British Journal of Educational Technology *Vol 33 No 5 2002* 551–569

SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects

Mark Freeman and Jo McKenzie

Associate Professor Mark Freeman is in the Faculty of Business at the University of Technology, Sydney. He joined the university in 1984 after a commercial career in accounting and financial analysis. While teaching financial management and regulation, his research extends beyond finance and insider trading to learning and assessment. Mark has lead learning and assessment projects and technology innovations and solutions. He has been active in disseminating student-centred learning programs and technologies within higher education and professional practice. Mark has received multiple awards for excellence in teaching from the university, the finance and educational technology professions, and the Australian government. Jo McKenzie is a Senior Lecturer in the Institute for Interactive Media and Learning at the University of Technology, Sydney. She works with academics and Faculties on course design, assessment, flexible learning and educational innovation to improve learning for students and academics. Her research interests include change in university teaching, academic development and variation and learning. Address for correspondence: Jo McKenzie, Institute for Interactive Media and Learning, University of Technology, Sydney. Email: Jo.McKenzie@uts.edu.au

Abstract

Students often enjoy learning in teams and developing teamwork skills, but criticise team assessment as unfair if there is equal reward for unequal contributions. This paper describes the design, implementation and evaluation in four subjects of the Self and Peer Assessment Resource Kit (SPARK), a web-based template which aims to improve learning from team assessment tasks and make the assessment fairer for students. Students benefit because the web-based template improves confidentiality and the potential for accurate assessment of relative contributions. Academics benefit through the potential for improving student learning from teamwork tasks, and saving time by automating the process of calculating self and peer adjustments of assessment grades, especially attractive for large enrolments. Benefits accrue to the institution and wider academic community because the template suits a range of group assessment situations. Based on experiences gained over five years of developing, evaluating and implementing SPARK, this paper aims to illustrate the potential benefits of the template to potential users and more critically, to use what was learned from implementing the template across a range of subjects to alert others to key issues for evaluating and disseminating educational technology innovations.

Published by Blackwell Publishers Ltd, 108 Cowley Road, Oxford, OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

Introduction

Many courses aim to develop students' ability to work as part of a team and include team assessment tasks such as presentations, projects, case studies, reports, debates and so on. Students often value the experience of learning in teams and developing teamwork skills, but criticise team assessment as unfair if team members are equally rewarded for unequal contributions. This paper illustrates the potential benefits of SPARK (Self and Peer Assessment Resource Kit), a web-based template that aims to improve the fairness of team assessment and enhance students' learning from team tasks. SPARK enables students to rate confidentially their own and their peers' contributions to team tasks and team maintenance, so that shared team marks can be moderated to acknowledge individuals' contributions. It also automates the logistics of data collection and calculation, making it less time consuming for academics to use self and peer assessment of team contributions, especially in large classes. One aim of this paper is to share with other academics the lessons learned from implementing and evaluating SPARK, to assist them to use similar approaches in their subjects. A second aim is to share with academic developers some further lessons that relate to the development and dissemination of generic templates, to increase the likelihood that others will use them effectively.

The paper is structured as follows. The next sections discuss previous research on self and peer assessment, briefly describe the context within which SPARK was developed then provide an overview of SPARK and its use in a typical context. We then present and compare four case studies of different subjects in which SPARK was implemented and evaluated. The following section synthesises the lessons that were learned, drawing out implications for the use of SPARK in other contexts and the potential for further developments. The final section discusses the implications for the dissemination of web-based templates across academics, subjects, disciplines and institutions.

Research on self and peer assessment of teamwork

There seems little argument about the importance of teamwork in university courses, but assessing it fairly has proved problematic (Conway, Kember, Sivan and Wu, 1993; Lejk, Wyvill and Farrow, 1996). One author has likened team assessment to a game, maintaining that the rules of the game advantage some students and disadvantage others, and that factors such as teamwork and contribution to a team are "essentially impossible to assess fairly" (Pitt, 2000, 240). However, assessment strongly influences students' learning (Ramsden, 1992; Biggs, 1999). Not assessing teamwork, or assessing only on individual components, may suggest to students that collaborative teamwork is not really valued. If our courses have the objective of developing students' capacity to work as part of a team, then we need some means of assessing teamwork in a fair and meaningful way that promotes peer collaboration (Sampson, Cohen, Boud and Anderson, 1999).

One way to increase the fairness of team assessment is to moderate team marks to reflect individual students' contributions. Peer and self assessment of individual contributions should be a useful way of doing this as it gives team members the

responsibility for negotiating and managing the balance of contributions and then assessing whether a fair balance has been achieved. After all, students are better placed than academics to know relative contributions and are keen to have differences in contributions reflected in differences in grades (Fallows and Balasubramayan, 2001). A number of peer and self assessment schemes have been reported in the literature, all involving a process where students rate their peers' contributions and the average ratings are then used to moderate team marks to reflect individual differences. The schemes differ on four main factors: the use of both self and peer assessment or peer assessment alone; the use of holistic or multiple assessment criteria; whether the criteria reflect task or teamwork elements or a combination; and the mathematical formulae used to calculate mark weighting factors.

