

# What *Exactly* Do You Want Me To Do? Analysis of a Criterion Referenced Assessment Project

**Tony Jewels**  
**Queensland University of Technology,**  
**Brisbane, Australia**

[t.jewels@qut.edu.au](mailto:t.jewels@qut.edu.au)

**Marilyn Ford**  
**Griffith University,**  
**Brisbane, Australia**

[m.ford@griffith.edu.au](mailto:m.ford@griffith.edu.au)

**Wendy Jones**  
**Queensland University of Technology, Brisbane, Australia**

[wc.jones@qut.edu.au](mailto:wc.jones@qut.edu.au)

## Executive Summary

In tertiary institutions in Australia, and no doubt elsewhere, there is increasing pressure for accountability. No longer are academics assumed *a priori* to be responsible and capable of self management in teaching and assessing the subjects they run. Procedures are being dictated more from the 'top down'. Although academics may not always appreciate 'top down' policies on teaching and learning, they should at least be open to the possibility that the policies may indeed have merit. On the other hand, academics should never be expected to blindly accept policies dictated from elsewhere. Responsible academics generally also need to evaluate for themselves the validity and legitimacy of externally introduced new policies and procedures.

At one Australian tertiary institution, the Academic Board recently endorsed a new assessment policy, to implement criterion referenced assessment (CRA) throughout the university. This new policy is being promoted by the university's teaching and learning support services and is to be implemented gradually throughout the university. As stated in the university's manual of policies and procedures (MOPP) the '*fundamental approach*' to the assessment of students was to be one based on '*criterion referencing*'. Such Assessment was to be free of any notion of awarding grades dependent on a bell curve, and the criteria against which a student's work was to be assessed would have to be communicated in a detailed and clear way in advance of the actual assessment. Thus, for each piece of assessment, criteria must be given to students, with performance standards for each criteria detailing clearly what standard must be reached on each criterion to achieve a certain result. It is stated in the MOPP that CRA will give the assessment process '*a great deal of objectivity*' and will contribute to the '*reliability and validity of the assessment task*'.

Moreover, it is stated that high, but attainable, standards will '*motivate students and focus their energy on learning rather than on competition with peers*'.

---

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact [Publisher@InformingScience.org](mailto:Publisher@InformingScience.org) to request redistribution permission.

The teaching team of the final year Information Technology Project Management subject had a strong commitment to continuous improvement processes in the presentation and running of the subject. Given that the university had en-

dorsed a new assessment policy that would soon relate to third year courses, the team decided to apply for, and were subsequently awarded, a faculty grant to implement and test the possible benefits of the endorsed CRA policy. An independent project, funded at faculty level, was undertaken to evaluate the effect of the new assessment policy on a final year unit.

The project took a team-based approach to studying the use and effectiveness of criterion referenced assessment. All seven members of the unit's teaching and marking groups were included in the decision making surrounding the design, development, and implementation of CRA and an environment was created in which the team members could openly communicate their experience of CRA. Members of the group initially held a variety of beliefs about CRA that ranged from a dismissal of CRA as merely the latest teaching and learning fad, to keen interest in how it might practically operate in the unit. This diversity of thought was regarded as valuable and in keeping with the attempt to take an impartial '*warts and all*' view of the CRA implementation, noting equally its positive and negative impacts.

It was found that many of the purported advantages attributed in the literature to CRA were not forthcoming. Moreover, some unexpected results were uncovered. While prior research suggests that CRA's failures are due to implementation errors, our work suggests that the real problem is that the assumptions behind CRA are flawed. This paper discusses the outcomes of the project for the students, for the academics, and for the institution.

