

Comparison of PBL Assessment Rubrics

Lyn Brodie

University of Southern Queensland, Toowoomba, Australia
brodie@usq.edu.au

Peter Gibbings

University of Southern Queensland, Toowoomba, Australia
gibbings@usq.edu.au

***Abstract:** This paper investigates the development and use of a comprehensive marking rubric for assessing team-based project work in a core first year problem-based learning (PBL) course. Students work in teams of up to eight to solve open-ended engineering problems and submit their solutions in a project report. The marking rubric is designed to assess all required aspects of the submissions including technical components and reflective aspects where teams reflect on their progress and problems to date and plan for future improvement. Team submissions may be assessed by any one of eleven different markers. Analysis of marking from earlier course offerings shows that this assessment was not consistent between markers, nor did it give constructive feedback to the students. As a consequence the marking rubric was redesigned and was evaluated against earlier marking schemes for consistency between markers and repeatability. Results indicate that new rubric is consistent and repeatable across markers with the exception of one criterion which needs further development..*

Introduction

Assessment, particularly in large classes can be problematic. Providing constructive, timely feedback is difficult, and so too is ensuring consistent marking standards when using several different markers. This is exacerbated when the assessment items are 'open-ended' and the answers are not well defined and depend on student assumptions, for example the initial scoping of a design brief.

ENG1101 has been a core course in all programs offered by the Faculty of Engineering and Surveying at the University of Southern Queensland since 2001. It is the foundation course in a strand of four courses which use a problem-based learning (PBL) methodology.

The course learning objectives include the development and application of skills in basic engineering science (math, physics and statistics), and it also has a large emphasis on the development of teamwork, communication (formal and informal), problem solving skills, self directed learning and reflective practice. In accordance with course learning objectives, it is essential that the assessment criteria used to provide student grades reflects these process skills and not just the outcome of a final technical report (Brodie 2008).

The course uses both criterion referenced and ipsative referenced assessments. Criterion referenced assessments seek a minimum standard of performance for each competency. This involves ordering skills and competencies in a coherent set and providing an overall interpretation of proficiency required. This is similar to standards-referenced which presents levels of performance against agreed quality levels (Griffin, 1991).

Marking schemes were initially established along these lines, with learning objectives established for each problem and four levels of proficiency indicated (poor, adequate, good, excellent) but no other descriptors were provided. Critical analysis of this marking scheme was undertaken. Through an audit and review process (quality control) several shortcomings with marking schemes and process were identified. Of particular concern was that the marking scheme:

1. lacked informative feedback to students;
2. was difficult to apply equitably across teams and with different markers resulting in

inconsistencies between markers; and

3. was not well supported by markers who found significant difficulties with interpretation and application of individual elements of the marking scheme.

Over several offers of this course different marking schemes and assessment methods have been tried in an attempt to deliver consistency between markers, equity for individual students, and quality informative feedback to students. The marking schemes attempted to minimise marker variation even where the content of submission might be quite different depending on the student teams' interpretation of the problem statement and subsequent assumptions. This led to the development of a marking rubric which offers clearer instructions and standards with each criterion often subdivided into several objectives, five levels of achievement for each objective with clear and consistent wording and a range of marks for each level dependent upon the weighting applied to each criterion. See Figure 1.

The new rubric was tested by having several past team submissions remarked by three experienced markers. Results were analysed to determine if consistency between markers was achieved. Markers perceptions to the new rubric were also noted via a survey and focus group. Student feedback surveys are also analysed and presented to determine if student perceptions on useful feedback from assessments has been improved by the new rubric.

This paper investigates the methodology of development of rubrics for PBL with respect to consistency, repeatability, ease of use and provision of feedback to students.

Background

Assessment information can be interpreted within different frameworks such as competency based, task referenced, goal based, and domain referenced (Griffin & Nix, 1991), however there are three major frames of reference that are relevant to this discussion:

1. **norm-referenced** or normative assessment:- This compares relative performances of individuals assessed against what is considered typical or average, hence 'norm' referenced.
2. **criterion referenced**:- This is a measure of competencies against well defined competencies or degree of mastery, both breadth or scope, and depth.
3. **ipsative referenced**:- This is self-referenced assessment of an individual's own interpretation of their performance and development in terms of their own indicators of progress (Griffin & Nix, 1991, p. 93).