One of the earliest schemes (Goldfinch and Raeside, 1990) involved peer assessment only and used a two-part assessment form. Students were prompted to identify peers who made the greatest contribution to particular task elements and then use these promptings to give peer ratings on multiple criteria related to team contributions. The academic used the ratings from both sections to adjust individual students' marks. A related approach using task ratings only was used by Conway *et al* (1993) and found to be reasonably well-accepted and regarded as fair by their students. Cheng and Warren (2000) also used peer assessment against multiple criteria to moderate team marks, noting that the approach "facilitates the benefits of group work while providing opportunities for peer assessment" (p. 253). Goldfinch's (1994) adaptation of the Goldfinch and Raeside (1990) approach simplified the rating process. The same type of two-part form was used to prompt students, but the ratings calculation was done only on the team contribution elements listed in the second part. Evaluation of the effect of this change showed that excluding the task prompts reduced the time taken to calculate mark adjustments but made little difference to the size of the adjustments or to the rank ordering of students.

Self as well as peer assessment was also introduced by Goldfinch (1994), to compensate for a problem that she had noted with students who were over-generous to their peers at their own expense. Goldfinch noted few instances of students inflating their self-assessment, but still included a manual process to check and compensate for this possibility. While there is some debate about the validity of using self assessment (Lejk *et al*, 1996), it does encourage students to reflect on their own contributions and capabilities. In fact, Boud, Cohen and Sampson (1999) favour self-assessment informed by peer feedback on specific criteria, in preference to peer assessment per se. More recent studies have compared the use of multiple criteria with a more holistic rating (Lejk and Wyvill, 2001) or have used different calculation approaches (Li, 2001). Most schemes use multiple assessment criteria, allowing for students to rate each other on a combination of task and/or teamwork elements. Lejk and Wyvill (2001) found that holistic ratings had higher inter-rater reliability than those using multiple criteria, however they noted the possibility of student collusion in the process of giving ratings. Furthermore, students in their study knew the multiple-criteria approach would not affect grades and may not have taken it as seriously. Different calculation approaches

have been proposed mainly to reduce possible biases in weighted marks as a result of students being "creative" with their ratings or unconsciously being over-generous. While Goldfinch (1994) applied a general scaling factor based on averaging, Li's (2001) procedure attempted to correct for individual biases.

While there are these differences in approach, overall the research supports the use of some combination of self and/or peer assessment of individual contributions to teamwork. Both students and academics are generally more satisfied with the fairness of this assessment process. However, all of the described schemes have been paper based and involve time consuming data entry and calculations to generate adjustment factors. These factors delay the provision of feedback to students, and create a disincentive for academics, particularly those who are busy or not mathematically inclined. The Goldfinch (1994) and Conway *et al* (1993) simplifications of the original Goldfinch and Raeside (1990) scheme were attempts to reduce the data entry time problem, but still proved extremely time consuming in large classes. Developing a computer-based approach to self and peer assessment seemed a logical solution.

Development of self and peer assessment software

SPARK, a web-based Self and Peer Assessment Resource Kit, was designed to reduce the limitations of paper-based systems and enable self and peer assessment of teamwork to be used with any number of students. The development of SPARK began in early 1996 after an unsuccessful search for similar software. To our knowledge, no equivalent system has been developed to date. SPARK is based on Goldfinch's (1994) well-designed and evaluated paper-based system. It was developed with the aims of improving the efficiency of the process and improving the learning experience for students. The efficiency goals were to automate the processes of collecting student ratings and calculating mark weighting factors, saving both academic time and paper costs. The learning goals reflected a desire to engage students in the self and peer assessment process throughout the team task. Using self and peer assessment encourages students to develop the capacity to reflect on and evaluate their own and others' contributions, and to develop awareness of their own strengths and needs as a team member. We wanted students to have access to the assessment criteria from the beginning of a team task and use them to negotiate their teamwork processes to achieve the best task result with balanced individual contributions by all students. We also wanted them to be able to practice rating each other during the progress of the task and be able to submit their final ratings confidentially and even change their ratings upon further reflection if prior to the rating cut-off date.

SPARK was also designed to be a relatively generic template, easily adapted to any learning context where teamwork and/or self and peer assessment are used. We wished to enable colleagues to use self and peer assessment of team contributions and to adapt the assessment criteria to suit their context. Any assessment process needs to be carefully aligned with learning objectives and teaching and learning activities (Biggs, 1999), and self and peer assessment is no exception. Making SPARK a relatively generic template affords alignment, but as we will show later in this paper, it does not guarantee it.

Description of SPARK and its use in a typical context
SPARK is designed for use in subjects in which:

- the objectives include developing students' capacity to work as part of a team and to reflect on their own teamwork skills;
- there is assessed teamwork;
- there are other forms of web-based learning embedded in the subject.

