Using Benchmarking to improve students' learning and make assessment more student centred.

Keith Willey
University of Technology, Sydney, Australia
Keith.Willey@uts.edu.au

Anne Gardner
University of Technology, Sydney, Australia
Anne.Gardner@uts.edu.au

Abstract: *If universities are to prepare students for professional practise then they should also prepare them for the type of learning that occurs in the workplace. Graduates' workplace learning often differs considerably from their university experience. Firstly there is typically no lecturer or tutor to instruct them, and workplace learning is often collaborative. Hence students' preparation for entering this workplace environment should include opportunities to practise collaborative learning with their peers. While designing collaborative learning tasks that involve students having to make and reflect on their judgements are extremely beneficial, these participatory exercises can result in an intolerable administrative burden especially for large classes. The authors have recently developed a specific benchmarking tool to assist academics to produce regular student centred assessments to improve students' judgement and learning with a manageable academic effort.*

In this paper we reflectively deconstruct a purposely designed collaborative learning activity, investigating the effect of each of its components on student learning. Furthermore we explore how effectively the educational technology used reduced both the marking and administrative burden of running these exercises particularly in large classes.

Introduction

In this paper we scrutinise a collaborative learning activity specifically designed to improve students' judgement. The motivation for the development of this activity was a desire to improve students' understanding of requirements in the context of engineering design and to provide an opportunity to practise their judgement based on this understanding. In the activity (described in this paper), prior to attending the tutorial students were required to make their own assessment of a requirements report. Hence, students had to test their individual knowledge, understanding and judgement before each group discussed and reflected on their individual assessments. The group then collectively reassessed the report. This peer feedback and collaborative learning was followed by feedback from the instructing academic. Students then had two weeks to submit a requirements report for their product. In this paper we reflectively deconstruct this activity, investigating the effect of each component on student learning. Furthermore we explore how successfully the educational technology used reduced both the marking and administrative burden of running these types of exercises particularly in large classes.

Background

Chung et al (2008) and Hargeaves (1997) discuss a gap between skills typically developed in engineering education and a range of skills required for professional practice such as communication, critical thinking, leadership, teamwork skills and life long learning capabilities. This requires not only considering **what** is taught but **how** it is taught (Hargreaves, 1997). Workplace learning and certainly practice is often collaborative. It follows that students' preparation for entering this environment should include opportunities to practise collaborative learning with their peers. Collaborative learning also provides opportunities to develop interpersonal and critical evaluation skills in addition to professional judgement. The skills to critically evaluate and clearly articulate your point of view are necessary to successfully participate in collaborative professional practice. Despite this students often receive only cursory training combined with infrequent opportunities to develop such skills.

Collaborative learning is also attractive from the perspective of the constructivist model of learning (Jawitz and Case, 2009). The constructivist view is that learning takes place when students construct their knowledge through individual engagement and social interactions with others (Wu, Beiber and Hiltz, 2008, Purzer, 2009). It is the students doing the learning rather than the teacher doing the teaching that determines whether learning takes place, and so this is a student-centred philosophy. Hagstrom (2006) argues that "…contexts for new knowledge construction include a blending of people … that gives rise to differences in interpretation and provides the occasion for the construction of new knowledge….If educators simply tell students what they need to know, they encourage reliance on memorization of facts. For students to make cognitive changes, the learning experience must begin with each student becoming aware of his or her own present understanding" (Hagstrom, 2006, p28).

Collaborative learning tasks in addition to supporting the constructivist theory, have been found to promote a community of learning in the classroom that has benefits in terms of "…retention, achievement and cognitive development" (Summers et al 2005, p.5). Furthermore, student participation in assessment activities provides an opportunity to develop a metacognitive awareness of what is meant by good work (Frederikson and Collins, 1989).

While designing collaborative learning tasks that involve students having to make and reflect on their assessments and judgements are extremely beneficial, Hargreaves (1997) reports that these participatory exercises can result in an "enormous increase in time required to mark…especially for large classes." (Hargreaves, 1997, p.408). As well as examining the learning benefits of the benchmarking activity we also test the use of educational technology to reduce both the marking and administrative burden of running these exercises particularly in large classes.

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 allow self and peer assessment of individual work and the facilitation of benchmarking exercises to develop students' judgement (Willey & Gardner, 2008a; Willey & Gardner, 2008b).

As this paper only discusses the use of the benchmarking mode we will only describe this aspect of the tool's operation.