Different methods can be used to collect assessment information within each of these three frameworks. Each method has relative advantages and disadvantages and in different contexts one may be more suitable and authentic than others. Thus it is important to consider a range of methods using more than one assessment approach to improve fairness and validity. Dennefer, Henson et al. (2005) also recognised the value of peer assessment for formative purposes (including teamwork and interpersonal skills) in undergraduate medical schools. A range of approaches, including peer assessment, assessment and monitoring of mentoring and reflection is used in ENG1101 to develop team and individual learning goals (Gibbins & Brodie, 2008).

Each of these assessment approaches needs an appropriate, reliable, fair, and equitable marking or grading method. Scoring or marking rubrics are often used for this task. They are popular because they can be adapted to a variety of courses and situations and they have the added advantage of providing feedback as well as a mark. They are especially useful in assessment *for* learning (as opposed to assessment of learning) where the assessment is an integral part of the learning process as it is in ENG1101 (Gibbins & Brodie, 2008). As rubrics contain qualitative descriptions of performance criteria, these can be useful in the formative function of the assessment item. This, according to Popham (1997) suggests that if appropriately designed, marking rubrics can become 'instructional illuminators'.

To achieve this, it is important that the marking rubrics are properly designed. Popham (1997) warned that many rubrics in use were not useful because of design flaws including inconsistencies in the

performance descriptors across the different scale levels. These flaws can affect the instructional usefulness as well as the validity of the marking results. Tierney & Simon (2004) offered some suggestions, examples, guidelines and principles of how to design effective rubrics. Their focus was on consistency of the language used to describe the performance criteria across the scale levels which are designed for both *learning* and *assessment*! The descriptors are important because the descriptive language used communicates the levels of quality expected of the students as well as assessing them. The descriptors and objectives of the assessment item relate to what is valued in terms of the course objectives and informs the students what performance is expected, what level they may be at now, and what level they need to get to. In addition, rubrics facilitate assessment marking and grading if carefully designed with appropriate weighting assigned to criteria and scales.

If graduates are expected to develop as lifelong learners to be prepared for an uncertain future, then they must also become adept at objectively assess their own learning (Williams, 2008). Rather than disempowering learners with strict summative assessments, greater emphasis should be placed on technology-supported tools and techniques to assess context based learning. This will provide opportunities for students to learn to use these tools to critically and objectively assess their own learning and for sustainable assessment of their continuing development throughout their professional careers. One viable alternative to the 'traditional' summative assessments is a well tailored assessment rubric that will focus students' attention on the learning objectives rather than getting marks (Woodhall, 2008). Such rubrics have recently been successfully used to assess, in an 'objective and unprejudiced manner' (Kumar & Natarajan, 2007, p. 100) students' oral presentations as well as contributions to team efforts in the PBL context. Rubrics must be properly designed to facilitate this student learning as well as provide objective assessment of learning objectives.

Design of marking rubrics for observation and assessment of learning is a challenge. But the challenges of doing this fairly, along with providing constructive feedback, are outweighed by the benefits in supporting learner understanding the individual or team progress (Tierney & Simon, 2004). Tierney and Simon (2004) offer examples of poor rubrics particularly those with negative or discouraging wording and vague descriptors. Rubrics should offer a positive view of every performance level on the continuum focussing on what the student can do and offer helpful suggestions for improvement in each of the categories. The literature also offers some 'guiding questions' for well designed and functional rubrics (Sigwart & Van Meer, 1985; Tierney & Simon, 2004). These include:

1. Are all performance criteria explicitly stated?
2. Are the attributes explicitly stated for each performance criterion?
3. Are the attributes consistently addressed from one level to the next on the progression scale?

These questions along with other aspects in the literature guided the design, review and improvement of the rubrics used in this investigation.