SPARK contains separate interfaces for "instructors" and students. The academic or "instructor" system is designed to reflect the sequence of processes that academics will go through in setting assessment due dates, creating teams, defining assessment criteria and calculating final adjustments. The student system enables students to practice using the criteria, view sample spreadsheets and information to help them understand the process and submit their assessments confidentially. Both interfaces include help information and frequently asked questions which are accessible from the login screen. For example, the instructor system includes information that academics can include in their subject outlines to help students to understand the rationale for using SPARK, how students can use it to assist their teamwork and how the self and peer assessment will affect their marks. Figure 1 shows the login screen for the instructor interface. Figure 2 shows a typical screen in this interface, showing the menu bar in the top panel and help information in the left-hand panel.

Figure 1: Academic/Instructor login screen

Figure 2: SPARK screen for instructors to enter or import student details

Academics begin using SPARK by entering subject and student details. Student details can be batch imported. Assessment criteria then need to be decided, and SPARK allows for either multiple criteria or holistic rating schemes. Academics also have the option of setting some criteria as prompting only (ie, to prompt students' memories of task and team components but not contribute to mark moderation), and setting criteria which will contribute to moderating a team mark. SPARK does provide a small bank of assessment criteria, but our experience suggests that academics will need to add others, ideally developing these in consultation with students.

How teams are formed and facilitated are important issues but not dealt with in this paper. If the academic chooses to form the teams, then they can be batch imported, otherwise the students can register their own teams. Students can modify self-chosen team memberships until a close-off date. (SPARK does not allow any overlap between the team registration period and the subsequent rating period.)

Once students and teams are defined and assessment criteria are chosen, students can access SPARK as often as they wish to view and discuss the assessment criteria with their team members. After the team task is over, a defined rating period allows students to confidentially rate each member of their team. Figure 3 shows an excerpt from a typical self and peer assessment form, showing some sample criteria and ratings where 0 = no contribution to the team for that aspect, 1 = below average for that team, 2 = average and 3 = above average contribution.

After students have done their assessment, academics can then use the system to calculate various self and peer assessment factors, based on the formulae published in Goldfinch (1994). These factors can be exported to a spreadsheet for calculating individual marks (if the purpose is summative assessment) or used as a source of feedback to students (if the purpose is formative assessment).

Efficient functioning of group:	John Smith	Katie Jones	James Cheung
Helping the group to function well as a team	0	3	3
Understanding what is required	2	3	2
Suggesting ideas	1	3	1
Level of enthusiasm & participation	2	3	2
Performing tasks efficiently	1	3	3
Organising the team and ensuring things get done	0	3	1

Figure 3: Excerpt from a typical SPARK self and peer assessment form

SPARK also has an administrator interface which enables instructor account creation and some other maintenance functions. The system has been designed to run on commonly available systems. The front-end which students and academics see is written in HTML for the web (and not some proprietary network or program), while the back-end database and operating system and programming approach have experienced multiple changes as technologies have advanced in the last 5 years. The version used in the trials described in this paper ran on a Windows NT server with a Microsoft Access database. In 1998, queries to the database were written in Java and Java script. More recent programming has refined and adapted the system to use ServletExec servlets and Microsoft SQL running on an NT server.

Case studies of development, implementation and evaluation

Development of the current generic template has involved an iterative process of developing, trialing, evaluating and refining SPARK over the past two years, based on prototype development over several years. Funding for the development of the generic version was provided through the Committee for University Teaching and Staff Development, an Australian government funding body. The development team comprised academics from five discipline areas across two universities, a programmer and academics from specialist academic development units. One of the disciplinary academics had particular expertise in graphic and interface design. Ongoing quantitative and qualitative feedback from students, academic users, experts and participants in workshop seminars has resulted in continuous improvements to the program. The team has utilised a web-based learning management system for project management, asynchronous discussion of issues, brainstorming and as a repository of key materials.

Educational specialists in self and peer assessment have been consulted for expert reference during the development period at conferences, by visits and by email. Although students are central to any educational context, our evaluation attempted to consider other stakeholders such as academics, technical and administrative staff, academic departments, the institution, and the wider context. This holistic approach reflects our perspective that unless all parties and issues are considered, generic systems are unlikely to be developed in ways which will lead to their effective dissemination and adoption outside the development context. While innovation does of course occur without this perspective, it is our view that advances made by passionate pioneers are more likely to be adapted and disseminated if a collaborative and holistic approach to development and evaluation is pursued. The process also provided information which enables those who adopt or adapt someone else's innovation to better understand the factors which encourage successful implementation. Issues of particular relevance to students and academics are woven through the following four cases. Further details of the findings and broader issues will be published in a later paper.

Subject A: Large first-year subject where the prototype was developed and evaluated

Development of SPARK prototypes began in March 1996 with the aim of improving learning from teamwork assessment, reducing student complaints about free-riding and improving administrative efficiency in a first year subject with a very large enrolment (Subject A). An assessable team task was introduced in the subject in 1992, and involved a case study worth 30%. The assessment task reflected a real world problem that students would face on graduation. The case study was submitted in three parts over the semester, with feedback on each part being built into the next stage. End-of-semester feedback from students indicated satisfaction with the nature of the case study but substantial dissatisfaction with team assessment, due to perceptions of unequal contributions. From 1993 to 1995, various approaches were used to adjust team marks to individual marks. These included student activity diaries, one-line summaries of self and peer assessment and asking teams to divide up the mark. Each of these yielded less than satisfactory results.

