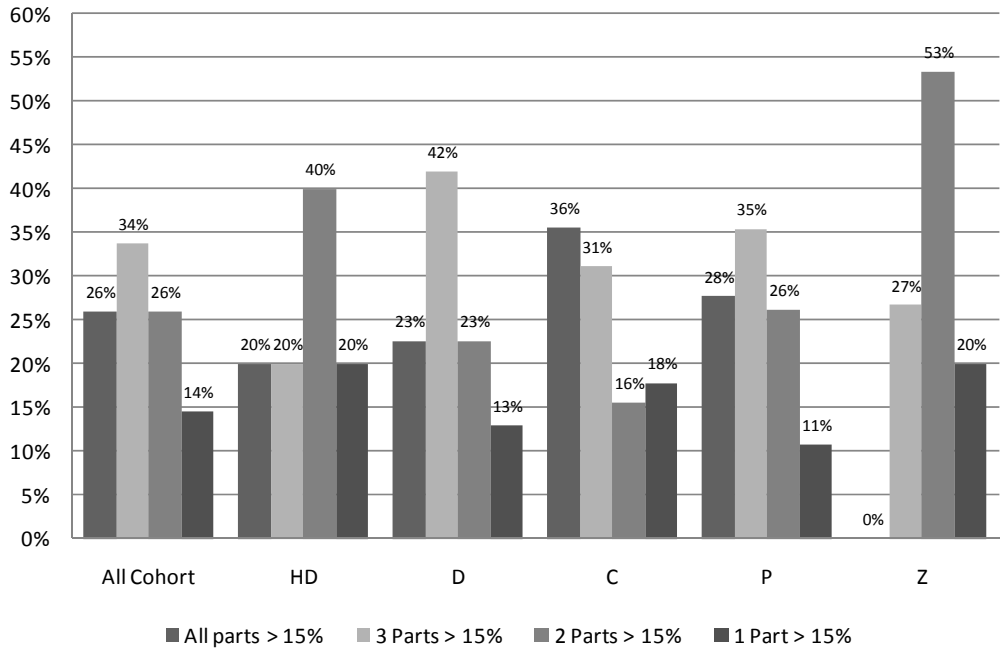

Figure 6. The Number of different parts of the exercise that contributed more than 15% to a student's overall improvement in their understanding/ability in relation to the activity learning outcomes, for the entire cohort and for individual grades.

Overall 60% of students found that at least three different parts of the exercise each contributed more than 15% to the overall improvement in their understanding and ability. This is also the case for students who received a distinction (D) (65%), credit (C) (67%) or pass (P) (63%). However, only 40% of students who received a high distinction (HD) and 27% of students who received a fail (Z) for the individual project concept found that at least three parts of the exercise each contributed more than 15% to their learning. Furthermore, not one student that failed the exercise found all four parts contributed at least 15% to their overall learning.

To explore these results further we calculated the percentage of students within each grade category that found each part of the exercise contributed at least 15% to their overall understanding. These results are shown in Figure 7.

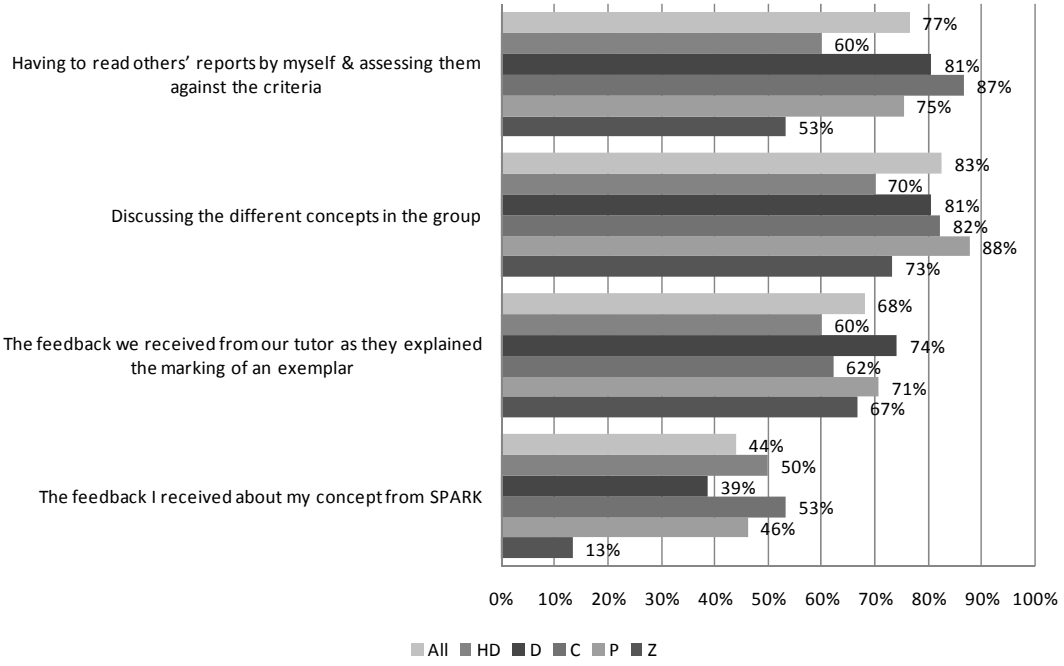

Figure 7. The number of students for each grade that found a part of the exercise contributed more than 15% to their overall improvement in their understanding/ability in relation to the activity learning outcomes.