Development of New Rubrics

Many different types of rubrics are commonly used in educational contexts. The rubrics developed for ENG1101 can be described as 'descriptive graphic rating scales' because they use generic traits as analytic performance criteria (Tierney & Simon, 2004). They guide the student teams, but without giving specific hints which were intrinsic in the old marking schemes e.g. "appropriate data analysis was done" or "explanation of the physics of heating applied to interior of car". The rubrics have been developed in accordance with guidelines provided in the literature and cover the technical and reflective requirements of the team submissions. They allow for the open ended nature of the engineering projects the student team's scope and undertake as well as PBL specific learning objectives. The PBL learning objectives are largely in the affective domain and have been difficult to assess with previous marking schemes. This is achieved by explicit performance criteria and attributes directly related to learning objectives

The new rubrics give guidance to students on performance criteria to be addressed, specific attributes within these criteria and the weightings applied. At the same time the rubrics are generic enough that

they can be applied to the different design tasks, scope and specifications chosen by different teams. An example of a small section of the rubric is shown in Figure 1.

All assignment submissions in the course are electronic and it is therefore important that marking schemes and feedback are also in an electronic format. The rubric was developed as an *electronic form* (a structured document with areas/spaces reserved for entering information e.g marks which are automatically added, specific comments from marker and tick boxes to indicate level of achievement or standard comment). This allows markers to select an appropriate level of achievement for each objective, add a typed comment, and allocate a mark with the specified range (each level of achievement has a range of marks depending on total assessment mark and a particular weighting e.g. checking and critiquing 5% and for level 5 achievement there is a range of mark – 11.25 to 12. The total mark for the assessment and the weighting for each criterion/objective can be easily changed in the original form document. See Figure 1. When these data are modified, the range of marks for each level automatically updates.

The performance criteria are clearly stated in the left hand column, for example in the ‘Team Reflection and Evaluation’ section one of the listed performance criteria is “Problem solving strategy is researched, documented, applied and tested”. Specific attributes and objectives of this criteria are “Strategy” – a problem solving strategy is research, documented, applied and tested and “Checking and Critiquing (more than simple proof reading) – evidence that team members supplied constructive feedback on critical aspects of the report”. Each of these attributes then has five levels of attainment, with consistent wording, where markers indicate student or team achievement.

Total marks available 250		Total mark for assessment can be modified					
Performance Criteria	Attribute	Level 1 –	Level 2 –	Level 3 –	Level 4 –	Level 5 –	
TEAM REFLECTION AND EVALUATION - 50%							
Problem solving strategy is researched, documented, applied and tested 5% General feedback:	Strategy 0% (feedback only for this report)	0.0 to 0.0 marks <input type="checkbox"/> Report is not submitted or discussion of problem solving strategy not clear or evident	0.0 to 0.0 marks <input type="checkbox"/> Problem solving strategy is poorly researched, documented, applied and tested	0.0 to 0.0 marks <input type="checkbox"/> Problem solving strategy is acceptably researched, documented, applied and tested	0.0 to 0.0 marks <input type="checkbox"/> Problem solving strategy is well researched, documented, applied and tested	0.0 to 0.0 marks <input type="checkbox"/> Problem solving strategy is extremely well researched, documented, applied and tested	0.00
	Checking, and critiquing (more than simple proof reading) 5%	0.0 to 2.5 marks <input type="checkbox"/> No obvious evidence of team members supplying constructive feedback on critical aspects of the report	2.5 to 5 marks <input type="checkbox"/> Few team members supplied constructive feedback on critical aspects of the report but not clearly demonstrated or discussed	5 to 8.75 marks <input type="checkbox"/> At least two* team members supplied constructive feedback on critical aspects of the report and could still benefit from internal critiquing. * No. of active students in the team will be considered in this section	8.75 to 11.25 marks <input type="checkbox"/> evidence that more than two* team members supplied constructive feedback on critical aspects of the report * No. of active students in the team will be considered in this section	11.25 to 12.5 marks <input type="checkbox"/> evidence that most team members supplied constructive feedback on critical aspects of the report	

Figure 1 Section of new marking rubric

The words used (for example: never, seldom, sometimes, usually, always) indicate the scale or level of achievement for each performance criteria attribute. The percentages represent a suggestion on the marks that might be attributed to each of these elements. In accordance with (Tierney & Simon, 2004) the scales that we used were generally: amount, frequency, and intensity as indicated by:

- An example of amount is: not, few, some, most, all.
- An example of frequency is: never, seldom, sometimes, usually, always.
- An example of intensity is: no, weak, some, strong, compelling.