By 1996, the students enrolled in the subject had risen to 850 and a series of "homebrew" web pages and simple applets was used to facilitate academic-student and student-student interaction and provide access to additional materials. The first prototype of SPARK was developed and trialed in the web environment. Assessment criteria were developed using a focus group of previous students, which identified 16 sub-tasks involved in the completion of the case study. These task criteria were supplemented by a further six criteria related to team maintenance and leadership roles. Students registered their team (of 3 to 5 students) and had access to the criteria before the case study began. A spreadsheet identifying the results of various ratings was demonstrated and made available. Following submission of the third stage of the case study, they submitted ratings of their own and their peers' contributions.

A number of benefits for students arose from using SPARK. Students perceived the process was fairer because the rating items reflected a range of aspects of the team task

and maintenance roles, and the self and peer assessment could be done confidentially and changed as many times as they wanted prior to the end of the rating period. The latter enabled students to change their ratings privately if others had publicly coerced them. Students also used the obvious nature of the items and method of calculation to manage team effort during the process and even affect their choice of team membership. Open-ended responses on the end-of-semester student feedback survey showed a dramatic reduction in complaints about the team assessment.

The main relevant cost to students was that of access to SPARK. Less than 10% had external web access and students were heavily competing for lab facilities around the time of the deadline. This problem has largely disappeared, as external web access amongst these students is now closer to 90%. The other access problem arose because of occasional bugs in the software. When it comes to assessment, even occasional bugs can be very frustrating and stressful.

Academics experienced a number of benefits from incorporating SPARK. Firstly, it was possible to retain the case study as a valuable learning task. Without a solution to the free-rider complaint, it was likely that the task would have been dropped. Secondly, academics felt satisfied that the process of assessment was fair. This was ethically important as well as increasing student satisfaction. Student comment about the case study being valuable for learning in the subject re-emerged. Thirdly, academics felt a sense of satisfaction that progress was made in a conscious attempt to develop students' ability to work in a team, a capability they knew was important in the profession. Without such easy data collection on multiple rating items, this would not have been possible. Another indicator of success was that despite the growing number of teams, the number of team problems needing academic intervention had reduced substantially. Essentially what remained for adjudication were several of the more challenging cases of free-riding. Time has not changed the latter experience. There were several costs to academics. Firstly, having to reprogram the system for several different platforms was time consuming. Technical bugs were bound to occur in such an environment, as they do in any development mode. But students are very unforgiving when it comes to technical bugs and assessment. Because student feedback was in a public web discussion forum and anonymous, even if only a few students experienced problems, their comments could be loud, strong and very evident to others. While this was very discouraging during the developmental phase, it was a strong incentive to improve. Secondly, in the first three years, the calculation of the factors was very time consuming, taking up to 3 days on Excel because there could be between 250 and 300 teams. The latter was rectified in the generic version. The lessons learned in this large class provided valuable insights into self and peer assessment and teamwork for the wider university community and also helped the subsequent development of the generic version.

Subject B: Mid degree undergraduate subject with one team task

Subject B was one of the first subjects to trial the initial generic version of SPARK, while it was being developed. It is a mid-degree subject in a different discipline from subject

A. Typical enrolments are around 50–80 students with day and evening classes both taught by the same academic. Students use a web-based learning management system within the subject to access announcements, interact with the academic and with each other and access a range of additional subject materials. The assessment in the subject includes a team case study.

SPARK was trialed in the subject in first semester 1999. Because of the development timeline, students first gained access to the system in the middle of the semester, at around the time they started the case study. The assessment criteria were taken directly from those used in Subject A, rather than being customised. The academic perceived that they were sufficiently appropriate as both assessment tasks involved a case study. Like in Subject A, students chose their own teams.

Evaluation of SPARK involved a survey followed by a focus group with both the day and evening classes, and a reflective diary kept by the academic. The student survey included rating and open-ended questions. It asked questions about useability, reactions to the system, and perceptions of learning from the self and peer assessment process. Table 1 shows student responses to some of the rating questions.

The percentage of students who reported that the process had helped them to learn more about teamwork was encouraging, considering that most students had encountered team tasks in previous subjects and the academic did not explicitly emphasise that SPARK could be used for this purpose. It was also interesting that 40% felt it encouraged them to make more effort whereas 33% disagreed. Students who disagreed often commented that they were self-motivated to contribute or wanted to do well and did not need the external incentive to make an effort, but some appreciated that it may have a motivating effect on other students.