Firstly, in all grade categories more students (>70%) found the collaborative discussions contributed at least 15% to their learning than any other part of the exercise. This result supports those previously reported in Figures 3 and 4 demonstrating the benefits of collaborative peer learning activities in assisting students to learn.

Secondly, not surprisingly students who received a failed grade (Z) for the individual project concept clearly preferred the parts of the exercise that did not rely solely on their judgement with 73% reporting that the group discussions and 67% the feedback received from their tutor contributed at least 15% to their overall learning. These parts of the exercise involve students receiving feedback and explanations from tutors and students who typically received higher grades than themselves. In comparison only 53% of students receiving a fail grade (Z) reported that assessing the concepts themselves and 13% the feedback they receive from SPARK contributed more than 15% of their overall learning. These two parts of the exercise require students to use their own judgement in both assessing the individual submissions and in interpreting the feedback factors produced by SPARK^{PLUS}.

Thirdly, while HD students also found the collaborative discussions to be the most beneficial (70%) compared to assessing the reports themselves (60%), feedback from the tutor in regard to the exemplar (60%), and the feedback they received from SPARK^{PLUS} (50%) except for the latter (feedback from SPARKPLUS) these are smaller percentages than any other passing grade (D, C and P).

Further analysis of the survey results for HD students showed that their learning tended to be concentrated to fewer parts of the exercise than students who passed the product concept with a lower grade. That is, if a task within the exercise contributed to a HD student's learning, then it typically contributed a lot, often > 50%. Hence those tasks deemed to have contributed less, were often considered to have contributed only a small amount, typically < 15%. It is interesting to note that although the learning for the HD students tended to be concentrated into fewer tasks than for other students, it was not the same combination of tasks for all HD students. In other words they did not all find the same tasks within the exercise useful.

In interpreting this result one needs to consider the role that HD students play in the collaborative exercise. Firstly there are only 10 respondents who received a HD for the exercise but there were more than 30 groups. Hence it is quite possible that a HD student didn't receive quality feedback from their group, as they were most likely the participant with arguably the 'best' knowledge. Furthermore HD students may not value or have confidence in the feedback they receive from their peers. This being the case some students may have

felt that the collaborative part of the exercise contributed only a small amount to their overall learning. However, HD students are also in the best position to adopt the role of a teacher and in explaining their thoughts or providing feedback to others they had an opportunity to consolidate and improve their own learning and understanding. Those students that adopted this teaching role may understandably have rated the collaborative discussions as the most effective part of the exercise in improving their understanding / learning. Similarly, there is a high probability that the HD student's concept was the one assessed and used as an example by the tutor when giving their feedback to each group. If this was the case it is understandable that this part of the exercise may have contributed a high percentage of the learning for these students.

Overall we were successful in designing an activity integrating self and peer assessment with the principles of learning oriented assessment to produce a minor assessment task that motivated significant learning. The integration of the different assessment processes meant that the exercise accommodated students' different learning styles. In addition, the collaborative learning component and the fact that students were mostly responsible for providing their own feedback and assessment required them to take responsibility for their own learning.

However, there are a number of factors that still need to be addressed. A minority of students felt that students should not be involved with assessing each other's work. This was due to a number of factors including they didn't think they had the knowledge or skill to make fair assessments, they had no confidence in the assessments or feedback provided by their peers or they thought it was not their responsibility and hence they should not be required to do it. These attitudes were evident in the following free response comments:

> "There should be less emphasis on other students marking your concept. I found some people were lazy and did not give each concept equal time to mark and overlooked some. It should be more a person who thoroughly knows the marking criteria (ie tutor) who has more weighting."
>
> "Students mark should be based on the mark given to them by their tutors. This is what we are paying them to do."
>
> "An individual project concept is individual work. The individual pays $2700 to enrol in this subject. The individual mark should be determined solely by a tutor or other person qualified to mark individual work."

There was also a group of students who while admitting they learnt a lot from the exercise were disappointed that they could not use this learning to improve their project concept and resubmit it for marking in the hope of achieving an improved grade. This attitude is typified by the free response comments below:

> "…. I learned a lot about making the concept general and not putting implementation details in, after I'd already completed and submitted my concept, so different timing on lectures may help."
>
> "I learned a lot from the project concept review, but this occurred after the assessable individual product concept was submitted."

In designing the exercise the authors deliberately chose not to allow resubmission. In the first instance this would significantly increase the burden on the academic staff. We also doubted that students would put in their best effort on their initial submission if they were allowed to resubmit it after receiving feedback from their peers. Additionally, given the competitive nature of students we suspected some may not wish to provide quality feedback if it could help one of their peers to achieve a higher grade than their own. More importantly we wanted the focus to be on learning rather than how to strategically achieve the best mark.

