Improving Self and Peer Assessment Processes with Technology.

Keith Willey
Faculty of Engineering and IT, University of Technology, Sydney; Keith.Willey@uts.edu.au

Anne Gardner
Faculty of Engineering and IT, University of Technology, Sydney; Anne.Gardner@uts.edu.au

Abstract

Purpose - As a way of focusing curriculum development and learning outcomes Universities have introduced graduate attributes which their students should develop during the course of their degree. Some of these attributes are discipline specific, others are generic to all professions. The development of these attributes can be promoted by the careful use of self and peer assessment. The authors have previously reported using the self and peer assessment software tool SPARK in various contexts to facilitate opportunities to practise, develop, assess and provide feedback on these attributes. This research and that of the other developers identified the need to extend the features of SPARK, to increase its flexibility and capacity to provide feedback. This paper reports the results of the initial trials to investigate the potential of these new features to improve learning outcomes.

Design/methodology/approach – This paper reviews some of the key literature in regard to self and peer assessment, discusses the main aspects of the original online self and peer assessment tool SPARK and the new version SPARK^{PLUS}, reports and analyses the results of a series of student surveys to investigate whether the new features and applications of the tool have improved the learning outcomes in a large multi-disciplinary Engineering Design subject.

Findings – We found that using self and peer assessment in conjunction with collaborative peer learning activities increased the benefits to students and improved engagement. Furthermore we found that the new features available in SPARK^{PLUS} facilitated efficient implementation of additional self and peer assessment processes (assessment of individual work and benchmarking exercises) and improved learning outcomes. The trials demonstrated the tool assisted in improving student's engagement with and learning from peer learning exercises, the collection and distribution of feedback and helping students to identify their individual strengths and weaknesses.

Practical implications – SPARK^{PLUS} facilitates the efficient management of self and peer assessment processes even in large classes allowing assessments to be run multiple times a semester without an excessive burden for the coordinating academic. While SPARK^{PLUS} has enormous potential to provide significant benefits to both students and academics we caution that although a powerful tool, its successful use requires thoughtful and reflective application combined with good assessment design.

Originality/value – We found that the new features available in SPARK^{PLUS} efficiently facilitated the development of new self and peer assessment processes (assessment of individual work and benchmarking exercises) and improved learning outcomes.

Article type – Research Paper
Keywords - self and peer assessment, groupwork, professional skills, graduate attributes, SPARK, SPARK^{PLUS}, benchmarking.

Improving Self and Peer Assessment Processes with Technology.

1. Introduction

Freeman and McKenzie (2002) reported the development of a confidential online tool called SPARK (Self and Peer Assessment Resource Kit), to both collect self and peer ratings and calculate an adjustment factor to convert group marks into an individual mark for a project. SPARK:

1. Solved most of the administrative issues associated with paper-based approaches such as data collection and analysis.
2. Enabled students to confidentially rate their own and their peers' contributions to a team project against criteria which can be written to include specific project tasks and/or good team practices.
3. Automatically generates both formative and summative assessment factors. The Self and Peer Assessment or SPA factor (see Equation 1) is a weighting factor that can be used to determine an individual's contribution to a team project as shown in Equation 2. A SPA factor of 1 indicates a student's contribution was rated as being equal to the average contribution of their team.

$$\text{SPA Factor} = \sqrt{\frac{\text{Total ratings for individual team member}}{\text{Average of total ratings for all team members}}} \quad (1)$$

$$\text{Individual mark} = \text{team mark} \times \text{Individual's SPA} \quad (2)$$

The second factor is the Self Assessment to Peer Assessment or SAPA factor (Equation 3). It is the ratio of a participant's own rating of themselves compared to the average rating of their contribution by their peers. For example a SAPA factor greater than 1 means that a student has rated their own performance higher than they were rated by their peers.

$$\text{SAPA Factor} = \sqrt{\frac{\text{Self ratings for individual team member}}{\text{Average of ratings for individual by peer team members}}} \quad (3)$$

As a result of research by the authors and other developers (see acknowledgement) we identified a number of features to increase SPARK's flexibility and capacity to provide feedback. This article compares the results of the first trials of these new features to trials using the original version of SPARK to investigate their potential to improve learning outcomes.

2. Background

It is often difficult for an academic to fairly assess the contribution of individual students to a team project, especially if most of the work has occurred outside of scheduled lecture or tutorial times. Self and peer assessment is often used as a means of handing over assessment of an individual's contribution to a team task to the team members themselves (Johnston & Miles, 2004). In addition to providing fairer assessment, self and peer assessment is reported as assisting students to develop important professional skills including reflection and critical thinking (Mello, 1993, Somervell, 1993). The literature also examines different methods of using self and peer assessment to derive individual grades for team assessments (Lejk et al, 1996, Goldfinch & Raeside, 1990, Goldfinch, 1994). Typically team members evaluate

themselves and each other. These evaluations are then incorporated into an individual student's assessment, either as an addition to the team assignment mark, or by adjusting the team assignment mark to produce an individual mark. Falchikov & Goldfinch (2000) compared peer and teacher marks and found that peer assessments more closely resemble teacher assessments when global judgements based on well understood criteria were used compared to marking that involved assessing several individual dimensions. They also reported a link between high quality design of assessment tasks and more valid peer assessments, a view supported by Freeman and McKenzie (2002). A great deal of the published research has focused on using self and peer assessment for discouraging group free riders ('free-riders' also known as 'passengers' are team members whose contribution is insufficient, inadequate and/or poor in comparison to their team peers), promoting team collaboration and for summative purposes. More recently Willey & Freeman (2006a) report using it to produce formative learning-oriented feedback to complete the learning cycle, promote academic honesty, encourage the ongoing development of skills and improve the quality of subsequent contributions, while Boud & Falchikov (2007) discuss its use for developing students' skills for lifelong learning.

As a way of focusing curriculum development and learning outcomes Universities have introduced graduate attributes which their students should develop during the course of their degree. Some of these attributes are discipline specific, others are generic to all professions. Similarly attributes are generally required by professional organisations to obtain degree program accreditation or for their members to obtain professional accreditation. For example Engineers Australia Accreditation Policy (Engineers Australia, 2004) lists attributes divided into three Units of Competency:

- Unit 1: Knowledge Base
- Unit 2: Engineering Ability
- Unit 3: Professional Attributes

Willey and Gardner (2008a) report the potential for using self and peer assessment to both assess and track graduate attribute development not only within a subject but potentially throughout a whole degree program.

The fact that SPARK is a criteria-based tool allows academics the flexibility to choose or create specifically targeted criteria to allow any task or attribute to be assessed. In addition, using common categories (like the three described above) throughout a degree program, to which academics link their chosen criteria, allows the results to be recorded, for example in an e-portfolio, providing a means for both academics and students to track students' development as they progress through their degree.

Receiving feedback is a crucial component of any learning process. To be productive and useful for student reflection, feedback must be both timely and focused. The thoughtful integration of SPARK allows feedback to be provided efficiently, even in large classes, by enabling students to receive feedback from their team peers.

The SPARK metrics are shared between all group members in feedback sessions to assist learning. Students are guided on how to both reflect on their own performance and learning, and to give constructive feedback to their team peers (Willey & Freeman , 2006b, Willey & Gardner, 2008b). The process focuses on improving students' judgement (Boud & Falchikov, 2007) moving them to be more expert in their ability to engage with subject learning outcomes. In addition, students are providing feedback, an essential ingredient of peer learning, with only a comparatively small implementation effort required by the coordinating academic.

As a result of research by the authors and the other developers a number of features were identified that would increase SPARK's flexibility, capacity to provide feedback and improve student learning outcomes. These features include:

- The ability for students to anonymously provide written feedback to their peers.
- The ability for both the SPA and SAPA factors to be either overall (assessment of the students' overall contribution or grade) or category based where the assessment and feedback is reported over a number of different categories or attributes.
- Provision for feedback on all assessment criteria so students can see how they rated themselves compared to the average rating they received from their team peers.