The benchmarking mode enables an instructor to create an assessment of an exemplar piece of work, activity or task against which students can compare their own understanding and/or judgement. Instructors can choose from a number of predefined rating scales or create their own, however typically for benchmarking a Standard Assessment scale is used (for example Unsatisfactory (Z), Pass (P), Credit (C), Distinction (D) and High Distinction (HD) (Figure 1).

Students are required to logon and enter their assessments by moving the sliders (orange bars) against a number of criteria (Figure 1). While we have found it useful to discuss the chosen criteria in class and post explanatory details online, to prevent screen clutter when the criteria are wordy we only record the heading of each criterion (eg Test Plan) in SPARK^PLUS (Figure 1).

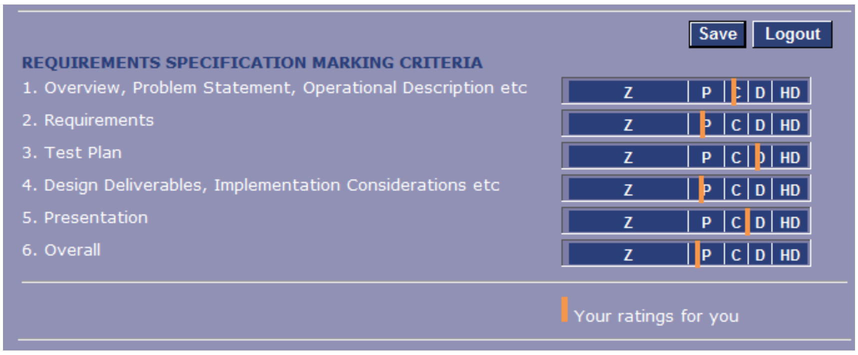

Figure 1: Students enter their assessments by moving the sliders to their chosen rating against each criterion.

In addition to entering their own assessments, instructors enter a written report explaining their marking against each criterion. The calculation of a student's score is generated using a weighted mean squared error of the differences between the student's assessments and the instructor's assessments. We recommend that it is best to moderate the results by adjusting the scores to fit to specified boundaries.

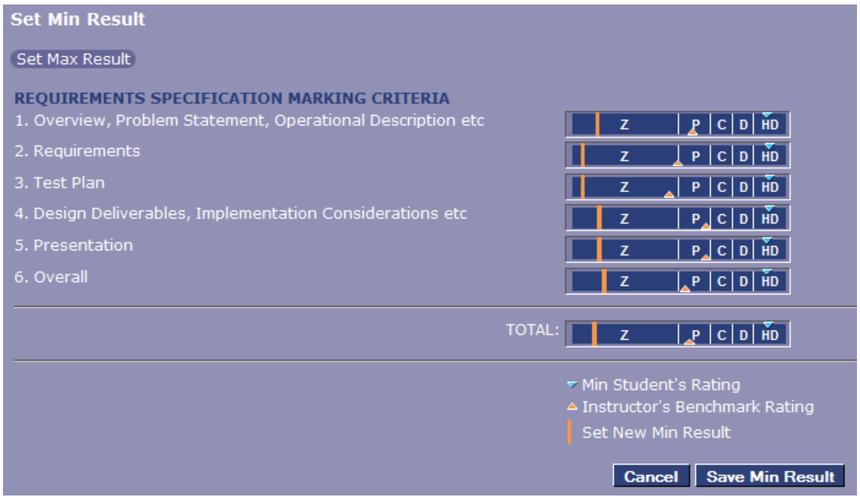

Figure 2: Mark moderation - Instructor screen showing the setting of the minimum student mark for the exercise.

SPARK[PLUS] facilitates moderation by providing instructors with two unidentified student submissions. One of these submissions is the student assessment that differs the most from the instructor's, and the other is the submission that is closest to the instructor's assessment of the report. These submissions are then used to set the minimum and the maximum student mark respectively. Figure 2 shows the screen used to set the minimum student mark for the exercise. The upper blue triangle shows the rating submitted by the student for each criterion. The lower orange triangle shows the rating submitted by the instructor. The instructor uses the orange bars to rate the student judgement against each criterion. The 'TOTAL' bar allows instructors to fine tune their marking encouraging a holistic approach (the individual criterion bars move proportionally tracking the movement of the 'TOTAL' slider). The maximum result is set in a similar way by grading the student submission that was closest to the instructor's assessment. The results for the remaining students are calculated according to the chosen formula (a number of different formulas are provided to accommodate different assessment objectives) by adjusting the raw mean squared error scores to fit within the specified moderation limits.