**Keywords:** criterion referenced assessment, accountability, student outcomes

## Initial Literature Review

The distinction is sometimes made in the literature, as it is in the university's policy description, between criterion-referenced and norm-referenced assessment (Brooker & Smith, 1996; Carlson, MacDonald, Gorely, Hanrahan, & Burgess-Limerick, 2000; Dunn, Parry, & Morgan, 2002; Neil, Wadley, & Phinn, 1999; Sadler, 2005). In norm-referenced assessment, people are ranked and given a position on a normal distribution curve, with grading depending on the position on the curve; the top few will get the top grade, the bottom few the bottom grade, and so on. In criterion-referenced assessment, everyone could theoretically achieve the same grade if they all satisfy the criteria to the same extent. As Neil et al. (1999) note, the trend in tertiary institutions to pit criterion-based assessment against norm-based assessment may be based on a lack of understanding of actual assessment practice at universities: norm referencing in our experience and that of Neil et al. is actually not used. It has probably been more used in the secondary sector, where students are typically ranked against each other. Still, it is probably good to have it officially declared that norm-referencing is not part of a university's assessment policy. While it is true that norm referencing is not the usual practice in universities, the same can be said for criterion-referencing. Despite the pitting of one against the other, the usual practice, from what we have experienced, is for academics to have their own criteria and standards for what would constitute a certain grade; the real difference from what CRA university policies want is that detailed written specifications of criteria are usually not given to students. The real question, then, is how well can CRA as endorsed by university policies be implemented and does it achieve positive outcomes.

Other advice in the literature concerned ways, apart from discussion, of educating students about the proposed criteria and standards to reach different grades. Box (2004) and Eccleston (2001) suggest distributing samples of work to students showing different standards. Isaacs (2001) and Rust, Price & O'Donovan (2003) suggest giving students opportunities to mark each others' work using the unit's criteria and standards.

## The CRA Project

From the literature it was quite apparent that all teaching and marking staff needed to be trained extensively in the use of CRA and preferably involved in its implementation. Thus, for example, Barrie, Brew & McCulloch (1999) suggest holding staff development workshops for CRA training. Carlson et al. (2000) suggest that collaboration is a key to success, with the opportunity provided for staff to critique evolving criteria and standards in a face-to-face discussion. It was thus decided that throughout the project, teaching staff would meet for discussion about CRA in general and how it could best be implemented in the unit. The staff would thus be well informed about CRA and about any evolving criteria and standards for different grades.

The team began by considering the unit's overall learning outcomes. Six areas were identified; four associated with technical areas of knowledge and two related to general academic capabilities expected of a third year student. These were:

1. Knowledge of the key components of a project plan.
2. Ability to apply these concepts to the unit case study.
3. Understanding of the key problems and potential solutions present in the case study.
4. Coherency of project plan.
5. Strength of thought / quality of argument.
6. Academic quality of presentation.

Next, the team set about creating a prototype marking grid for Assignment 1 that would reflect different criteria and standards of achievement in these areas. This was a much greater challenge. Within the teaching team, substantial debate took place around the question of granularity. The university's seven-point grading system, together with the six identified areas, suggested that a 6 x 7 grid should be developed. The teaching team viewed this degree of complexity as artificial and confusing because it over-complicated the marking process. In the end, the group agreed to take a simpler approach and defined three standards of performance – *Unsatisfactory*, *Satisfactory* and *Very Satisfactory* – and thus produced a 6 x grid. Further discussion focused on the precise wording of criteria for reaching different standards in the six areas. The final marking grid passed through ten or so iterations before reaching its final form. This prototype marking grid is given in Table 1 (some terms in the marking grid would be clearly obvious only to those who had taken the course e.g. Dag-Brücken, DB, and PMI).

The aim of the prototype marking grid was to allow the identification of weaknesses with the grid and to familiarize the team with CRA concepts and processes. The teaching team used the prototype grid to mark 126 student assignments. After the marking was completed, a focus group was held to obtain feedback from markers about how the marking grid could be improved. It should be noted that for this iteration, students were not provided with the marking grid prior to their assignments being completed. However, as part of student feedback, their marks were provided on the grid.

It became apparent that granularity and the exact marks associated with each standard – *Unsatisfactory*, *Satisfactory* and *Very Satisfactory* – were the main issues that needed to be addressed. Markers felt that they needed more than three levels of achievement. Also, it was noted that the lower marks in the range of marks for *Satisfactory* were *less* than half of those possible for the different areas being considered, which seemed too low. The marking grid was altered to take these problems into account. The new marking grid is given in Table 2.