More recently as a result of ongoing research the authors have added additional features including:

- A graphical interface to assist interpretation of results.
- The addition of different formulas and rating scales to increase SPARK's flexibility to facilitate assessment tasks with different objectives.
- The ability for students to self and peer assess and receive / provide feedback on their and their peer's individual submissions.
- The ability for students, academics and / or tutors to benchmark their judgment, assessment and evaluation skills.
- A third factor being a percentage mark/grade, the calculation of which depends on the type of task that has been selected (e.g. benchmarking exercise or marking individual work).

3. SPARK Improvements

SPARK^{PLUS} allows students to provide anonymous written feedback comments to their peers. An example of this feature can be seen in Figure 1 which shows a partial screen shot of the results screen for a student named George. This screen shot also displays the overall assessment factors and the assessment factors for the three Engineers Australia categories previously described (note the assessment criteria within each category are hidden behind the radar diagram). These factors not only allow students to identify their strengths and weaknesses but enables academics to provide specific coaching to assist students to improve their performance in identified areas of weakness.

For example referring to Figure 1 where the three categories used (in the order they appear in the Figure) were *Knowledge Base, Engineering Ability and Professional Skills* and the assessment rating key was WB - well below average to WA- well above average:

Take in Figure (1) Paper 1

The overall (aggregate) performance factor (0.97) indicates that George is performing well, contributing only slightly lower than the average performance of his team peers. The formative feedback or SAPA factor of 1.18 indicates that the George's opinion of his own performance is significantly greater than the average opinion of his performance by his team peers (Shown in radar diagram by SAPA envelope exceeding 1). More insight is gained by looking at the individual factors for each attribute category. These report that:

George's contribution to the team's required Engineering Knowledge is only slightly below the average of his team peers (SPA = 0.96). Perhaps more importantly the SAPA factor of 1.27 indicates that George does not share his peers' opinion of his performance, believing he has contributed much more to this category. The reasons for this difference in opinion would generally be explored in a feedback session, and maybe attributed to a number of factors including:

1. George's real contribution has not been fairly assessed by his peers.
2. George's peers have not provided feedback to George in regard to his performance and hence he is unaware of the differences between his self and his team peers perceptions.
3. George may be aware of his performance level but deliberately chose to inflate his ratings to increase his overall mark.

Conversely George's contribution to the team in regard to Engineering Ability is equivalent to the average of his team peers (SPA = 1). Furthermore the SAPA factor of 0.99 indicates that George's assessment of his contribution is shared by his peers.

Hence analysis of the category factors indicates that George's strength in this project was his Engineering Ability, while his weaker areas of contribution were his Engineering Knowledge and Professional Skills. Furthermore the category SAPA factors suggest that prior to receiving this feedback George may have been unaware that he was underperforming in these areas.

The factors provide feedback for George to reflect on his performance and for the whole team to reflect on their assessments. Early stage students are guided in the interpretation of their results during tutor facilitated feedback sessions. In these sessions students are guided through a feedback process to assist them in identifying their strengths and weaknesses, encourage feedback between team members and reaching an agreement on how the team can improve their performance for the remaining assessment tasks (Willey & Freeman , 2006b, Willey & Gardner, 2008b).

Two of the assessment graphical feedback options are also shown in Figure 1. The student's radar diagram reports the feedback factors for each category identifying a student's strengths and weaknesses. Performance in a particular category is depicted by the position of the SPA factor envelope compared to 1, while the SAPA envelope identifies any discrepancies between a student's self perceptions and the perceptions of their performances by their peers. The recording of these diagrams in an e-portfolio allows students to track their attribute development throughout their degree program.

An instructor can also choose to provide students with feedback in regard to the differences between their own self assessments and the average assessment of their performance by their peers for each individual criterion. In Figure 1 the upper (blue) triangle shows the student's self rating for each individual criteria, while the lower (orange) triangle shows the student's average rating for each criteria received from their team peers, providing students with detailed formative feedback on their performance in regard to each individual criterion.

Take in Figure (2) Paper 1

SPARK^{PLUS} allows the use of different formulas to accommodate the design of assessment tasks with different objectives. For example Figure 2 shows the relationship between the two

new and original (Equation 1) method of calculating the SPA factor: Linear SPA = (Equation 1) ^{2}, Knee SPA = (Equation 1) ^{2} if ≤ 1 or Equation 1 if > 1.

SPARK^{PLUS} also allows the use of different rating scales (unsatisfactory to high distinction, well below average to well above average etc) to accommodate the design of assessment tasks with different objectives.

Take in Figure (3) Paper 1

Students can also use SPARK^{PLUS} to assess and receive / provide feedback on their own and their peer's individual submissions against a number of specified criteria or to benchmark their judgment, assessment and evaluation skills. Figure 3 shows a student's results screen for an exercise in which students were required to mark their own and their group peers' individual submissions. Note the rating scale now reflects grades Z (unsatisfactory) to HD (High Distinction) and in addition to the SPA and SAPA factors the student receives an overall mark in this case 73%. Figure 4 shows a radar diagram reporting the overall results for one group for this exercise. These Radar diagrams are distributed to groups in feedback sessions that are integrated into peer learning activities. Note how Andrew and Fran rated their own submissions much higher than the average rating of their submission by their team peers (SAPA > 1). In comparison, James and Ed whose contributions were rated the best by the group (highest marks and SPA factors) rated their contributions lower than they were rated by their team peers (SAPA < 1).

Take in Figure (4) Paper 1

4. Method

In this article we compare the results from two trial implementations, in which self and peer assessment was used to assess, assign marks and provide feedback on a combination of both discipline specific and generic professional attributes. The first trials were conducted using the original version of SPARK in Autumn semester 2007. The second series of trials conducted in Spring semester 2008 used SPARK^{PLUS}. In both cases the trials were run at the University of Technology, Sydney (UTS), within the subject Design Fundamentals. Design Fundamentals is a second year subject undertaken by all engineering students at UTS. The subject's aim is to develop students' understanding of the engineering design process and to provide them with the skills to develop a small engineering project from initial concept through the design stages of requirements analysis, system design, and detailed design to the development of an alpha prototype. In addition the subject aims to build on students' skill and generic attribute development begun in the first year of their degree. Students are required to work in teams of 4 to produce a prototype product, two reports and make an oral presentation. The group assessment represents 50% of a student's overall grade.

Self and Peer assessment was integrated into distinct peer learning assessment tasks that combine to form a major design project. In Autumn 2007 using the original version of SPARK only task 3 and 4 (see below) were used. With the advent of SPARK^{PLUS} in Spring 2008 tasks 1 and 2 were added. The tasks were as follows:

1 Assessment of Individual Submissions: Outside of class students use SPARK^{PLUS} to assess their own and seven of their peer's individual project concept submissions rating each

student's work against a number of specified criteria (approx 1.5 hrs). In the next tutorial (2 to 3 hours) the group of eight students debate the merits of each individual submission (discussing their individual strengths and weaknesses) and collectively place them in order from best to worst awarding a mark for each. Students then receive the results from SPARK^{PLUS} and are asked to reflect on any differences between results produced from their individual assessments (SPARK^{PLUS}) and those produced collectively in their peer group. The tutor marks the best report from each group (as identified by the students) and determines marks for the other reports using the weighting produced by SPARK^{PLUS}.

The peer learning groups are divided into two groups of four students. These groups of four students then work together to complete the design project.

2 Benchmarking Exercise: Students are provided with a Sample Requirement Specification produced by a student group from a previous semester. After discussing the marking criteria each student has to individually assess the report using SPARK^{PLUS} (approx 45 minutes). In their next tutorial (approx 2 hours) each group of four students discuss their marking of the report and re-mark it collectively against the criteria. Students then re-combine into their peer learning groups (two groups of four students) and discuss their group's marking of the report, reflecting on any differences and collectively re-mark it. Tutors then discuss how they marked the report. After the tutorial students may log on to SPARK^{PLUS} and compare their individual marking to the tutor's marking of the report for each individual criterion. In addition, SPARK^{PLUS} produces a mark (several formulas and moderation methods are available) related to how close the student's individual assessment was to the academic's assessment.