The new performance criteria identify the dimensions of the required performance of a particular skill. This example illustrates how the different levels refer to the development of the skill on a continuum. This can be seen from the main words highlighted in the individual performance criteria.

The criteria that best describes the observed performance is highlighted electronically or annotated in some way. A range of marks is indicated for each level dependent on the overall marks for the assessment piece and the weighting to each criterion. In addition some criteria may be listed 'for feedback only' indicating no contribution to the final marks of this particular assessment, but something that may need to be addressed in subsequent submissions. The main criteria represent broad learning targets, and this increases the usefulness of the rubrics because they can be used universally for each of the projects. Because of this the rubric does not contain specific descriptions related to individual projects or problems, so comment fields and annotation the project report were used to provide this level of feedback. Variability in the use between facilitators is reduced by having facilitator meetings where examples are used to provide consistent interpretation of what is expected as exemplars in each of the criteria. For example, 'clear and concise' becomes much easier for the facilitators to interpret when given some examples of what to look for as possible indicators of when a report might fall into this category rather than one either side of it.

Methodology

Six student team submissions were chosen from a total cohort of 61 teams. These reports were blind marked by three experienced facilitators using the original marking scale. Level of achievement (poor, adequate, good, excellent) along with marks for each section or criteria were recorded. In addition a survey to determine the markers perceptions of the marking scale was administered. These perceptions included:

- The rubric allowed you to assess the report efficiently with respect to time spent on each team report
- The rubric made it easy to identify what element or criteria of the report was being assessed
- The rubric made it easy to choose the appropriate level of achievement
- The rubric made it easy to give an appropriate mark to indicate the achievement
- I am confident in the repeatability of the assessment if I were to mark this same assignment in the future using this rubric
- I am confident that another marker would achieve a similar grade from the same assignment using this rubric
- Overall the grading determined by the rubric gave an accurate indication of the quality of the report.

Over the course of the following semesters a new marking rubric was developed. A review of the problems and course objectives led to listing of specific performance criteria. Clear levels of achievement were added with consistent language for amount, frequency, and intensity. The rubric continuously revised based on literature and input from facilitators. When the new rubric was finalised, the original six team reports were remarked by the same experienced markers. Again the perceptions of the markers were compared using the same questions.

The analysis included:

- The perceptions of the markers with respect to time, repeatability, ease of use, validity and accuracy.
- Comparison of the actual marks for each criteria
- Comparison of the level of achievement for each criteria and objective.

Results

Old rubric

Analysis of the marks and levels of achievement allocated by markers using the old marking scheme indicated a wide range of views and interpretation of the marking scheme despite a face to face meeting prior to starting. The marking scheme could not be considered consistent in any listed criteria in either mark or level of achievement. Analysis of the final mark (total mark 200) for the team report showed a variation of between three and 21 percent between markers for the *same* report. There were discrepancies in feedback on the level of achievement for each criterion, with the possible exception of the criteria of “Spelling and grammar”.

For this criterion the indicated levels of achievement varied only by a maximum of 2 levels e.g. good to adequate or poor to adequate. Marks varied across the three markers from a maximum of 5% to a 2% difference for the total marks allocated for that criterion.

Overall, mark differences and variation in feedback are of considerable concern from a moderation equity and quality control perspective.

The maximum variation for the old rubric was accorded to the criteria of the ‘experimental methodology’ devised by the teams. For this criterion, marks and levels of achievement varied as indicated in Table 1.

There were similar discrepancies for the mark attributed to the team reflection with marks varying from 16 to 30 for team 10, 15 to 25 for team 1 and 4 and smaller variations for the remainder of the teams e.g. 20 to 25, 20 to 24 etc. Results of this variation for each criterion obviously affected the overall mark or grade for the team.