**Table 1: The prototype marking grid**

| Area of Assessment                                                                                                                                                                                                    | Weight % | Criteria for reaching different standards                                                                                                                                                                      |                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                       |          | <i>Unsatisfactory</i>                                                                                                                                                                                          | <i>Satisfactory</i>                                                                                                                                                                                                                                     | <i>Very Satisfactory</i>                                                                                                                                                                                                           |
| <b>1. Knowledge of the key components of a project plan.</b><br>Identification of the 9 major areas of knowledge required by the PMI in a project plan.                                                               | 9        | 0-3<br>0-4 areas of knowledge identified                                                                                                                                                                       | 4-6<br>5-8 areas of knowledge identified                                                                                                                                                                                                                | 7-9<br>All 9 areas of knowledge identified                                                                                                                                                                                         |
| <b>2. Ability to apply these concepts to the case study.</b><br>Application of these areas of knowledge to the Dag-Brucken project environment.                                                                       | 21       | 0-7<br>Demonstrates limited understanding of the principles and accepted use of these areas of knowledge.                                                                                                      | 8-14<br>Demonstrates understanding of the principles and accepted use of these areas of knowledge.                                                                                                                                                      | 15-21<br>Demonstrates a high degree of understanding of the principles and accepted use of these areas of knowledge.                                                                                                               |
| <b>3. Understanding of the key problems and potential solutions present in the Dag-Brucken case study.</b><br>Evidence that steps have been taken to avoid the problems encountered in the Dag-Brucken case study.    | 35       | 0-11<br>Provides weak / no evidence of an understanding of the project issues present in the Dag Brucken case. Provides weak / no evidence that preventative action has been taken in the current project plan | 12-23<br>Provides some evidence of an understanding of the project issues present in the Dag Brucken case. Provides some evidence that preventative action has been taken in the current project plan                                                   | 24-35<br>Provides strong / plentiful evidence of an understanding of the project issues present in the Dag Brucken case. Provides strong / plentiful evidence that preventative action has been taken in the current project plan. |
| <b>4. Coherency of project plan.</b><br>The project plan is coherent and contributes to the possibility of a successful project outcome. Includes innovation and lateral thinking in contributing towards success ... | 10       | 0-3<br>The project plan is not presented as a coherent document and/or does not convince the reader that the project has a likelihood of success.                                                              | 4-7<br>The project plan is presented as a relatively coherent document that convinces the reader that the project has a reasonable likelihood of success.                                                                                               | 8-10<br>The project plan is presented as a coherent document that convinces the reader that the project has a high likelihood of success.                                                                                          |
| <b>5. Strength of thought / quality of argument.</b><br>Points are clearly stated, justified and logically argued.                                                                                                    | 15       | 0-5<br>Arguments do not necessarily lead to clear conclusions and / or demonstrate critical thinking and justification of points.                                                                              | 6-10<br>In the majority of instances, arguments lead to clear conclusions and demonstrate critical thinking and justification of points.                                                                                                                | 11-15<br>Arguments lead to clear conclusions and demonstrate critical thinking and convincing justification of points. All stated issues have been thoroughly considered.                                                          |
| <b>6. Academic quality of presentation.</b><br>The report is presented to academic standards in terms of spelling, grammar, correct use of figures / tables and correct referencing.                                  | 10       | 0-3<br>There was evidence of many errors in spelling and / or grammar. Some errors in the presentation of figures and tables. Referencing was non-standard, inconsistent, inaccurate and / or absent.          | 4-7<br>Some errors were apparent in spelling and / or grammar. Report was readable. A few errors were made in the presentation of figures and tables. Standard referencing is used throughout the document with a few inconsistencies and/or omissions. | 8-10<br>Spelling and grammar used throughout the document is correct and appropriate for a (referenced) business report. Figures and tables are correctly presented. Referencing is accurate, consistent and complete.             |