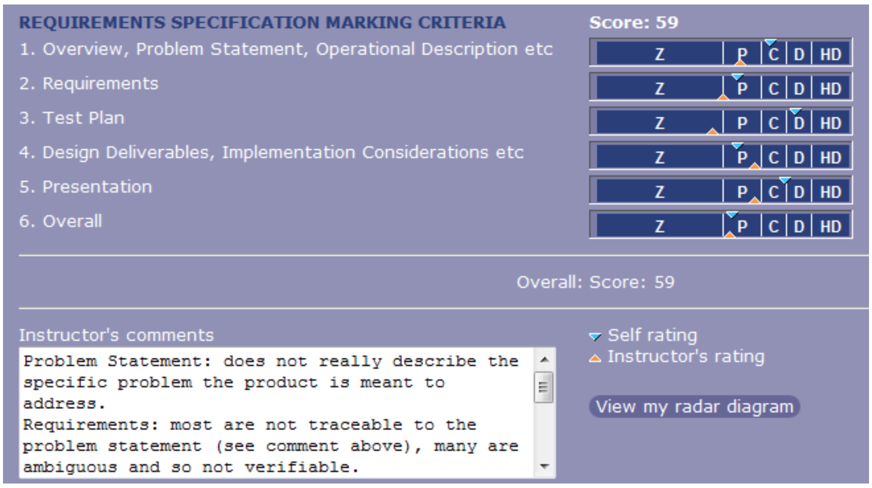

Figure 3: Individual student results screen reports the student's overall score. In addition, the upper blue triangle shows the rating submitted by the student and the lower orange triangle shows the rating submitted by the instructor for each criterion. The feedback box explains the instructor ratings.

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. In the results screen the sliders display two triangles. The upper blue triangle shows the rating submitted by the student and the lower orange triangle shows the rating submitted by the instructor. In addition a feedback box explains the instructor ratings for each criterion (Figure 3).

Method

The benchmarking 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 with tutorial classes being limited to a maximum of 32 students and project groups 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, and
- Continue the development of students' professional skills including teamwork, critical evaluation, feedback and communication commenced in earlier subjects.

The benchmarking activity consisted of a series of distinct processes:

1. Students were provided with a Sample Requirement Specification. After discussing the marking criteria each student individually assessed the report using the benchmarking function of SPARK^{PLUS}.
2. In the following tutorial each project group of four students discussed their individual marking of the report and re-assessed it collectively against the criteria.
3. Two project groups then combined and discussed each group's marking of the report, reflecting on any differences between the assessments by the two groups and collectively re-marking it.
4. Tutors then discussed how the academic had marked the report.
5. After the tutorial students logged on to SPARK^{PLUS} to compare their individual marking to the academic's assessment of the report for each separate criterion and to read the academic's comments. In addition, they can view their mark calculated based on how close the student's individual assessment was to the academic's assessment of the sample report.

The order of the processes in the benchmarking activity meant that students were required to individually think and engage with the assessment criteria for the report before they came to the tutorial and discussed the report with their peers. So students knew the details of what they were coming to class to talk about and discussions could relatively quickly focus on areas where there was a difference of opinion.

The motivation to actively participate in the benchmarking activity was that students in their project groups were required to write a Requirement Specification report for their group's product. Furthermore, students are encouraged to explore the different opportunities to learn. They may choose to teach others and in the process improve their own understanding or alternatively being taught by their peers to address gaps in their learning. It is our experience that in collaborative exercises most students tend to adopt a combination of these learning methods, but we strongly encourage those that feel they have nothing to learn from their peers to take the opportunity to teach.

On completion of the benchmarking activity students were asked to complete a questionnaire to assess how effective the process was in improving their learning. The 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 201 (eligible cohort 256) students responding. Of these, student one student only completed the background questions and hence this submission was excluded from the results making the participating cohort 200.

Results

The survey results relevant to this article are shown in Figures 4 – 6 and Table 1. 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%.