Table 1 Comparison of marks and level of achievement for criterion of experimental methodology

Team	Marker 1		Marker 2		Marker 3	
	Level (of achievement)	Mark/40	Level	Mark/40	Level	Mark/40
4	Good	35	Poor	10	Adequate	20
1	Poor/adequate	15	Adequate/good	25	Poor	12
10	Adequate	20	Excellent	40	Good	25

Perceptions of the markers supported the data which clearly shows the results are unrepeatable, inaccurate and inconsistent. There was no consist response from the markers with respect to efficiency and ease of identifying a particular element to assess. Overall markers believed that it was difficult to give an appropriate mark to indicate a particular level of achievement given the information and guidance provided on the marking scheme.

New rubric

The reactions and perceptions of markers to the new rubric were much more positive. The markers agreed the rubric was efficient to use (with respect to time) even given the increased complexity of the marking matrix. They agreed that the rubric made it easy to

- Identify what element or criteria of the report was being assessed
- Chose the appropriate level of achievement
- Give an appropriate, repeatable and consistent mark for each criterion

In summary they agreed that the over all grading determined by the rubric gave an accurate indication of the quality of the report considering all criteria and objectives that were assessed.

Analysis of the marking data from each of the criteria and objectives supports the postulation that the new rubric is more consistent and repeatable. Four of the teams (X01, 2, 3, 7) showed a total deviation of less than five percent across the three markers, which is considered acceptable. However two teams (4 and 10) showed a deviation of 14% and 13% respectively between marker 1 and the other two markers. Markers 2 and 3 were consistent with each other. See Figure 2

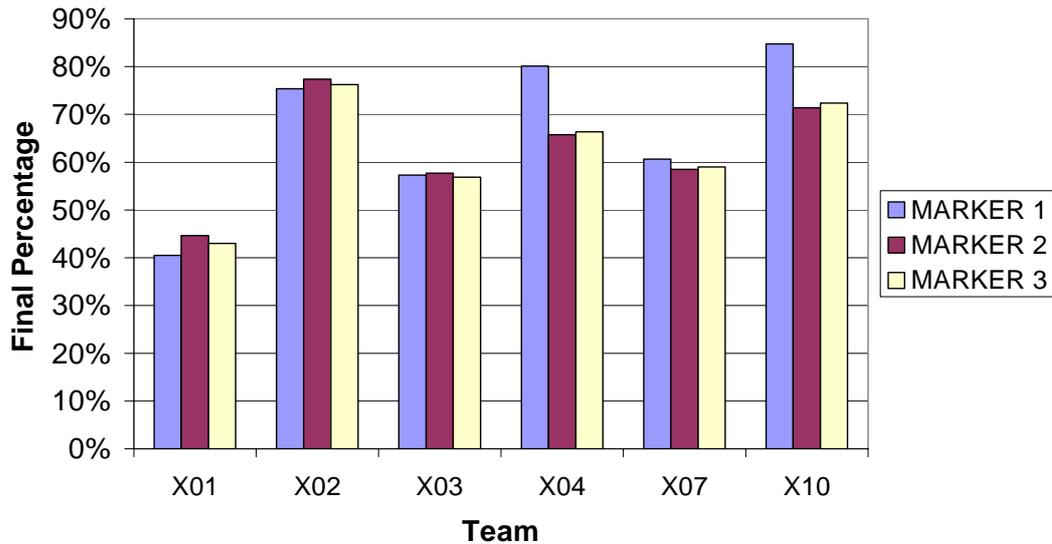


Figure 2 Summary of final marks for each team

The majority of the differences can be accounted for by just two criteria on the report section of the rubric – depth and completeness. These two objectives account for eight percent of the difference in marks. Minor differences can also be traced to the Presentation criterion (and in particular the Language objective) and the Graphs, diagrams and graphics criterion.

When using the new marking rubric, there was consistency between markers in the level of achievement for each criteria and objective. The discrepancy described above relates only to the marks and this is due to the wide range of marks available for each level.

Student feedback and evaluation is a major driver for change in curriculum, assessment and feedback. Whilst there are questions raised over the validity of student evaluations to improve teaching and learning, they do play a critical role in tertiary education. A number of identified purposes of student feedback include diagnostic feedback that will aid in the development and improvement of the course and providing research data to underpin design and improvement to courses (Bennett, Nair, & Wayland, 2006).