Responses to the open questions and the focus group suggested that many students perceived the purpose of the system as encouraging equal contributions by team members, or controlling free-riders. While most students perceived that the system was fair, some disagreed, particularly if they had worked in teams of three rather than four. Some students clearly did not understand exactly how the self and peer assessment

Table 1: Student responses to SPARK in Subject B (n = 48)

	SA/Agree	*SD/Disagree*
The system was accessible	79%	8%
The system was easy to use	70%	13%
The process helped me learn more about teamwork	40%	24%
Identified aspects of teamwork I hadn't thought about before	41%	27%
Items were appropriate for assessing contributions	69%	9%
Encouraged greater effort	40%	33%
Able to give an honest assessment	78%	11%
Fair way of assessing team contributions	69%	18%

ratings would affect their marks. These perceptions appeared to reflect the way that the academic introduced and explained the system. The academic perceived the main benefit to be reducing free-riding, and described it accordingly to students.

Disadvantages for both the academic and the students focused on the useability of the system, in particular bugs and other technical problems which happened during the development phase. Discussions of useability resulted in some changes to the system, including simplifying the password system to make it the same as that used in the web-based learning management system, and providing feedback messages to confirm that students' ratings had been submitted successfully.

Subject C: Mid-degree undergraduate subject with two team tasks

Subject C is an intermediate stage subject in a different discipline again and is taught by a team of several academics. The subject was offered for the first time in the semester when SPARK was trialed and had around 200 students enrolled. Students work on two major assignments in cross-disciplinary teams, and developing the capacity to work in these teams is an important learning objective. Team membership is allocated by the subject academics to ensure that a range of disciplinary majors is represented in each team. Students participated in team development activities in the tutorials before they commenced work on their assignments.

SPARK was introduced and used for formative feedback to teams at the end of the first team assignment and summative assessment at the end of the second task. This was an innovative use and some of the teaching team perceived that it should encourage teams to discuss the way they worked and work more effectively for the second task. The assessment criteria were the same as those used in Subject A. Students gained access to SPARK shortly before the end of the first task. Evaluation of the subject included rating and open-ended questions on SPARK as part of a standard evaluation survey, a student focus group, academic reflection and a focus group with the teaching team. Students were asked fewer questions than in Subject B, because the teaching team wanted to ask many questions about other aspects of the new subject. Table 2 shows some responses.

Introduction of SPARK in this subject had more disadvantages than benefits for students and academics, resulting in some valuable lessons learned for the

Table 2: Some student responses in Subject C (n = 187)

	SA/Agree	*SD/Disagree*
Self and peer assessment feedback after assignment 1 helped the team to work more effectively on assignment 2	20%	47%
Items were appropriate for assessing contributions to the team assignments	42%	37%
Team development tutorial activities helped me learn more about teamwork	36%	35%

development team. Formative use of the system after assignment 1 created breakdowns in some teams when team members who perceived themselves to have contributed equally ended up with different peer assessment ratings. Academics reported more of these team problems than they usually experienced in similar subjects. Several factors appeared to contribute to this. The assessment criteria were taken directly from subject A rather than specifically chosen to reflect the team tasks that students had to do. While some items on the "subject A" form might be said to be generic qualities of teamwork (see Figure 3), others were not. Only 42% of students agreed that the items were appropriate for assessing contributions, compared with 37% who disagreed. In subject B, 69% had agreed and 9% disagreed. Further analysis of the open-ended responses and discussion in the focus group yielded other reasons. Students did not fully understand how the system worked, and in particular how ratings on each of the assessment criteria would affect the overall self and peer ratings. They also felt that they had spent tutorial time in discussing how their teams would work, and their discussions were not reflected in the assessment criteria used in SPARK. Formative feedback was given only in the form of the overall self and peer adjustment factor, rather than as a profile of contributions which could be discussed in a team.

Despite the problems however, students generally perceived that SPARK had a useful purpose if it were appropriately implemented, as illustrated in the following quotes:

"made you think about how much each member and yourself contributed to different aspects of the assignments"

"to evaluate contributions from each team member by team members to get a fair distribution of marks. It still didn't work."

Students also described difficulties with accessing SPARK, and complained about having to make time to access the web and complete the process in a subject where they did not otherwise use the web. Teaching team members also complained about technical problems and difficulties in calculating the required factors for formative purposes. While some teaching team members sought to improve SPARK's use and maintain it in the subject, others sought to drop it entirely. For the development team there were some significant lessons learned, which are discussed more fully in later in this paper.

Subject D: Postgraduate subject with integrated flexible learning

Subject D is a postgraduate subject taught jointly by two academics from different faculties. About 30 students take the subject, which has been significantly adapted from the typical on-campus 13 week semester mode. The "weekly lecture and tutorial" format has been replaced by 4.5 Saturdays of face-to-face contact, about one day per month. In between the block sessions, students complete a range of learning activities by themselves and in teams, with most activities involving interaction using a web-based learning management system.

The subject has eight objectives—two related to knowledge development, four related to the development of capabilities for using that knowledge (eg, critically evaluate

problems and alternative solutions; effectively use analytical tools; competently use technology; communicate effectively to develop and maintain personal and professional relationships) and two objectives related to values (ie, able to work self critically in a group or autonomously; respecting different cultures, ethical and disciplinary approaches). Assessment is aligned carefully with learning activities to ensure subject objectives are achieved. 50% of the grade is allocated to individual work. The remaining 50% of the grade is based on four team assessment tasks. A team presentation worth 20% and 3 team tests where the average of their best two is worth 10% are conducted in class but require significant preparation out of class. A team debate worth 10% and a team topic tracking exercise worth 10%, are completed out-of-class time but submitted online.