3 & 4 Assessment of Individual's Contribution to a Team Task: each group of students produces two reports, makes an oral presentation and produces a prototype as part of their design project. Students use SPARK^{PLUS} to rate their own and their team peers' contribution to the project which was divided into two stages. The SPARK^{PLUS} SPA factors are used to produce individual marks by moderating the mark for the group's submission. The group's radar diagrams and a table of categorised factors (similar to Figure 4) are distributed to each group and discussed in the next tutorial (only the table of factors were distributed in the Autumn 2007 trials). Groups are guided through a feedback process. This process begins with students sharing positive feedback with the focus not just being on what their peers did well but also on what they learnt from their peers. This is followed by a process of self evaluation where students share with their group what they have learnt or discovered about their strengths, weaknesses or performance from the exercise. Students are encouraged to identify how they could improve their own performance and in what way they would approach the task differently if they had to do it again. The final stage in the feedback process is the provision of constructive criticism to team peers. Students are asked to suggest how others in their group may have approached their tasks differently to achieve a better group result, how aspects of their behaviour affected the team and the benefits of changing that behaviour and to reflect on how team peers could have learnt more from the process. Furthermore, students are asked to share what they consider to be the weaker aspects of a peer's contribution and how this could have been improved.

The in-class discussion (approx 1 hour) concludes by teams agreeing how to improve their overall team and individual performance for the remaining parts of the project and /or in future group work opportunities.

In both the semesters a number of subject surveys were conducted to assess the effectiveness of the self and peer assessment processes used. The questions in all surveys were a mixture of free response and Likert format (5 point in Autumn 2007, 4 point in Spring 2008).

While all students undertaking the project (eligible cohort 220(2007) and 256 (2008)) were required to participate in the assessment exercises, in accordance with our ethics approval, participation in the surveys was voluntary. The 2008 surveys associated with the Assessment of Individual Submissions and Benchmarking exercises were conducted in tutorial classes resulting in 209 and 201 students responding respectively. The 2007 and 2008 surveys associated with the Assessment of Individual's Contribution to a Team Task (tasks 3 and 4) were conducted as a post subject survey. The post subject survey was much longer, covered a number of topics, was conducted online just before the exam period and took students at least 30 to 45 minutes to complete. Of the eligible cohort of 220 (2007), 256 (2008), 95 (37%) (2007), 83 (32%) (2008) students volunteered to complete the online survey.

5. Results

The survey results relevant to this article are shown in Figures 5 – 9. Where applicable the 'Strongly Agree' and 'Agree' responses were combined to give an aggregate result (SA/A), as were the 'Strongly Disagree' and 'Disagree' responses (D/SD). The percentage of any unanswered responses are generally not shown but can be calculated by subtracting the provided results from 100%.

Take in Figure (5) Paper 1

Take in Figure (6) Paper 1

Take in Figure (7) Paper 1

Take in Figure (8) Paper 1

Take in Figure (9) Paper 1

6. Discussion

The advent of SPARK^{PLUS} allowed the introduction of individual submission assessment and benchmarking exercises to the subject. While it would have been possible to partially introduce these exercises with other software (eg Calibrated Peer Review (2009)) or using a paper-based approach the academic overhead in our particular case would have been excessive. Hence we have no pre-SPARK^{PLUS} results to compare for these two exercises.

6.1 Effect on students

The results presented in Figures 5 and 7, report that the majority of students (ranging from 79% to 91%) felt that all aspects of the group marking of individual submissions and the benchmarking exercises improved their ability to choose and report a product concept and write a requirement specification respectively. While this does not mean that the self and peer assessment processes used cannot be improved, it does demonstrate that each distinct process within each assessment task contributed significantly to improving a student's ability to achieve the prescribed learning outcomes. Perhaps more importantly these results demonstrate that the new features of SPARK^{PLUS} were able to successfully facilitate the introduction of two different types of self and peer assessment implementations, being the marking of individual submissions and the benchmarking exercise.

Students (47%) in the individual project concept exercise reported that *discussing the different concepts in the group* was the most effective part of the exercise in improving their understanding and ability (Figure 6). This was followed by *reading the reports themselves and assessing them against the criteria* (31%). Somewhat surprisingly only 17% of students reported that their tutors explanation of their *marking of an exemplar* was the major contributor to improving their understanding and ability of this particular exercise.

In the benchmarking exercise 37% of students reported that *discussing the specification marking within the group and then re-marking it collaboratively* was the part of the process that improved their understanding and ability the most (Figure 8). This was followed by *discussing and re-marking the report within the combined group* (25%) and *feedback guidance and explanation from the tutor* (23%). Only 14% of students reported that their understanding and ability was most improved by *reading and assessing the specification by themselves*.

While the fact that students found different parts of these tasks to be the most beneficial in improving their understanding and ability may be partly explained by differences between individual learning styles, the results do suggest that collaborative peer learning activities are generally the most beneficial. This is not surprising given that the group discussion activities have a social element, which tends to promote engagement. Conversely the fact that 31% (individual project concept marking) and 14% (benchmarking exercise) of students reported that individual work provided them with the most benefit supports our deliberate intention to design assessment tasks with different types of opportunities to learn that build on each other and accommodate the differences in students' abilities. In addition, the process of individually marking the work using SPARK^{PLUS} before the tutorial did mean that students had thought about the assessed work before the tutorial and so were able to make useful contributions to the discussion (for a more comprehensive discussion of these results see Willey and Gardner 2009a).

Furthermore the group discussion exercises were specifically designed to provide incentive for peer learning to occur. For example in the individual project concept exercise, the marking scheme is such that it is in each student's interest to choose the 'best' concept rather than to just argue for their own idea (academic moderation is achieved by marking the report identified as being the best by the peer learning group. This caps the marks received for the other submissions). While there were some complaints from students that it took too long to complete all the parts of the individual project concept and benchmarking exercises, generally speaking most students were positive in line with the survey free response comments below:

Peer review: *"Allows you to see what people think of your work and how you can improve"* [sic].

Benchmarking: *"Reviewing and marking a previous piece of work helped to understand the theory from the lectures. Knowing we need to write a Requirements Specification that is unambiguous is easy enough to know, but WHAT that actually looks like, and doing it is hard. Getting a picture of what NOT to do first, helps developing that knowledge"* [sic].

Figure 9 reports the results for the two self and peer assessment exercises used to determine a team member's contribution to the last two stages of the project. The 2008 results suggest that the use of self and peer assessment facilitated using SPARK^{PLUS} made a significant contribution to students' learning outcomes with 74% of respondents agreeing that it encouraged them to put more effort into their assigned work, 73% agreeing it improved their ability to make assessments, 75% agreeing it improved their ability to both give and receive feedback, and 69% agreeing that the feedback they received improved their contribution.

The majority of the survey free response comments were also positive, a sample of which are reported below:

"Peer assessment facilitated by SPARK improved my group work experience by facilitating and giving me peer feedback with regards to the contributions by the team. It gave all team members an opportunity to give fair and constructive feedback (mostly) to each other, thus improving the performance in projects throughout the semester, and most likely in later subjects also."

"Improved my group work experience as SPARK enables a fairer assessment, I was driven to participate and function with my team as a group. It gave me the opportunity to see my effort (by my SPA rating) and also to know what other team members thought about my performance from feedback received. I really enjoyed working in a group for this subject and I think SPARK had a big influence in that" [sic].

However some free response comments like those below also highlighted issues that should be considered in the design of future assessment tasks.

"*It's still difficult to give negative feedback, for fear of people being defensive and resentful."*

"Feedback couldn't be used to improve mistakes and consequently improve the assessment marks. I feel its a big waste when this is the case as the feedback isn't taken as serious as it should be as you cant use it to improve your marks. Even though it helps you to learn, as it doesn't show through in the assessment marks which is ultimately the students number 1 aim,…." [sic].