Assessment is a key aspect of student evaluations, covering appropriateness of assessment tasks, clear assessment criteria, and feedback provided. Figure 3 shows the results of student evaluation surveys over three years, 2005 to 2007. The original marking scheme was used in 2005. Continuous development of the marking rubric took place throughout 2006 using the feedback from both facilitators (markers), students and

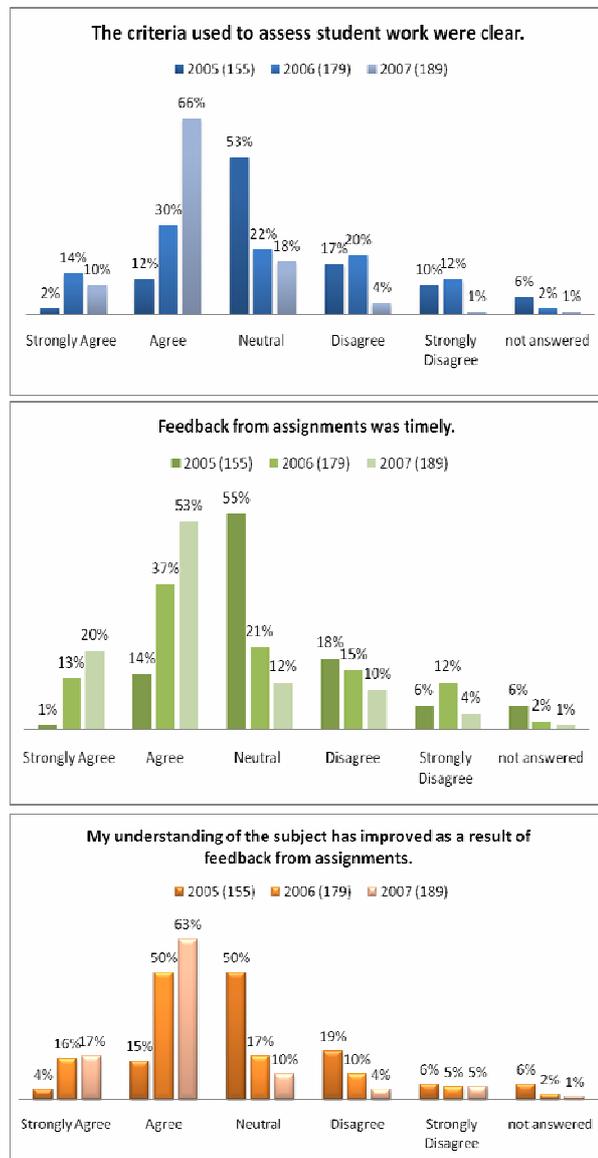


Figure 3 Student survey results relating to assessment over a three year period of development

some analysis of results. The 2006 data informed the development of the new rubric and is included here to demonstrate the temporal changes during the period of rubric development. The new rubric was finalised for use in 2007.

Over this three year period, student evaluations with respect to assessment and feedback continuously improved with results for all three questions showing a positive trend e.g. Neutral and Disagree to Agree and Neutral etc.

Discussion

The new rubric is much more comprehensive than previous marking schemes and spans three pages. It includes a comprehensive set of performance criteria covering teamwork, team reflection, peer mentoring, communication (formal and informal) and the technical components of the tasks. Each performance criteria has specific attributes and more consistent levels indicate achievement levels in all attributes.

Initially, when presented with the new rubrics, markers were somewhat apprehensive and daunted. However, the comprehensiveness of the scheme was soon realised as an advantage since each element and objective is easily identified and the consistent descriptors are easily interpreted.