With 50% of the grade comprising team assessment tasks, students need to seriously deal with their own and others' abilities to work in a team. Not only do teams in subject D face the possibilities of free-riders, but the potential for dysfunction in teams is higher because of language and cultural differences. Up to 70–80% of the student cohort are international students, coming from a large variety of countries where English is not their native tongue. To optimise the potential benefit of working in a team, the membership is static for the duration of the semester.

Following completion of the final team assessment task, students undertook to rate each team member. In 1998 the self and peer assessment process was completed on paper at the final face-to-face session and then manually entered by academics into an Excel spreadsheet which calculated the self and peer adjustment factor identified by Goldfinch (1994). In 1999, SPARK was used for data entry when the students rated each other online over a one-week period, and academics used it for the subsequent calculation of the self and peer assessment adjustment factors. Sixteen "prompting" criteria were specifically chosen for the four team tasks. Students for example evaluated their own and their peer's on two aspects for the topic tracking task (ie, "quality of postings" and "quantity of postings"), five aspects of the debate, six for the presentation, and three for the tests. This was followed by six "final" criteria relating to an effective team. Evaluation of the process was carried out using student surveys, a structured phone interview with almost all students and reflective journals kept by the two academics who jointly taught the subject.

The phone interviews revealed that the SPARK rating items were appropriate and that most felt it was a fair and honest solution for encouraging teamwork overall. Only some 10–14% disagreed with any of the questions. Most students thought SPARK should be implemented wherever teamwork is used. Interestingly, some 40% said that they did not contribute a greater effort because self and peer assessment was used. Combined with the previous data, this is a positive outcome since it means that free-riding was discouraged without pressuring already committed students to do more work.

Academics found SPARK saved them considerable time previously spent on data entry and calculation. They also felt satisfied that the process had encouraged students to

achieve the subject objectives, including the development of their ability to work in a team, in the more flexible learning mode.

Lessons learned and future potential

Clearly there have been different experiences for both students and academics in using SPARK in these different subjects. These differences pointed to a range of important lessons for the development team. Some are specific to SPARK and others have implications for the use of generic templates more widely. We will firstly discuss those specific to the use of SPARK.

Implications of SPARK for self and peer assessment

Students' feedback on SPARK suggests that most see it as a fair way of assessing team contributions, if it is implemented effectively. There was general support for a system of self and peer assessment. Students in focus groups appreciated being able to rate both their own and their team members' contributions, although there were some comments about individuals who might inflate or be too modest about their own contributions. The SPARK development team has now built in the ability for academics to export separate self ratings and peer ratings and a ratio of these, both to help students to learn about how they are perceived compared with how they perceive themselves, and to allow academics to identify and take action where there are serious discrepancies. Not all students supported the use of SPARK, nor do we expect this will ever be the case. It would be unusual for any form of assessment, particularly team assessment, to receive 100% student approval, but we do know that how the assessment is implemented can have a strong effect on its acceptance and its value for learning. Even if teamwork problems were to be reduced completely for a given cohort following the introduction of SPARK, it is neither our expectation nor experience that this will occur in the typical context. Academics will therefore be required to intervene and adjudicate in the harder cases of free-riding when SPARK fails to motivate appropriate behaviour.

SPARK works best when students can see the valid reasons for having a team task in the subject, and for using self and peer assessment of team contributions. It is only one approach for ascertaining students' contributions to teamwork, and, like all approaches, needs to be educationally justifiable. Academics wanting to use SPARK need to align its use as a learning activity and assessment tool with subject learning objectives such as developing students' teamwork and evaluative capabilities in the context of the discipline. Alignment is important in any subject, as it focuses students' learning towards desired outcomes (Biggs, 1999). Criteria for self and peer assessment need to be aligned with the relevant subject objectives and with the task and teamwork activities necessary for teams to complete their task effectively. Relevant criteria are crucial to the success of SPARK, as illustrated in the differences between subject C and the others where SPARK was trialed. Involving current and/or past students in negotiating the criteria can greatly enhance students' understanding and their perception of relevance and fairness of the self and peer assessment process.

Once criteria are decided, the academic (or academic in conjunction with students) then needs to decide which items will be used in the calculation of final marks. Some criteria may simply prompt students' memories of the task activities that the team undertook rather than affect the self and peer assessment adjustment factor (cf, Goldfinch and Raeside, 1990). On the other hand, all criteria may be used in the final calculations. Whatever approach is chosen, students need to be fully informed about how each of the criteria affects the self and peer ratings which are used to adjust their marks. This is a critical point for increasing students' perceptions of the fairness of the system.