Table 2: Final marking grid

| Criterion                                                                                                                                                        | Weighting | 85%-100%                                                                                                                                                                                                                                  | 75%-84%                                                                                                                                                                         | 65%-74%                                                                                                                                                             | 50%-64%                                                                                                                                                                      | 0%-49%                                                                                                                                                                                   |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Understanding of the purpose of the report The ability to: -- understand the rationale for the report. -- identifying and discussing issues central to the topic | 10%       | There is a full and clear understanding of the topic and the reasons for the report. Rationale for the report is clearly stated. There is a full and clear identification and analysis of the key issues.                                 | Central ideas are clearly apparent but may not have been explicitly stated                                                                                                      | Discussion is generally on topic but the ideas have not been clearly identified nor explicitly stated                                                               | Inadequate grasp of the topic although the rationale has been identified There is some attempt at stating the rationale for the report.                                      | Lacks evidence of an understanding of the topic. Is unable to clearly state the purpose of the report and shows no clear understanding of the rationale/purpose                          |
| Identification of mistakes made by DB and the rationale behind those mistakes                                                                                    | 25%       | Clearly identifies all or most of the mistakes made by DB and the rational behind those mistakes, clearly linking the mistakes with the theoretical foundations of each                                                                   | Clearly identifies many of the mistakes made by DB, and links those mistakes with the theoretical foundations of each                                                           | Identifies many of the mistakes made by DB but fails to link those mistakes with the theoretical foundations                                                        | Identifies some of the mistakes made by DB but fails to link those mistakes with the theoretical foundations                                                                 | Failure to identify many of the mistakes made by DB or failure to identify the theoretical foundations                                                                                   |
| Understanding of how to avoid the mistakes of DB with the incorporation into a project plan of relevant practices and procedures                                 | 30%       | Clearly identifies how most of the mistakes of DB can be addressed by incorporation into project plan relevant practices and procedures. Clearly shows how these practices and procedures will help prevent the mistakes from reoccurring | Incorporates into project plan practices and procedures that will help prevent many of the identified mistakes from reoccurring, and identifying why these practices are needed | Includes within project plan practices and procedures that indicates a clear understanding of how they would prevent a reoccurrence of many of DB mistakes.         | Includes within project plan practices and procedures that would help prevent the major DB mistakes and acknowledging why these practices/procedures would be needed         | Inclusion into project plan few practices and procedures that would help prevent DB from making the same mistakes and a failure to acknowledge how practices/procedures would be needed. |
| Structure and strength of argument -- contribution to success The ability to: -- present a persuasive and logical argument with a system approach                | 15%       | Clearly persuades the target audience that the proposed project plan overview will help contribute to success, clearly linking the set of components to overall outcomes                                                                  | The target audience is reasonably convinced that the project plan overview would lead to success and the set of individual components are linked to overall outcomes            | The target audience is somewhat convinced that the project plan overview would lead to success without linking the set of individual components to overall outcomes | The target audience is only partially convinced that the project plan overview would lead to success with no linking of the set of individual components to overall outcomes | Fails to convince the target audience that the project plan overview would lead to success with no linking of the set of individual components to overall outcomes.                      |