Elements to note on the new marking rubrics are:

- better clarity of the descriptors leading to easier use and greater consistency and more reliable interpretations by both students and markers;
- the performance levels are much clearer and are plainly differentiated;
- there is only one element to look at in each objective whereas the older rubric often had two or more and sometimes new criteria were introduced across the levels;
- good balance between general wording to make it universally usable for all projects;
- easier use and detailed enough descriptions especially when coupled with feedback on the main project report; and
- consistency across the levels of achievement for each of the attributes by the use of 'parallel language' (Tierney & Simon, 2004, p. 94)

There is generally a positive tone in the rubrics in terms of what was achieved rather than what was not done. This provides motivation to achieve higher levels and puts a positive spin on the expectations to promote learning. However, the rubric does set clear standards and expectations so, in particular, the lower levels do use words such as 'never', 'not present' or 'no evidence provided'. This is clear feedback to missing documentation in the report.

The descriptors for each level deal with the same performance criteria and attribute so the progressive scale is meaningful. Older versions sometimes introduced new attributes or criteria across the levels and this led to some confusion and inconsistencies of markers and generally made it more difficult to use. In the examples above (Figure 1) the same attribute and performance criteria are present; it is just the degree (in terms of amount, frequency or intensity) that changes from level 1 to level 5.

Conclusion

In large classes where multiple markers are used, consistency, repeatability and reliability are integral to equity of assessment. When using the PBL approach, clear assessment criteria for students is required without allowing students to either 'reverse engineer' the solution or guide the direction of research. Furthermore, the same criteria need to be suitable for numerous teams, problems/projects and solutions.

The new marking rubric developed allows the marker to give clear feedback to the students on the current level of achievement whilst effectively guiding students to address the course learning objectives. For example, when addressing the team reflection component the rubric, released at the beginning of the semester, guides teams to the main areas on which they need to focus, e.g. a team code of conduct, not just written but providing evidence of its implementation, review and improvement.

The implementation of a quality review cycle in the course has helped, not only the development of the assessment scheme, but also other general learning and teaching components. It has forced the academic coordinator as well as facilitators to reflect on, review and continuously improve the course objectives, problem objectives and resources and equitable assessment procedures which promote learning.

Considering the improved consistency of both marks and level of achievement, feedback provided to the students and endorsement of the markers the new rubrics are considered successful and far superior to the original. However, further work needs to be done on the criterion of depth and completeness to minimise variation between markers.

References

- Bennett, L., Nair, C. S., & Wayland, C. (2006). *Love it or Hate it: Participation a Key Ingredient in Closing the Loop*. Paper presented at the Proceedings of the AUQF 2006 - Quality Outcomes and Diversity. from http://www.auqa.edu.au/files/publications/proceedings_2006_final.pdf.
- Dannefer, E. F., Henson, L. C., Bierer, S. B., Grady-Weliky, T. A., Meldrum, S., Nofziger, A. C., et al. (2005). Student assessment: Peer assessment of professional competence. *Medical Education*, 39(7), 713-722.
- Gibbings, P., & Brodie, L. (2008). Assessment Strategy for an Engineering Problem Solving Course. *International Journal of Engineering Education*, 24(1, Part II), 153-161.
- Griffin, P., & Nix, P. (1991). *Educational Assessment and Reporting: A new approach*. Sydney: Harcourt Brace Jovanovich.
- Griffin, P., Nix, P. . (1991). *Educational Assessment and Reporting: A new approach*. Sydney: Harcourt Brace Jovanovich.
- Kumar, M., & Natarajan, U. (2007). A problem-based learning model: showcasing an educational paradigm shift. *Curriculum Journal*, 18(1), 89-102.
- Popham, W. J. (1997). What's Wrong-and What's Right-with Rubrics. *Educational Leadership*, 55(2), 72-75.
- Sigwart, C. D., & Van Meer, G. L. (1985). Evaluation of group projects in a software engineering course. *ACM SIGCSE Bulletin* 17(2), 32-35.
- Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels [Electronic Version]. *Practical Assessment, Research & Evaluation*, 9. Retrieved 18 July, 2007 from <http://www.pareonline.net/getvn.asp?v=9&n=2>.
- Williams, P. (2008). Assessing context-based learning: not only rigorous but also relevant. 33(4), 395 - 408.
- Woodhall, T. (2008). *Redesigning Assessment: The Design and Implementation of a Rubric-Based Assessment System to Improve Engineering Design Education*. Queen's University, Kingston, Ontario, Canada.