As with any assessment, some students may query their result. Self and peer assessment is no different and some team members may dispute the outcomes of the self and peer assessment process. The message from this is that SPARK is not a hands-off tool that an academic puts into the subject to manage teamwork. Academics are still required to think critically about its usefulness, make the process as transparent and open as possible for students and maintain hands-on processes for communicating with teams and resolving conflicts. Academics need to help students to clearly understand the SPARK process before the team task begins and during the task. To assist in communicating the effects of ratings to students, the student interface for SPARK incorporates a series of examples. Other preventative measures or resolution mechanisms can also help. For example, students can be required to keep an individual and/or team diary of effort and events. This would be the first resource in the event of a dispute.

The above issues point to the need for academics to think carefully about how SPARK is integrated into the assessment for the subject and how it is made clear for students. Another implementation lesson relates to access. If SPARK is the only subject activity which requires access to the web, as in Subject C, students tend to regard it as an add-on and see access as much more of a problem. In subject D where many of the students' learning and assessment activities took place in a web-based learning management system, access was not seen as a problem. Our recommendation is that SPARK only be used in subjects where the web is already an integrated part of the learning environment.

A further major point relates to the context of trialing and evaluating a system while it is still in development. This has some major benefits for progressively improving the system, but also some disadvantages if development work does not keep to a planned timeframe. It is critical for any assessment-related system to be accessible to students as early as possible in the semester and to remain accessible and easy to use throughout the assessment process. Downtime and bugs in SPARK were frustrating for students and stressful for academics if they were unable to gain instant solutions. With a small project under development and one programmer providing technical support it was not possible to provide 24 hours a day, seven days a week support for students and academics. This is an increasing expectation when systems are available via the web. Students and academics need to be aware of this, and academics need to have clear

alternatives available if students find that they cannot gain access to systems at critical periods during the assessment. This can be alleviated in part by providing an extended rating period so students can complete their ratings at a later time.

In summary, the following factors were identified as characteristic of subject environments where SPARK was more successfully implemented:

- Assessment criteria were designed specifically for the specific team task and were aligned with subject learning objectives. Task and team management roles were identified, preferably in negotiation with students;
- The system and assessment criteria were available from the beginning of the team task and could be accessed by students as often as desired;
- Academics were convinced of SPARK's usefulness for learning;
- Academics helped students to gain a clear understanding of how SPARK worked, why it was introduced, how it could be used and the effects it would have on marks;
- Students regularly used other web resources for learning in the subject and doing team tasks, so that web access was not just required for SPARK;
- The system was reliable and accessible throughout the semester;
- Academics felt well supported at all levels by their colleagues and academic departments and by technical support personnel.

All of these points are applicable to the design and implementation of any educational technology tool aimed at improving either learning or efficiency (Alexander and McKenzie, 1998). Most, such as appropriate criteria and clear communication with students, simply relate to overall good teaching practice (Ramsden, 1992; Biggs, 1999). We offer them here both as feedback from our evaluations, and as advice for academics who may wish to use SPARK in their own contexts.

The current version of SPARK has now been used for several semesters, and this use has suggested future developments to extend its functionality. These include:

- Developing ways of providing students with formative profiles of their self assessments and the combined peer assessments against each of the individual criteria. This may enable students to see the differences between their own and their peers' perceptions of their contributions, and discuss these in their teams. This would be considerably more informative and hopefully more constructive than the approach used in subject C where students were given the numbers only;
- Incorporating a normalisation factor to correct for known biases (ie, over-generous or creative raters) along the lines of Li (2001);
- Using SPARK for self and peer assessment of student outputs rather than contribution inputs. Using this online environment, students could assess each other's assignments against summative marking criteria irrespective of whether the academic does as well. This use has not yet been trialed, and careful thinking will need to be done before moving in this direction.

We also believe it has possible uses in other areas beyond academic contexts where teamwork is used, and this may also be investigated in future developments.

Implications for the development and evaluation of web-based templates

The process we used in implementing and evaluating SPARK in a series of subjects also yielded some important insights for those involved in developing other web-based templates. We learned that there is considerable value in implementing and evaluating across multiple subjects and disciplines and with multiple academics. The variation in implementation contexts between subjects served to raise awareness of some taken-for-granted assumptions made in the innovation context, particularly those which relate to the intentions of SPARK, its integration into a subject and communication with students.

For those involved in implementing a template in their subject we recommend:

- Understanding and valuing the learning principles underpinning the template design, and communicating this to students. In the case of SPARK, this meant seeing the template as a tool which could be used to help students to learn about successful teamwork in the subject, rather than merely seeing it as an efficiency device or way of controlling team free-riders.
- Understanding the differences between their own context of use and the template design context. Academics need to adapt the template to integrate it into the learning context in their own subjects—templates are rarely likely to be so generic that they can simply be picked up and used, and the way that they are integrated is critical for student acceptance and learning. Academics need to consider all aspects of the learning context, including subject objectives, assessment, learning resources, students' prior experiences and expectations and the expectations of other academics teaching in the subject.
- Evaluating and disseminating the success or otherwise of their implementation. This should help others think through the issues that they may encounter in introducing SPARK or other generic tools into the learning environments they design for students.