|                                                                                                                                                                                                                                                                     |     |                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                         |                                                                                                                                                                          |                                                                                                                                                                                          |                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Referencing The ability to -- consistently use the correct referencing style -- reference all sources                                                                                                                                                               | 10% | Referencing is consistently accurate and complete and is used appropriately in context                                                                                                                                                                            | Referencing is mostly accurate and is used appropriately in context                                                                                                                                     | Referencing is fairly accurate and is used appropriately and in context                                                                                                  | Referencing is generally accurate and is used appropriately or in context                                                                                                                | Referencing is not accurate and not used appropriately nor in context                                                                                                   |
| Expression and presentation The ability to: -- use fluent language and correct grammar, spelling and punctuation use appropriate report format. Paragraphing and sentence structure -- use appropriate style and tone of writing -- produce a professional document | 10% | Consistent use of standard grammar, spelling and punctuation. The sentence structure is coherent concise unified... Fluent professional style No proof reading errors Professional standard presentation                                                          | Grammar and spelling are correct Language fluent Few proof reading errors Professional presentation                                                                                                     | Language mainly fluent but some minor grammatical or spelling errors Presentation not of a professional standard Tone and style of writing inappropriate eg too informal | Meaning is apparent but language is not fluent. Grammar and or spelling contains errors. Inappropriate vocabulary, style or tone of writing Presentation not of a professional standard. | Meaning is unclear due to frequent grammatical/spelling errors. Disorganised and incoherent writing.                                                                    |
| Personal quality of project team members                                                                                                                                                                                                                            | 40% | Provide comprehensive, appropriate and logical set of personal qualities. Provide good evidence of the importance of each individual quality in contributing to success Provide evidence of how the set of individual qualities (combined) contributes to success | Provide significant, mainly appropriate and logical set of personal evidence. Provide evidence of the importance of each individual quality and the combination of qualities in contributing to success | Provide a range of personal qualities that would contribute towards success and show how they would contribute.                                                          | Provide a range of personal qualities that would contribute to success but without showing how they would contribute                                                                     | Provide an unsatisfactory, inappropriate or illogical list of personal qualities that are not backed up by any evidence as to how they might contribute towards success |

With a marking grid that had been developed with best practice in mind and that had undergone many iterations and pre-testing, the teaching team was ready for a full implementation of CRA with the following research questions in mind:

1. How do students view their experience with CRA?
2. How does the teaching team perceive their experience with CRA?
3. Are student outcomes improved with the use of CRA?

## **Methodology**

For assignment 2 of the third year IT Project Management course held one semester after the development and pre-testing of the marking grid, students were given the assignment and the criteria sheet given in Table 2. There were 148 students enrolled in this course, 35 females and 113 males.

As there was some concern whether students provided with marking criteria would ignore the assignment question itself and simply write to the criteria, a decision was taken to provide the assignment question one week in advance of the release of the marking guide. As in the previous semester, the assignment asked students to consider a case study of a failed project that had been discussed throughout the semester and to devise an overview of a project plan that could have been used to increase the likelihood of success of the project. Also as in the previous semester, they were asked to give the personal qualities of project team members that they would consider important in achieving project success.

Rather than just providing the students with the grid, the primary instructor of the unit discussed the criteria with students. To avoid any confusion, students were specifically asked to direct all questions about criteria and standards to the primary lecturer and course coordinator.

After completion of the unit, a student focus group was held to determine students' views of their experience of CRA. The student focus group consisted of 7 students who volunteered to be part of the group and included representation from international, domestic, part-time, full-time, male and female students. To determine how the teaching team viewed their experience, there was also a focus group for them. Both focus groups were run by an independent person, experienced in running focus groups. Student outcomes were compared with outcomes in the same subject for the same piece of assessment from the semester before.

## **Results**

### ***The Students' Experience***

#### **Costs**

The students who participated in the focus session were, in the main, negative about their CRA experience. Instead of clarifying what was required to achieve different grades, the students said the criteria sheet caused confusion, as indicated by the following three quotations:

*Actually I was confused by the criteria. It was almost as if you if you had received the criteria, the assignment specifications or just the Coordinator's remarks separately you would have understood what to do but all three together sort of created the confusion and I think they were just inconsistent.*

*The criteria for the assignment didn't come out until after we received the assignment specs so we'd had a couple weeks to get our thoughts together, then we got the criteria and then we had to discuss it all over again in my group and so actually I was confused by the criteria.*

*I thought the assignment specifications were clear but then when we got the marking criteria it confused me and it would have been better not to have the criteria.*

The focus group, without any lead from the facilitator, spent a significant amount of time examining the relationship between the marking criteria and creativity. They felt strongly that the criteria sheet stifled creativity, as shown by the following quotations:

*In this situation I don't think there's room for creativity. [The Unit Coordinator] is teaching us project management processes and we understand that, but there's no creativity at all in the assignment, there's only the criteria that they will use to mark us on.*

*I think the criteria should add in an extra section called "x-factor" or creativity which may be 20% of the actual marks so people can do other things and include their own ideas.*