"I feel SPARK did not improve my group work experience, i was lucky enough to have a group of great guys, where we took it upon ourselves to work as a team, it may have affected my experience greatly if i had group members that did not do their share." [sic]

The latter comment indicates that there are still students who regard the major function of self and peer assessment is to deter free riders. This perception needs to change if students are to receive the potential benefits from the feedback these processes provide to assist their ongoing learning and professional development (Willey & Gardner 2009a, 2008b). We also find that students need more training and support to develop their team skills in particular dispute or conflict resolution and the ability to give constructive feedback.

As part of our response to these findings we are currently working on expanding SPARK^{PLUS} to facilitate students receiving a mark for the quality and usefulness of feedback they provide to group peers. It is our opinion that given the competitive nature of some students, unless we assess the quality of the feedback they provide, they may be reluctant to provide beneficial feedback to tasks that allow resubmission, for fear of helping a fellow student to exceed their own final grade. While we agree that this situation is not ideal with our desire being for students to focus on learning and not grades, for this to occur work is required to change the attitude of some students. As indicated in the previous free response comment improving their grade *"is ultimately the student's number 1 aim"*.

6.2 Effect on academics

The results clearly show the improvements to SPARK are effective in helping students to learn and in particular develop generic attributes including reflection, critical evaluation, ability to give feedback and an ability to respond to feedback to increase learning and improve performance. However, like teamwork, often only a student's peers are in a position to evaluate how well a student has developed these skills. Even when assessment tasks are designed to allow an academic to make informed judgements, in large classes this requires considerable time and effort often imposing an intolerable burden. The use of self and peer

assessment facilitated using SPARK^{PLUS} allows students to not only develop their critical evaluation / judgement by assessing each other's work and collaborate in peer learning but with careful design they are also provided with frequent opportunities to practise and develop a whole range of generic and discipline specific attributes. Furthermore after some initial training and support the academics and tutors in our trial reported they were able to use the metrics produced by SPARK^{PLUS}, in particular the category factors, to identify at a glance students/groups that were having problems achieving the specified learning outcomes. This facilitated academics moving into a coaching role, supporting students to address gaps in their development and learning.

The academic coordinators (2) and tutors (8) felt the knee formula for calculating a student's SPA factor provided them with the most useful option, combining the best features of the original and linear calculation methods. The coordinators reported that the knee formula helped promote teamwork and fair division of the assessment task between team members. For example Table 1 shows the SPA factors for a team of four students where initially Student A only contributes half of the work contributed by their three team peers. In the second instance Student A contributes twice as much work as their team peers. The knee formula does not reward students who might be tempted to do most of the work (Table 1 shows a student who did twice as much work as their peers would only get an SPA factor of 1.26) while providing incentive for those who are tempted to underperform (Table 1 shows a student who did half as much work as their peers would get an SPA factor of 0.57).

Table 1: SPA factors using different formulae for groups with over and under performing team members.

Assessor	Student A	Student B	Student C	Student D	SPA Original	SPA Knee	SPA Linear
Student A Contribution Half that of other Team Peer's							
Student A	1	1	1	1	0.76	0.57	0.57
Student B, C & D	2	2	2	2	1.07	1.07	1.14
Student A Contribution Twice that of other Team Peer's							
Student A	2	2	2	2	1.26	1.26	1.60
Student B, C & D	1	1	1	1	0.89	0.80	0.80

Students also reported that they preferred the knee formula as it provided a fairer distribution of marks and sent a stronger feedback message to underperforming students than with the factors calculated using the original formula. Some students had previously expressed concern that, using the original formula, underperforming students received an inflated mark that they were satisfied with, and hence were not motivated to improve their performance for the remaining parts of a project (Willey & Freeman, 2006b). For example Table 1 shows that with the original formula a student who only did half as much work as their team peers would get an SPA factor of 0.76 and hence receive 76% of the group mark. Using the knee formula this student's mark would be reduced to 57% of the group mark, a value that more closely reflects their true contribution.

The academics and tutors also reported that the new features of SPARK^{PLUS} were not only effective in facilitating the introduction of new self and peer assessment processes but produced efficiencies allowing self and peer assessment to be used multiple times a semester without an intolerable academic burden. These efficiencies include being able to clone subjects, tasks and criteria, easily produce tabulated results and radar diagrams for distribution to students within feedback sessions, and the ability to exclude or include student ratings when calculating factors assisted significantly in the detection of saboteurs (students who

intentionally submit dishonest ratings to distort the results).

However while these advantages are beneficial for a successful outcome, careful planning and assessment design is still required and support materials should be produced for both staff and students. For this reason we recommend a staged introduction of self and peer assessment into any subject, perhaps initially starting with a formative exercise to allow both academics and students to become familiar with the software. In particular, we caution against trying to implement all of the advanced procedures discussed in this article at once.

6.3. Original SPARK (2007) – SPARK^{PLUS} (2008) Comparison

The results reported in Figure 9 show that in 2007, when the original version of SPARK was used for students to rate their own and their team peers' contribution to the group project, a five point Likert format was used in the post-subject survey. In 2008 when SPARK^{PLUS} was used to facilitate the same assessment tasks the post subject survey used a four point Likert format. To allow comparison of these results we have divided the neutral responses to the 2007 survey equally between 'agree' and 'disagree' (ie: we have assumed that if the neutral option had not been provided, half of the respondents who chose it, would have chosen 'agree' while the other half would have chosen 'disagree'). The modified Strongly Agree/Agree (SA/A) results have been reproduced in Figure 10. While having to modify the data prevents the formation of a definitive opinion it does provide a useful comparison.

Take in Figure (10) Paper 1

Somewhat surprisingly the results (Figure 10) are almost identical. This is despite the authors being the only common tutors between both trials with the remaining six tutors being different. In fact five of the tutors in the second (2008) trial were very inexperienced having never previously tutored in any subject, however the tutor training, assessment design and feedback processes were better developed in 2008 compared to 2007. The results suggest that irrespective of whether SPARK or SPARK^{PLUS} is used as the tool to facilitate self and peer assessment of a student's contribution to a team task there is no statistical difference in the percentage of respondents who reported improvement (as described in the criteria) from these processes. While the authors report (via observation of tutorial classes, improved assessment task marks, student survey free response comments and feedback from tutors) that the learning and improvement achieved by students was considerably higher using SPARK^{PLUS} than when using SPARK (indeed the reported questions didn't ask how much students improved, just whether they did) the correlation between the results cannot be ignored. While more research is needed to further investigate this correlation, the results suggest that better tools do not necessarily lead to better outcomes, reinforcing that in teaching there is no substitute for good assessment design.

In previous research the authors have found that in exercises where self and peer assessment was used to determine an individual contribution to a team task a significant number of students perceive its use to be an instrument to facilitate fairness, focusing on its free-rider deterrent capacity, rather than providing opportunities for reflection and feedback to complete the learning cycle. This narrow focus reduced both student engagement and the quality of feedback provided (Willey & Gardner 2008b). The assessment of individual submissions and benchmark tasks were in part introduced to increase student engagement and change their perceptions in regard to self and assessment processes (Willey & Gardner 2009a). These two exercises produced individual marks and feedback on individual work while the assessments in task 3 and 4 were used to assess and provide feedback on individual contributions to a team project.

The results show that 25% to 30% of participants reported no improvement in their ability as a result of using self and peer assessment processes to determine their team contribution in tasks 3 and 4 (2nd column in each data set in Figure 9). In contrast approximately only 12% of respondents reported no improvement in their ability as a result of participating in the parts of the individual project concept assessment and benchmarking exercises that used SPARK^{PLUS} (first data sets in Figures 5 and 7). While these results need to be interpreted with caution since the practices were both new to students and only tested in a single trial they do suggest more students are engaged when self and peer assessment processes are focused on peer learning than when used to facilitate fairer group work assessment.