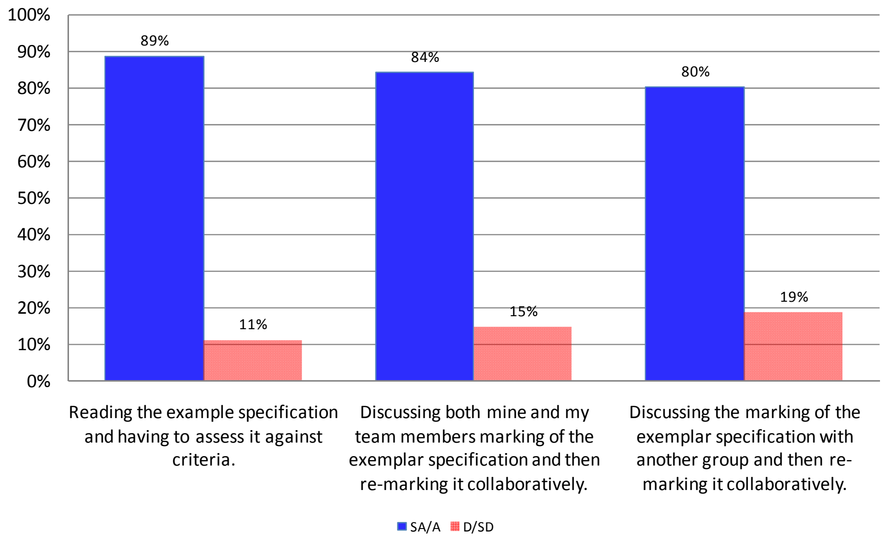

Figure 4: Student survey results for Self and Peer Assessment Benchmarking Exercise in response to the question "My ability to write a quality requirement specification has increased as a result of:"

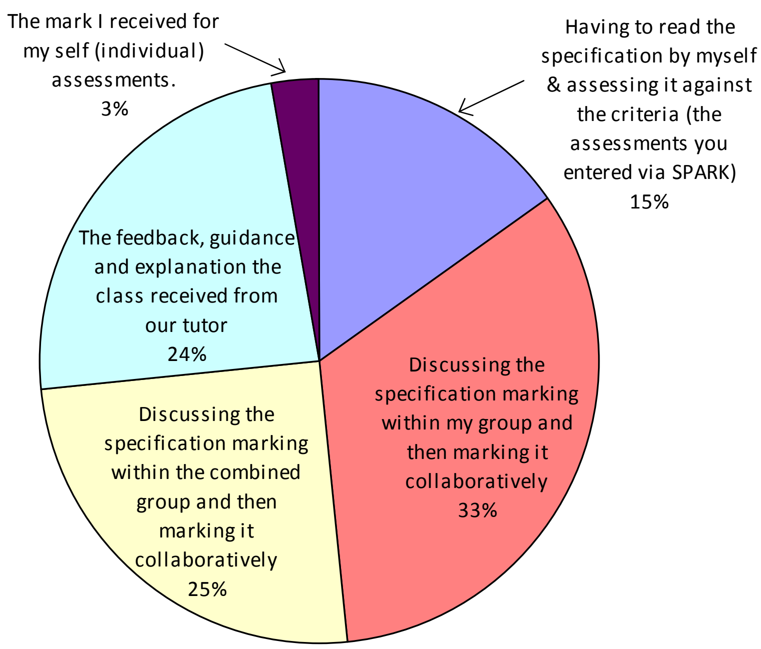

Figure 5: Results from student survey of Self and Peer Assessment Marking of Individual Project Concepts in response to the question: "*Which part of the whole process improved your understanding / ability the most?*"

Table 1: Student survey results in response to the question "If you consider the amount that your understanding / ability to write a quality requirement specification increased as a result of this exercise as 100%, how much (expressed as a percentage) did each of the following contribute to improving your understanding / ability?"

Process	Average % improvement in understanding / ability	Standard deviation
Having to read the specification by myself & assessing it against the criteria (the assessments you entered via SPARK)	23	22
Discussing the specification marking within my group and then marking it collaboratively	28	20
Discussing the specification marking within the combined group and then marking it collaboratively	24	19
The feedback, guidance and explanation the class received from our tutor	25	21