*It's a double-edged sword in a way because if you don't have criteria you have no way to mark the assignments but if you do have it you are (directing) everyone down a particular path, the one they have to go down if they are to get 100%.*

*Personally I don't like the criteria at all because it stops creativity but if you have to use it I think you should focus the core elements you need to pass at around 50% then allow a certain amount like 30% for recognizing the value of different methods of approaching the problem.*

*If we are always to use criteria how are we supposed to come up with new things? How can we students come up with new ideas if we always have to stay within the criteria.*

*I understand that in this unit [the Unit Coordinator] wanted us to develop your own ideas and approach to the assignment but because there's such a strict marking criteria you can't use your creativity and write something that you think you know might be appropriate, you've got to stick to the marking criteria.*

The students also believed that the criteria sheet had not enhanced their learning:

*For learning purposes, I think the criteria didn't help at all. It didn't get you to explore new ideas and do research. I think the marking criteria is mainly for people who are just interested in marks and getting good grades for the subject.*

*I don't think you can use the criteria for learning purposes. I thought the tutorials were a lot more creatively inspired. In tutorials you got to sift through all the possibilities and talk with the tutor to find out what they thought. There was more feedback and more room for creativity. The assignment just said this is the criteria and you have to write to that and that's what you're going to be marked on. So as far as learning more of the content from the assignment I don't think it really happened.*

*I don't know whether I'm going to find out if I know the subject better or worse through the marks I get for my assignment. I'll just find out whether I conformed or not to the criteria.*

## Benefits

The students all agreed that some marking criteria were better than none at all and that the weightings given to various criteria had helped them apportion their effort:

*The criteria were useful because they were weighted differently. For example, the referencing part is only worth like 10% or 20%, whereas other issues are more heavily weighted but without a criteria that you wouldn't know that and you might get a low mark because you should have spent more time on a heavily weighted part of the assignment.*

*I did a professional studies unit last year. The first assignment came out without marking criteria and all the students said, "You need to tell us what you want". Despite all the feedback going back to the tutors they still didn't end up putting out an actual criteria sheet for it and everybody did badly.*

*Having some marking criteria at least gives you a bit of something to work with rather than flying completely blind.*

## **The Teaching Team's Experience**

### **Costs**

Staff felt that the implementation of CRA required an enormous amount of time. The development of the marking criteria and standards through iterations and the pre-testing took great time and effort. Literally hundreds of hours were dedicated by the teaching team to these tasks, none of which was paid.

The teaching team was also concerned that after the release of marks to students, student complaints peaked. Students argued that the wrong criteria had been specified for the assignment, disputed their rankings on the associated standards, and protested that there was a gap between the assignment specifications and the marking criteria that had caused confusion. In the two weeks following the release of marks to students, the unit coordinator's usual five hours of consultation per week rose to 13 hours in order to handle the volume of student complaints. The project team was extremely disappointed. As a teaching team member explained:

*This was a real blow as our reading of the literature had led us to expect the very opposite. We believed that the release of marking criteria would reduce any ambiguity surrounding the assessment item and that the students' understanding of why they had received their final marks would be vastly improved. What was most interesting about the level of student complaints was that the usual level of questions leading up to the assignment due date was way down. I think these two changes were connected, that is the marking grid gave students a degree of false confidence about their assignments which was shattered when they received their final grades.*

Apart from concerns about time and student complaints, some markers felt that the marks obtained when using the criteria sheet were inappropriate:

*Students could receive good marks without doing a quality assignment.*

*It was too easy for students to end up with good marks.*

*In some cases, I wanted to mark lower, but the criteria didn't allow it.*

*I would rather use professional discretion. I know what an assignment is worth.*

Also, some markers felt that the criteria forced a focus on pieces of the assignment, so that the "whole" was missed:

*I don't like giving a mark for independent pieces of the whole.*

*The criteria break it into bits when we want students to see the whole.*

The teaching team was also still concerned with the granularity of the criteria sheet:

*It still needs more definition for the 0-49 area*

*The marking grid needs more granularity and more categories.*

## Benefits

The unit coordinator felt that the teaching team experienced a substantial improvement in the consistency and quality of marking through the use of marking criteria and standards. He thus achieved productivity gains in the area of the moderation of marks. The majority of markers also realized higher levels of productivity and increased confidence in their assessment work.