7. Future Directions

Our aim is to eventually combine the use of self and peer assessment with traditional academic assessment to assess, monitor, track and provide feedback on graduate attribute development throughout a degree program. Our intention is to provide students with two transcripts on graduation, one identifying their academic achievement and the second their professional attribute/skill development. Respondents (70% in the post subject survey) reported that this form of reporting would increase their motivation to address their weaknesses as they were identified.

Note: While we acknowledge that the assessments provided by students may not always accurately reflect the actual level of a student's attribute development, as they are the result of subjective assessments by peers, it does provide a useful guide and indeed mirrors the professional situation where an individual's contribution and performance is judged by their peers. In addition, if multiple peers are used (eg groups of size 4 to 8) in the evaluation process any bias from a single member of the evaluation group tends to be averaged out.

8. Conclusion

We found that using self and peer assessment in conjunction with collaborative peer learning activities increased the benefits to students and improved engagement. Furthermore we found that the new features available in SPARK^{PLUS} facilitated efficient implementation of additional self and peer assessment processes (assessment of individual work and benchmarking exercises) and improved learning outcomes. The trials demonstrated the tool assisted in improving student's engagement with and learning from peer learning exercises, the collection and distribution of feedback and helping students to identify their individual strengths and weaknesses. However we caution that while SPARK^{PLUS} is a powerful tool, its successful use requires thoughtful and reflective application combined with good assessment design.

Acknowledgements

The redevelopment of SPARK was a joint research project between the University of Technology, Sydney and the University of Sydney. The main developers were Dr Keith Willey UTS, A/Prof Mark Freeman USyd (also chief architect and developer of the original SPARK) and Mr Darrall Thompson UTS. We would also like to acknowledge the contribution of Mr Mike Howard who has continued to work with the authors in developing SPARK^{PLUS}.

9. References

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

Calibrated Peer Review (2009), http://cpr.molsci.ucla.edu/ , Last visited May 27th 2009.

Engineers Australia (2004), *Australia Engineering Competency Standards – Stage 1 Competency Standards for Professional Engineers*, downloaded from http://www.engineersaustralia.org.au on 30 June 2007.

Falchikov, N., and 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. and 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*, vol. 33, pp. 551-69.

Goldfinch, J. (1994). Further developments in peer assessment of group projects., *Assessment & Evaluation in Higher Education* Vol. 19, pp. 29: Carfax Publishing Company.

Goldfinch, J., and Raeside, R. (1990). Development of a Peer Assessment Technique for Obtaining Individual Marks on a Group Project. *Assessment and Evaluation in Higher Education*, 15(3), 210- 231.

Hazelton, P., Malone, M., & Gardner, A (2008) A multicultural, multidisciplinary shortcourse to introduce recently graduated engineers to the global nature of professional practice. *Proceedings of the 36th Annual Conference of the European Society for Engineering Education*, 2-5th July, 2008, Aalborg, Denmark.

Lejk M, Wyvill M, and S. Farrow (1996). A survey of methods of deriving individual grades from group assessments, *Assessment & Evaluation in Higher Education*, vol. 21, pp. 267–280, 1996.

Johnston L and Miles L, (2004). Assessing contributions to group assignments, *Assessment and Evaluation in Higher Education*, vol. 29, pp. 751, 2004.

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

Scott G and Yates K. W, "Using successful graduates to improve the quality of undergraduate engineering programmes," *European Journal of Engineering Education*, vol. 27, pp. 363, 2002.

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

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. *Proceedings of the 17th Annual Conference of the Australasian Association for Engineering Education*, 10 -13th December 2006, Auckland, New Zealand.

Willey K, & Freeman M. (2006b), "Improving teamwork and engagement: the case for self and peer assessment", *Australasian Journal of Engineering Education.* Online publication 2006-02 http://www.aaee.com.au/journal/2006/willey0106.pdf

Willey, K and Gardner, A. (2008a) Using Self Assessment to Integrate Graduate Attribute Development with Discipline Content Delivery. *Proceedings of the 36th Annual Conference of the European Association of Engineering Education (SEFI)* 2-5 July, Aalborg, Denmark.

Willey, K. & Gardner, A. (2008b), 'Using self and peer assessment for professional and team skill development: do well functioning teams experience all the benefits?', Adelaide, November 2008 in *ATN Assessment Conference 2008: Engaging Students in Assessment*, University of South Australia, Adelaide,

Willey, K. & Gardner, A. (2009a), "Investigating the capacity of Self and Peer Assessment to Engage Students and Increase their Desire to Learn" *SEFI 37th Annual Conference*, Rotterdam, The Netherlands, July 2009

Willey, K. & Gardner, A. (2009b),"Changing Student's Perceptions of Self and Peer Assessment", *Research in Engineering Education Symposium* July 20-23, 2009, Queensland, Australia

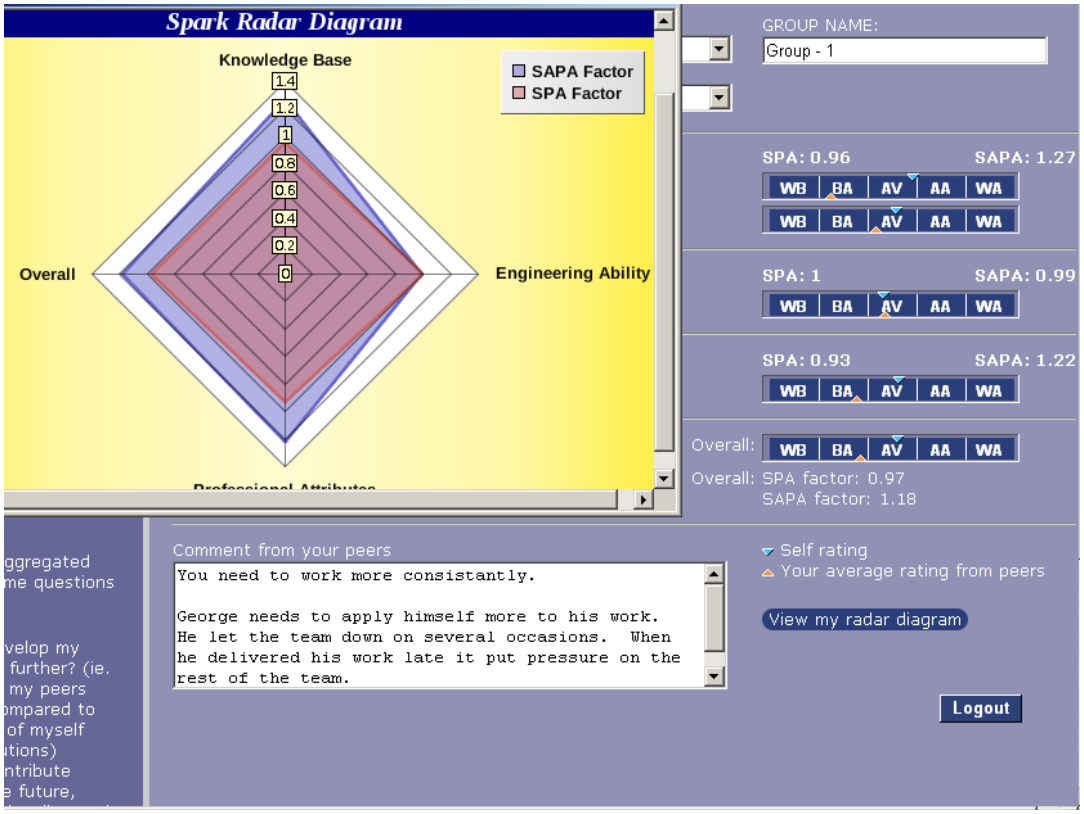


Figure 1: Screen shot showing overall and category feedback (SAPA) and performance (SPA) factors, radar diagram and feedback comments.