While the survey did not directly ask for positive free response comments, that is it did not ask students what they liked about the process (the authors recognise this as a serious omission), many students still took the opportunity to report what they liked about the exercise. An example of these comments are reported below:

> "I could figure out the flaws in my concept. I could get to know certain things I didn't think about while designing the product"
>
> "Tutorial was beneficial because it allowed 2 different groups to assess project concept. It was a good way to review because a variety of answers were provided towards the Individual Concept."

While we recognise the results of this initial implementation are not definitive, the following points are clear:

- we were successful in using a minor assessment task which was completed outside of class time to motivate students to achieve considerable learning. The collaborative discussions were not directly assessed yet students perceived that this was the most effective part of the exercise in improving their understanding and ability.
- the fact that students found different parts of the exercise to be most useful to them supports the aim of designing assessment tasks with multiple opportunities for learning to accommodate different learning styles.
- the authors are of the opinion that if the collaborative discussions had been left to students to conduct themselves or were held outside of scheduled classes both student attendance and participation would have been significantly reduced, in turn reducing the effectiveness of the exercise. In designing the task we deliberately chose to have these collaborative discussions conducted during scheduled class (tutorial) time and the individual components conducted out of class.
- SPARK^{PLUS} allowed the instructors to conduct the exercise with very little academic effort. The set up overheads were minor and the program automates the distribution of feedback and results.

Conclusion

Creating assessment tasks to cover all we want students to learn may impose a significant workload for academics (especially in large classes) and over assessment of students.

In this paper we reported how a small assessment task followed by collaborative phases of a learning exercise, promoted significant learning. The vast majority of students reported that most of their learning occurred in the collaborative parts of the exercise, even though these were not directly assessed. In addition, the exercise encouraged students to take responsibility for their own learning by requiring them to explore their understanding and provide their own feedback. Furthermore, the integration of different reflective processes within the exercise accommodated different learning styles.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5 (1), 7.

Boud, D., & Falchikov, N. (2007). Rethinking Assessment in Higher Education Learning for the Longer Term. Routledge.

Brown, S. (2004). Assessment for Learning. Learning and Teaching in Higher Education, 1, 81-89.

Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical implications. Innovations in Education and Teaching International, 44 (1), 57-66.

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. Review of Educational Research, 70 (3), 287-322.

Freeman, M., & McKenzie, J. (2002). SPARK, A Confidential Web-Based Template for Self and Peer Assessment of Student Teamwork: Benefits of Evaluating across Different Subjects. British Journal of Educational Technology, 33, 551-69.

Gibbs, G., & Simpson, C. (2004). Conditions Under Which Assessment Supports Students' Learning. Learning and Teaching in Higher Education, 1, 3-31.

Hanrahan, S., & Isaacs, G. (2001). Assessing Self- and Peer-assessment: the students' views. Higher Education Research & Development, 20 (1), 53-70.

Hargreaves, D. J. (1997). Student Learning and Assessment are inextricably linked. European Journal of Engineering Education, 22 (4), 401-409.

Keppell, M., & Carless, D. (2006). Learning-oriented assessment: a technology-based case study. Assessment in Education, 13 (2), 179-191.

Keppell, M., Au, E., Ma, A., & Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. Assessment & Evaluation in Higher Education, 31 (4), 453-464.

Mello, J. A. (1993). Improving individual member accountability in small work group settings. Journal of Management Education, 17 (2), 253-259.

Michaelsen, L., Knight, A., & Fink, L. (2004). Team-based Learning – A transformative use of small groups in college teaching. USA: Sylus Publishing.

Ramsden, P. (2003). Learning to Teach in Higher Education (2nd ed.). London: Routledge.

Somervell, H. (1993). Issues in assessment, enterprise and higher education: the case for self-, peer and collaborative assessment. Assessment & Evaluation in Higher Education, 18, 221-233.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. Assessment in Education, 14 (3), 281-294.

Willey, K., & Freeman, M. (2006a). Completing the learning cycle: The role of formative feedback when using self and peer assessment to improve teamwork and engagement. In Proceedings of the 17th Annual Conference of the Australasian Association for Engineering Education. Auckland, New Zealand.

Willey, K., & Freeman, M. (2006b). Improving teamwork and engagement: the case for self and peer assessment [Electronic version]. Australasian Journal of Engineering Education. Retrieved from http://www.aaee.com.au/journal/2006/willey0106.pdf.

Willey, K., & Gardner, A. (2008a). Using Self Assessment to Integrate Graduate Attribute Development with Discipline Content Delivery. In Proceedings of the 36th Annual Conference of the European Association of Engineering Education (SEFI). Aalborg, Denmark.

Willey, K., & Gardner, A. (2008b). Improvements in the self and peer assessment tool SPARK: Do they improve learning outcomes? In Proceedings of the ATN Assessment Conference – Engaging Students in Assessment. South Australia.