Some members of the teaching team felt that the simple exercise of developing and implementing a marking grid had benefited them, both as individual teachers and as a teaching team. This benefit came as a pleasant surprise to the project team as it had been unexpected and was not mentioned in the literature. One tutor explained the situation:

*The very best thing about the CRA project was that we talked as a group about our marking practices. Time is in such short supply that it's easy to let this type of reflection slip and miss the learning opportunities it provides. The CRA project gave us a reason to get together and really focus on what we hoped to achieve as a team. I learnt a lot from the others and feel that I have now taken a step forward in my professional skills.*

Some markers said they appreciated having the marking criteria:

*The marking process was faster.*

*It gives a good backup for defending marks.*

*I really enjoyed having the MC. I found it clear and very helpful.*

## Student Outcomes

### Costs

The range and distribution of grades were compared with those from the preceding semester and no difference could be observed. At first glance, this might be taken to indicate that the standards reached by the students remained the same as in previous years. However, the markers, for the most part, felt that they were *marking more easily* with the criteria sheet, that is, marking generously. With this in mind, it would seem that in actuality the standards dropped. In fact, markers complained that they felt that students' creativity was stifled and that the criteria did not raise standards:

*I felt that some students wrote to the criteria.*

*Students follow the criteria and it's too mechanical.*

*The X-factor needs to be considered.*

*Students seemed to be comfortable with the criteria sheet, but they were doing the wrong thing.*

*I got less questions, but the reality is they got it wrong.*

*Standards have definitely dropped*

The unit coordinator agreed. He felt student performance had dropped significantly; in particular, that despite the team's best efforts, students were responding to the assignment criteria rather than the assignment question. It was as though students didn't have to 'think' anymore, but instead went directly to the criteria and lost what had been seen with previous cohorts in the area of creative and original thought:

*In our experience in ITB272, CRA closed down the out of the box element in student assignments, the original take. It could be argued that CRA in some ways defines creativity as the deviant behavior described by Csikszentmihalyi [1994]. Students faced lots of risks in going outside the box of the marking grid for no apparent rewards. Our marking grid acted as a disincentive to original thought by limiting students' expression. This is at odds with deep learning which generates creativity. We want our final year students to be creative so in some ways the outcomes of the CRA project were contradictory.*

## Benefits

There were no observed benefits of CRA to student outcomes.

## Discussion

The study reported here came about because of the endorsement of a new assessment policy by a university and its promotion as a very positive thing by the teaching and learning support services. The results were, however, disappointing.

Let's first consider the fact that the staff felt that the CRA implementation was very time consuming. Although it would be expected that the first implementation of CRA would lead to an initial increase in workload, the teaching team was surprised at the level of the increased workload. Further, despite many iterations, the team still felt there were problems with the granularity and marks awarded using the grid. When this is added to the fact that teaching teams are not stable, in that staff members come and go and continual training is required, the overheads in terms of time are enormous. For the staff, and hence the institution, there are significant negatives in terms of resources.

The results concerning the students were also disturbing. Apart from the fact that some students had complaints and others were confused, faculty believed that the quality of student work dropped and both faculty and students believed that creativity was stifled.

Given the somewhat negative nature of the students' perceptions and their outcomes and given the concerns of the teaching team, we need to consider why our results were not those expected by the university.

It was noted in the introduction that Box (2004) and Eccleston (2001) suggest distributing samples of work to students to show them examples of different standards. Also, Isaacs (2001) and Rust, Price & O'Donovan (2003) suggest giving students opportunities to mark each others' work using the proposed criteria and standards. Neither of these tactics was taken in educating our students about the criteria and standards because they were not feasible. The criteria to be written would relate specifically to assessment for the unit on IT Project Management and would refer to items relating to particular cases studied. Thus, samples of work on the assessment could not be distributed. One could spend lecture and tutorial time discussing such topics as, for example, 'what makes an argument logical' and showing examples of work that to varying degrees give a logical argument. However, that material seemed more appropriate for another subject itself and the allotted time for the core topics of the unit was already taken up.