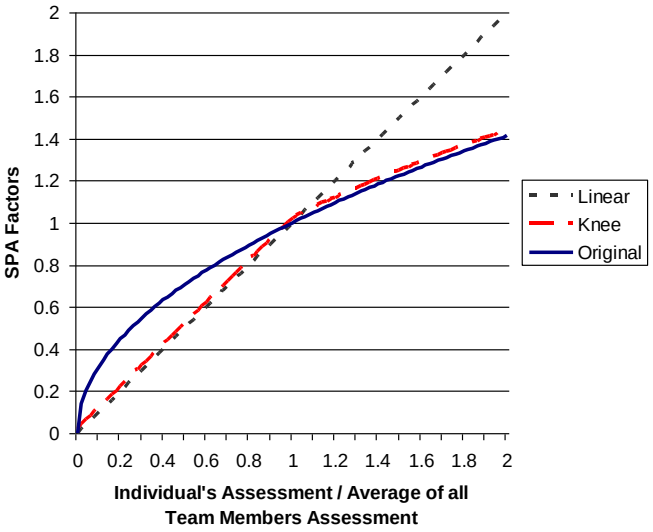


Figure 2: The relationship between the two new and original method of calculating the SPA factor. Note the Knee plot has been slightly offset to increase readability.

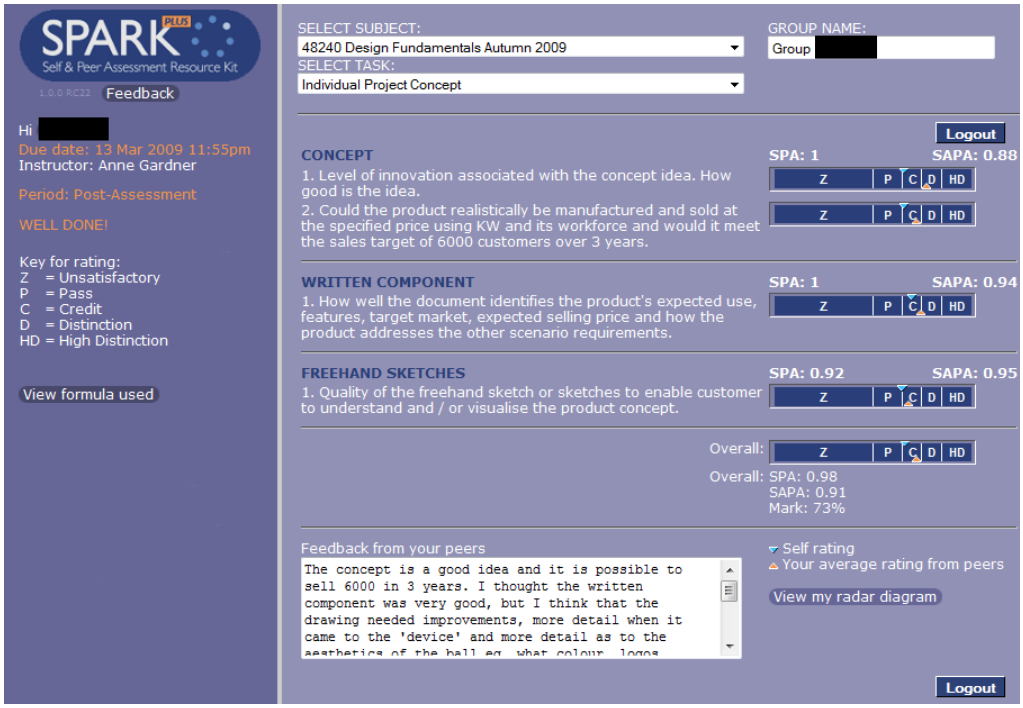
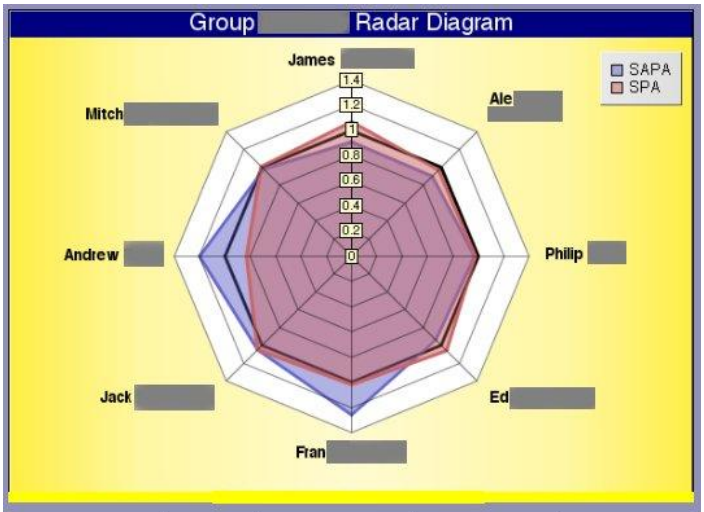


FIGURE 3: A student's SPARK^{PLUS} results screen for a task where each student had to self assess their own submission and peer assess the submissions of their team peers.



Student	Concept		Written Component		Freehand Sketches		Overall		
	SPA	SAPA	SPA	SAPA	SPA	SAPA	SPA	SAPA	Mark
James [redacted]	1.06	0.91	1.08	0.87	1.08	0.96	1.07	0.91	84%
Ale [redacted]	0.99	0.9	0.99	0.85	0.92	0.97	0.97	0.9	76%
Philip [redacted]	0.98	0.95	0.94	0.95	1.02	1.04	0.98	0.97	77%
Ed [redacted]	1.06	1.01	1.03	0.83	1.09	0.92	1.06	0.94	83%
Fran [redacted]	0.99	1.3	1.05	1.23	1.03	1.23	1.02	1.26	80%
Jack [redacted]	1.05	1	1.07	1.09	1.06	1.08	1.05	1.05	83%
Andrew [redacted]	0.84	1.21	0.85	1.14	0.78	1.24	0.83	1.2	65%
Mitch [redacted]	1.02	1	1	0.92	1.02	1.01	1.01	0.98	80%

FIGURE 4: SPARK^{PLUS} Group Radar Diagram and table.

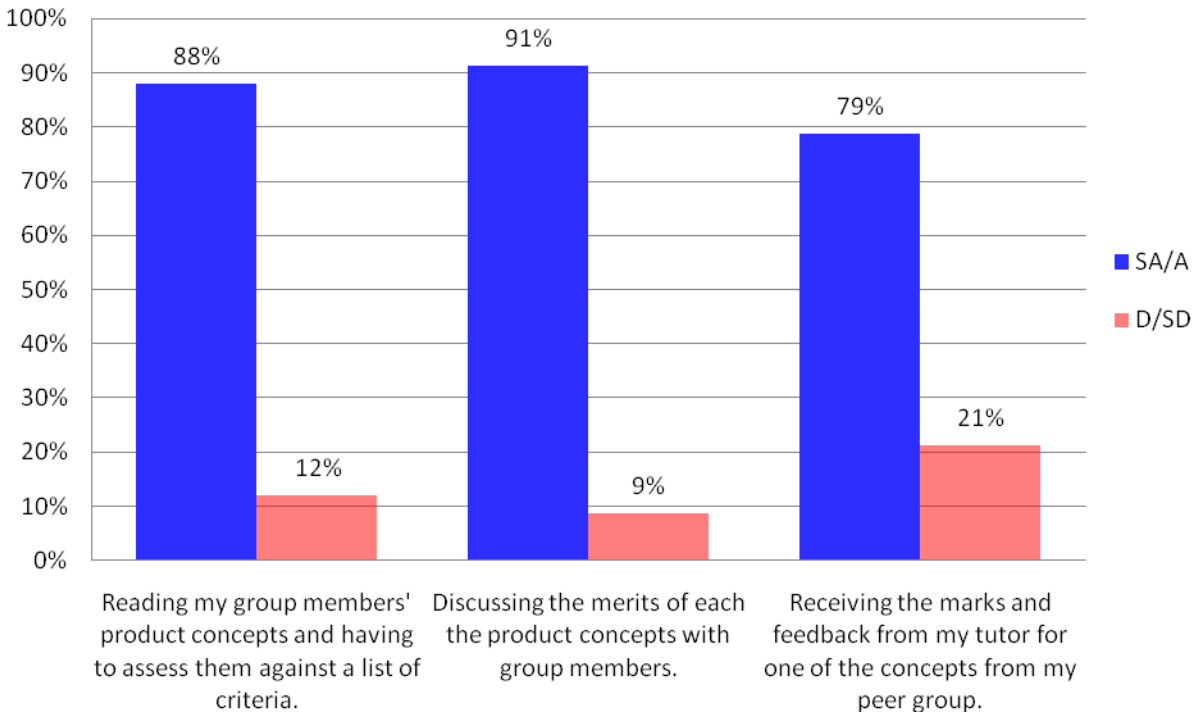


FIGURE 5: Results from student survey of Self and Peer Assessment Marking of Individual Project Concepts in response to the question "*My ability to choose a product concept and write a concept document to meet a list of requirements increased as a result of:"*