For those involved in developing and evaluating a template we recommend:

- Designing templates that are grounded in student-focused conceptions of teaching and learning (Prosser and Trigwell, 1999; Prosser, Trigwell and Taylor, 1994; Samuelowicz and Bain, 1992), even when they are also designed to improve efficiency. This means that developers need to consider the potential learning benefits of the template tool, and evaluate learning outcomes along with efficiency outcomes.
- Becoming more aware of at least some of the different kinds of contexts in which their template might be used. One way of doing this successfully is to include academics from different disciplines and teaching contexts as part of the development, implementation and evaluation team.
- Making the learning design of the template clear. If the design principles are hidden, academics working with teacher-focused conceptions in their subjects may not implement student-focused templates effectively. Student-focused template designers therefore should try to help potential users become aware of the differences between teacher-focused and student-focused intentions for using the template, and make

explicit links between student-focused intentions and useful strategies for implementation.
- Accompanying the templates with case studies that highlight the critical student-focused integration features which academics could compare with the features of their own context.
- Including a range of examples of implementation and evaluation, both successful and problematic. This should encourage teaching staff to become more aware of the student-focused ways in which SPARK can be used to encourage students to develop their teamwork and self-evaluation capabilities.
- Collaborating with academics from a range of different contexts in the development, implementation and evaluation phases. Evaluations of funded educational technology projects (eg, Hayden and Speedy, 1995; Alexander and McKenzie, 1998) find that dissemination of such projects is very limited, with few being adopted outside the development context. We suggest that dissemination could be assisted by having academics from multiple disciplines as part of project development, implementation and evaluation teams. These teams could then identify critical factors for effective use and provide a range of models or cases to assist in wider adoption.

On the basis of our development, implementation and evaluation, we believe that SPARK has considerable potential as a "generic" template for improving team-based assessment and students' capacity to work as part of a team. Like any assessment, it needs to be thoughtfully integrated into the learning context and its use made transparent to students. If this happens, it has the potential to improve students' learning of teamwork skills and reduce team problems, particularly when used with very large classes. We wish to encourage our colleagues to use any method which encourages students to develop their capacity to work in a team and engage productively in teamwork tasks, and offer SPARK as one approach.

Acknowledgements
Development of the SPARK template was supported by grants from the Business Faculty of the University of Technology, Sydney, and the Committee for University Teaching and Staff Development, funded by the Australian Commonwealth Government. The authors acknowledge the commitment and efforts of other members of the development team: Duncan Ford, Rob McLaughlan, Michael Adams, Darrall Thompson and Bob Spencer. We also are appreciative of the time and efforts of these and other academics who used SPARK in the subjects which formed the basis of the case studies.

Readers interested in installing and using SPARK in their local context should contact one of the authors for details on obtaining a copy of the software and administrator's guide, and may access further information from the instructor home page at http://ntweb1.itd.uts.edu.au/spark/StaffLOG.html.

References
Alexander S and McKenzie J (1998) *An evaluation of information technology projects for university learning* AGPS, Canberra.

Biggs J (1999) *Teaching for quality learning at university* SRHE and Open University Press, Buckingham.
Cheng W and Warren M (2000) Making a difference: using peers to assess individual students' contributions to a group project *Teaching in Higher Education* **5** (2).
Conway R, Kember D, Sivan A and Wu M (1993) Peer Assessment of an individual's contribution to a group work project *Assessment and Evaluation in Higher Education* **18** (1) 45–56.
Fallows S and Balasubramayan C (2001) Multiple approaches to assessment: reflections on use of tutor, peer and self-assessment *Teaching in Higher Education* **6** (2) 229–246.
Freeman M (1995) Peer assessment by groups of group work *Assessment and Evaluation in Higher Education* **20** (3) 289–300.
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) 21–31.
Goldfinch J (1994) Further developments in peer assessment of group projects *Assessment and Evaluation in Higher Education* **19** (1) 29–35.
Hayden M and Speedy G (1995) *Evaluation of the 1993 national teaching development grants* Southern Cross University, Lismore.
Lejk M, Wyvill M and Farrow S (1996) A survey of methods of deriving individual grades from group assessments *Assessment and Evaluation in Higher Education* **21** (3) 267–280.
Lejk M and Wyvill M (2001) Peer assessment of contributions to a group project: a comparison of holistic and category-based approaches *Assessment and Evaluation in Higher Education* **26** (1) 61–72.
Li L K Y (2001) Some refinements on peer assessment of group projects *Assessment and Evaluation in Higher Education* **26** (1) 5–18.
Pitt M. J (2000) The application of games theory to group project assessment *Teaching in Higher Education* **5** (2) 233–241.
Prosser M and Trigwell K (1999) *Understanding learning and teaching: the experience in higher education* SRHE and Open University Press, Buckingham.
Prosser M, Trigwell K and Taylor P (1994) A phenomenographic study of academics' conceptions of science learning and teaching *Learning and Instruction* **4**, 217–231.
Ramsden P (1992) *Learning to teach in higher education* Routledge.
Sampson J, Cohen R, Boud D and Anderson G (1999) *Reciprocal peer learning: a guide for staff and students* University of Technology, Sydney.
Samuelowicz K and Bain J D (1992) Conceptions of teaching held by academic teachers *Higher Education* **24**, 93–112.