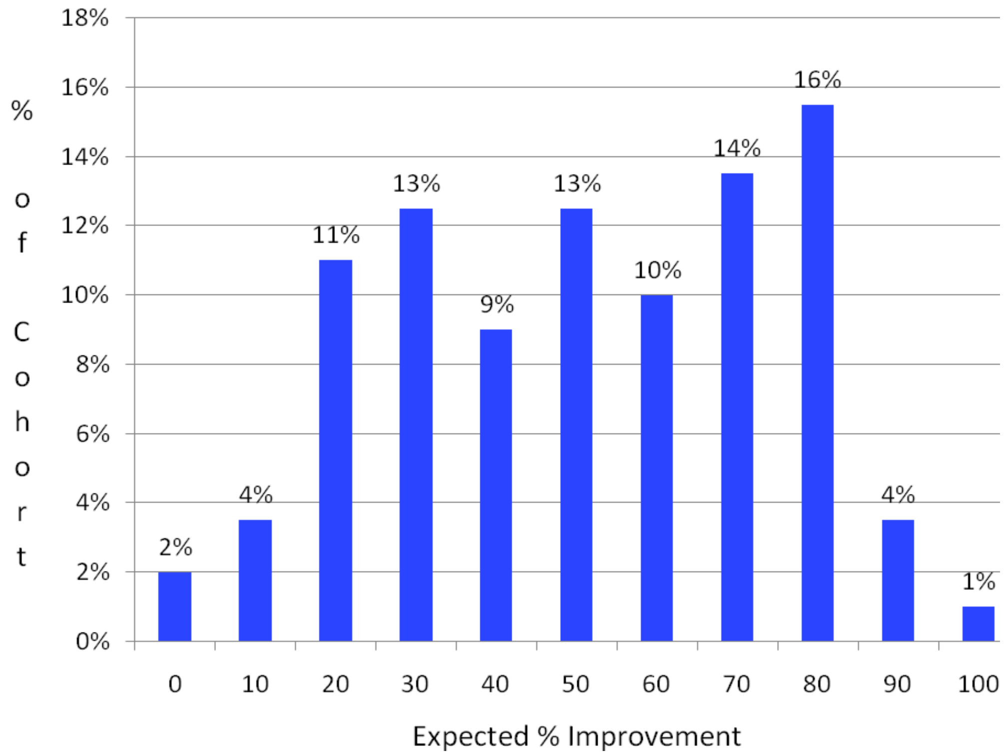

Figure 6: Student survey results in response to the question "Overall how much do you think this exercise will improve the mark you will get for your requirement specification, compared to the mark you think you may have received if you hadn't participated in this exercise?"

Discussion

The results (Figure 4) show that the majority of students (ranging from 80% to 89%) felt that all aspects of the benchmarking exercise improved their ability to meet the prescribed learning outcomes.

Furthermore 33% of students reported that discussing the report 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 5). This was followed by discussing and re-marking the report within the combined group (25%) and feedback, guidance and explanation from the tutor (24%). Only 15 % 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 this task to be the most beneficial in improving their understanding and ability may be partly explained by differences between individual learning styles, the results do demonstrate the effectiveness of collaborative activities to promote students learning with students reporting that the collaborative peer learning tasks (2 and 3) contributing twice as much as the self assessment component to improving their ability and learning. Conversely the fact that 15% of students reported that the individual component 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.

The group discussion activity is also different to other parts of the process in that it has a social element, which tends to promote engagement. However, 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. While receiving feedback from the tutor has a personal communication element, this interaction is much different to the interaction that students have with each other, and probably not as much fun. It is not unreasonable to suggest if the collaborative discussion part of the process is where students are most engaged, then this may also contribute to their perception that this is the part of the process where they learnt the most.

Another factor that needs to be considered in interpreting these results is that it tends to be the middle activity in the exercises that most students feel produced the most benefit. This is not surprising as students develop some understanding in the individual segment of the task, they build on this learning and explore their understanding in the collaborative group exercise while the last component involving interaction with the tutor happens after most of the learning has already occurred and serves mainly to clarify any outstanding questions and issues.

However the benefits of providing a mixture of learning opportunities should not be underestimated. The results shown in Table 1 report on average that the percentage of students' overall improvement in their understanding / ability attributed to each part of the benchmarking exercise were remarkably similar. Students reported that the second part of the exercise where they *discussed the specification marking with their group and re-marked it collaboratively* contributed on average 28% to their improved understanding. This was only slightly higher than the average contribution made by *having to read the specification themselves & assessing it against the criteria* in SPARK^{PLUS} (23%), *discussing the specification marking within the combined group and then marking it collaboratively* (24%) and *the feedback, guidance and explanation the class received from our tutor* (25%). Furthermore, the similar standard deviations for each data set suggest a similar distribution in each case.