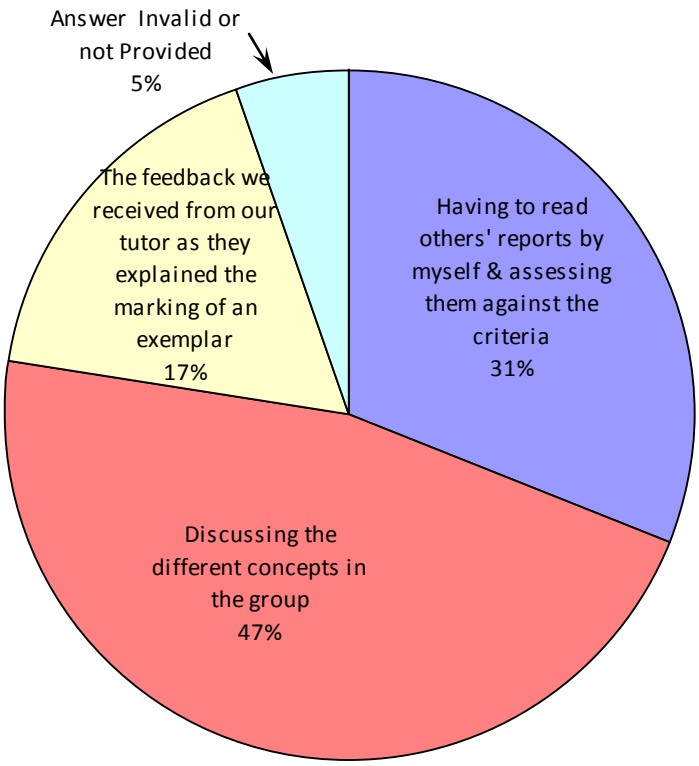


FIGURE 6: Results from student survey of Self and Peer Assessment Marking of Individual Project Concepts in response to the question: "*Which part of the whole process improved your understanding / ability the most?"*

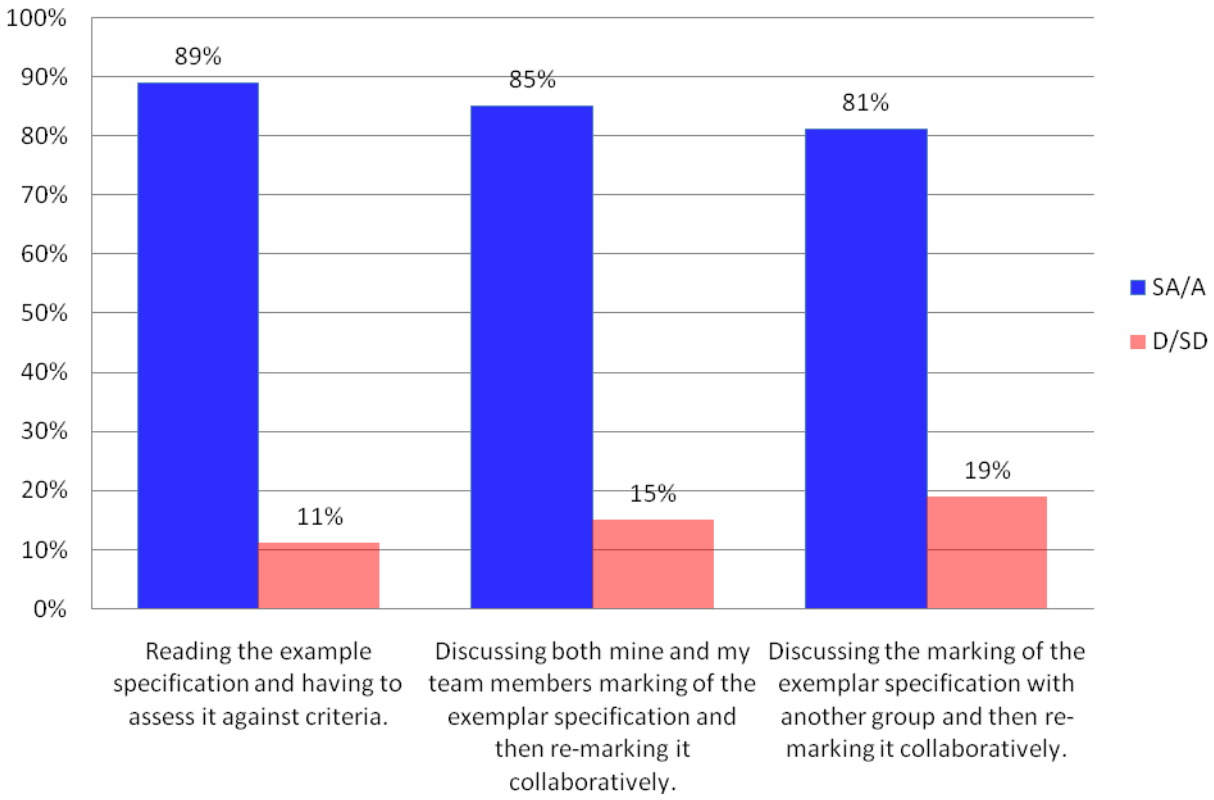


FIGURE 7: Results from student survey of Self and Peer Assessment Benchmarking Exercise in response to the question *"My ability to write a quality requirement specification has increased as a result of:"*

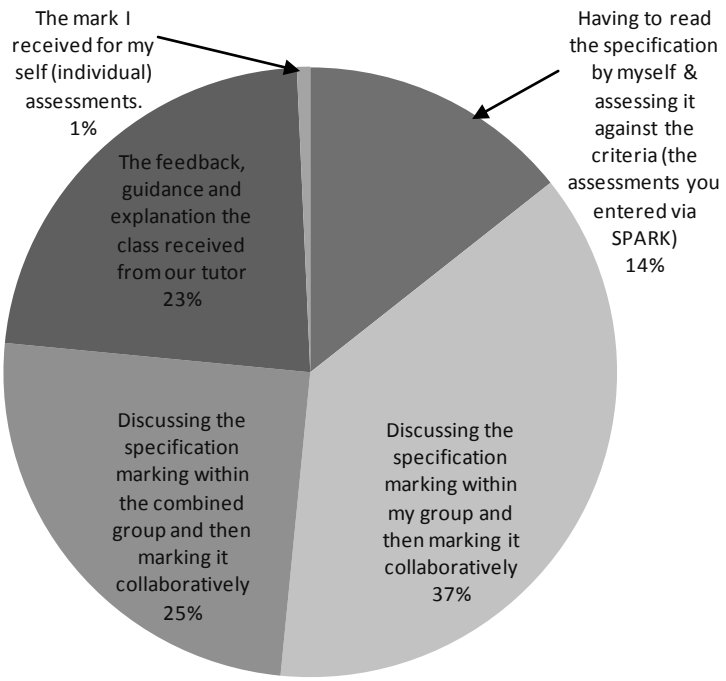


FIGURE 8:Results from student survey of Self and Peer Assessment Benchmarking Exercise in response to the question *"Which part of the whole process do you feel improved your understanding / ability the most?"*