Thus a comparison of the results reported in Figures 4 and Table 1 shows that while more students considered 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 on average all parts of the exercise contributed almost the same to improving the understanding of the collective cohort. An additional benefit of integrating the benchmarking with a number of interactive collaborative exercises is that it makes the whole process more student centred, providing students with an opportunity to take responsibility for their own learning and providing each other with feedback.

Figure 6 shows that 16% of respondents reported that as a result of participating in the benchmarking exercise they expected to improve their mark for their first assignment by 80%. We were quite surprised by this figure and suspect that this exercise might have been the first time most of the students had engaged with the subject material and hence as a result of the exercise their understanding improved significantly. Probably more representatively Figure 6 shows that the majority of students felt that the benchmarking exercise would improve their mark by between 20% and 80% with the average expected improvement of all respondents being 51%. The fact that 88% of the respondents expected their improved understanding would result in the least a 20% improvement in their grade for their group's requirement specification report is quite remarkable and clearly demonstrates the potential of both the SPARK^{PLUS} tool and in particular benchmarking exercises to improve student learning outcomes. Furthermore, the tool enables these exercises to be run in large classes with relatively little overhead for the coordinating academic.

While there were some complaints from students that it took too long to complete the entire exercise, generally speaking most students were positive in line with the following survey free response comment: "*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].*

The results clearly show that students believe that the exercise improved their ability or understanding of the required learning outcomes. While this single trial does not allow us to determine directly how successful the process was in improving a student's judgement (as this would require either the use of a control group or a second evaluation exercise) the fact that students were able to practise, test and receive feedback in regard to their understanding compared to that of the instructor should improve their judgement. Compared to previous semesters, running the benchmarking exercise allowed students to practice and test their judgement twice, once during the exercise and once when they submitted their actual report. Furthermore, SPARK^{PLUS} allowed the instructors to conduct the exercise with very little academic effort. The assessment criteria used were the same as those already prepared for the final report. The set up overheads were minor and the program automates the distribution of feedback and results. In addition, the moderation method requires the instructor to only mark two individual submissions.

Conclusion

The results show that benchmarking was successful in assisting students to achieve the desired learning outcomes. The majority of respondents, greater than 80%, reported that its use improved their ability to meet the required learning outcomes. Furthermore integrating the benchmarking with a number of interactive collaborative exercises made the whole process more student centred, providing students with an opportunity to take responsibility for their own learning. The value of each component was demonstrated by the fact that on average students reporting that the overall improvement in their understanding / ability attributed to each part of the exercise was nearly the same (approximately 25%).

Furthermore the instructors involved with the subject reported that SPARK^{PLUS} efficiently facilitated the running of the exercise, with only minor academic effort being required to collect, assess, collate and report (feedback and results) to a cohort of 256 students.

References

Chung W., Stump G., Hilpert J.,Husman J., Kim W., Lee J. E. (2008) Addressing Engineering Educators' Concerns: Collaborative Learning and Achievement *Proceedings of the 38th ASEE/IEEE Frontiers in Education Conference* 22-25 October, Saratoga Springs, NY.

Frederickson, J.R. and Collins, A. (1989) A Systems Approach to Educational Testing. *Educational Researcher*, Vol. 18, pp. 27-32.

Hagstrom F. (2006) Formative Learning and Assessment *Communication Disorders Quarterly*. 28, 1. p. 24

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

Jawitz, J. and Case, J. (2009) Communicating your findings in engineering education: the value of making your theoretical perspective explicit ,*European Journal of Engineering Education*,34:2,149 — 154

Purzer S. (2009), Learning Engineering in Teams: Perspectives from Two Different Learning Theories *Proceedings of the Research in Engineering Education Symposium* Palm Cove, QLD

Summers J. Beretvas S. N., Svinicki M. D. Gorin J. S. (2005) Evaluating Collaborative Learning and Community *The Journal of Experimental Education*; 73 (3). p. 165

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) Improvements in the self and peer assessment tool SPARK: Do they improve learning outcomes? *Proceedings of the ATN Assessment Conference – Engaging Students in Assessment, November, 2008. South Australia.*

Wu D., Bieber M. Hiltz S. R., (2008) Engaging Students with Constructivist Participatory Examinations in Asynchronous Learning Networks *Journal of Information Systems Education*, Vol. 19(3).

Acknowledgements

The initial 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}.