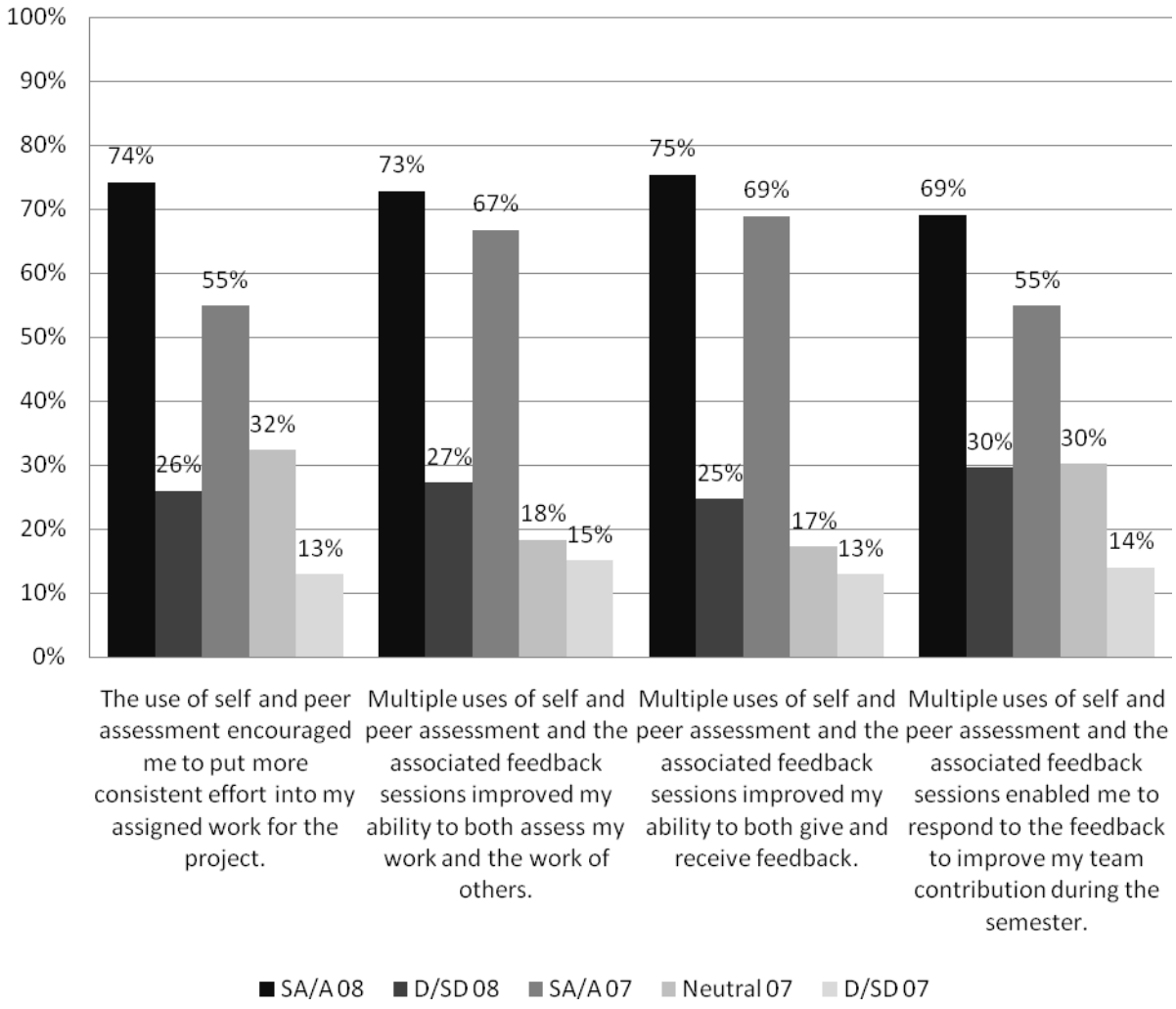


FIGURE 9: Results from student post-subject survey

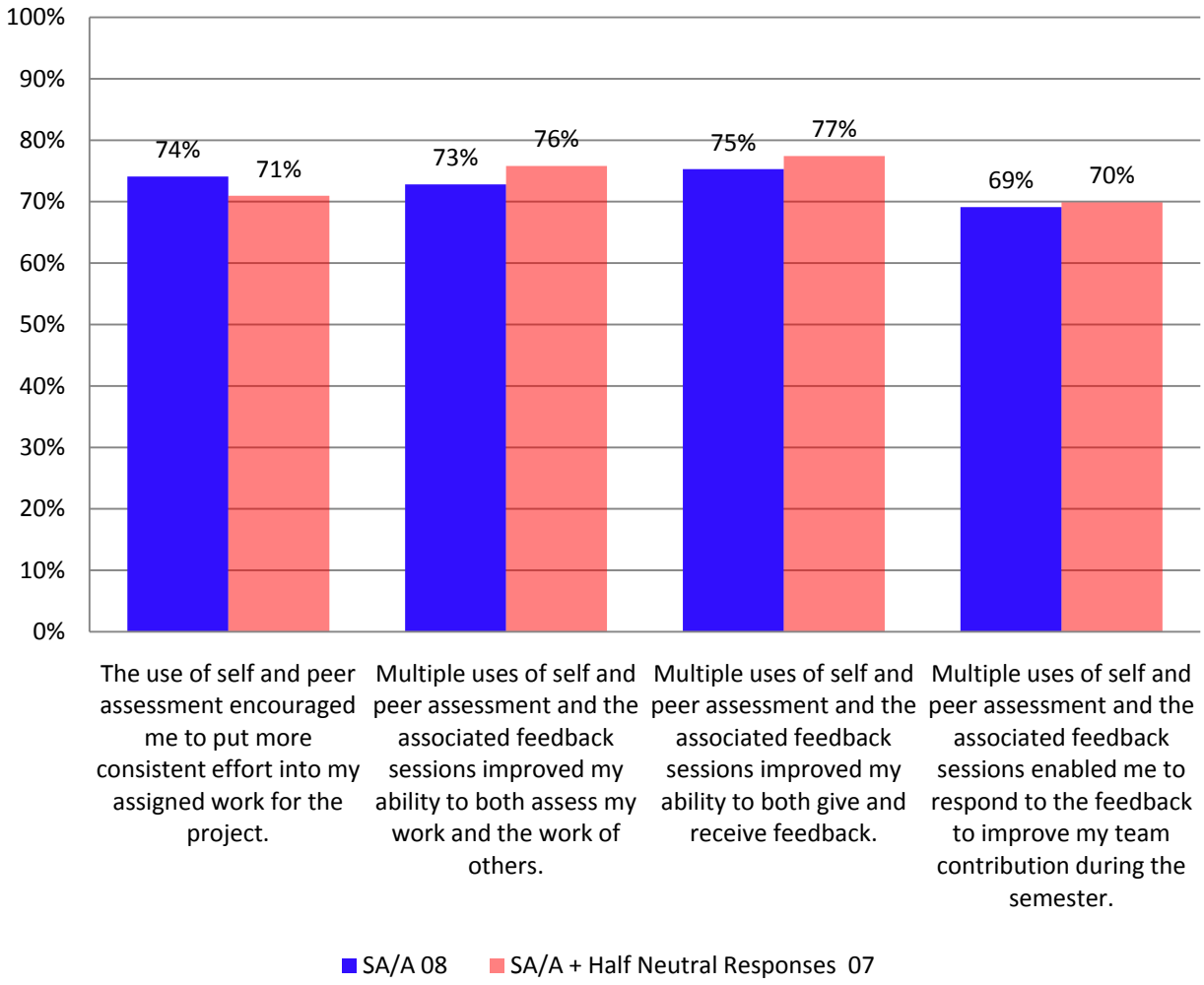


FIGURE 10: Modified results from student post-subject survey