The following questions now arise:

- How crucial is the giving of examples of work to be examined or marked by students?
- How have other educators coped with giving examples of work marked with CRA?
- Although CRA was being promoted, a fact which prompted this study, what are the experiences of other tertiary educators using CRA.

## ***A Closer Examination of the Literature***

It would be normal practice to give a review like the one to be presented here in the introduction to a paper. We give it at this point instead because we were prompted to examine the literature in more detail after our study, having found a major discrepancy between our findings and the benefits promoted by the university's teaching and learning support services. It turns out, in fact, that there are very few studies examining the use of CRA in tertiary institutions and that what does exist indicates a less than positive effect of CRA.

Brooker & Smith (1996, p.166) noted that "although there is little, if any direct empirical evidence demonstrating the educational value of making explicit assessment criteria available to students in advance of the task, a variety of theoretical frameworks would suggest its benefits". In their study, they interviewed lecturers who used pre-specified criteria in their assessment and students who had done the assessment. They found that while 79% of the lecturers thought their criteria were well understood, only 39% of the students said they understood the criteria well. Further, while 67% of the lecturers felt that the requirements for doing the assessment were well understood, only 15% of the students agreed with this. Brooker and Smith noted that no lecturer had used examples of work to illustrate the standard required for a particular grade. They suggested that this raises the issue of the extent to which assessment criteria should be made explicit. They also noted that it could be argued that "too much specificity in this regard constrains students by limiting their scope for creativity and personally meaningful approaches ..." (Brooker & Smith, 1996, p. 173). In spite of their results and the issue of stifling creativity, Brooker & Smith still seem, in the end, to be in favour of CRA, suggesting that lecturers need to be better educated about assessment practices.

Neil, Wadley, & Phinn (1999, p. 307) discuss the use of CRA in tertiary institutions and, noting that the development and application of CRA "requires considerable time and effort", they present a "generic framework for criterion-referenced assessment ... to reduce [the] load". Perhaps the framework is of use, but there is no study of its usage. Moreover, Neil et al. note that Hay (1995), who had been a proponent of CRA (Abbiss & Hay, 1992), reversed his support of CRA, giving the following reasons for his 'retreat' from CRA:

- it is inflexible, not allowing for individual students
- students are encouraged to conform, potentially discouraging exploration and experimentation
- the establishment of appropriate weighting systems defies experts
- different markers might assess criteria in different ways, though still agree on the final mark
- students might strive to satisfy individual criteria and lose sight of the 'whole'
- establishing the relationship between criteria and ideas/originality/innovation/genius is virtually impossible.

Like Brooker & Smith (1996), Neil et al. did not implement their own usage and study of CRA and despite being aware of negative features of CRA pursued the task of developing a framework to reduce the load when implementing it.

O'Donovan, Price, & Rust (2001) did do a study at tertiary level, examining students' perceptions of a quite generic CRA grid that had previously been developed. They found that one quarter of the students reported difficulty using the grid. The general consensus was that the grid could be helpful in showing what is expected and the relative weightings of different components. However, there were a number of criticisms of the grid. A major criticism was that the criteria were too broad and vague and that there needed to be clarification of what the terms and phrases on the grid meant; for example, what does "address them comprehensively" mean? They also stated that students regarded CRA "as a well-conceived assessment tool that is, however, of limited practical use if presented on its own without the benefit of explanation, exemplars and the opportunity for discussion". O'Donovan et al. also note the "imprecision inherent in verbal description" of criteria and standards. Again, like Brooker & Smith (1996) and Neil et al. (1999), O'Donovan et al. remain in favour of CRA and believe effort should be put into somehow facilitating a common understanding of the criteria and standards between staff and students, perhaps through the use of exemplars and more discussion.

Dunn, Parry, & Morgan (2002) give a theoretical discussion of CRA in tertiary institutions. They note that "criterion referencing requires considerable negotiation to arrive at agreed criteria and standards" (Dunn et al., p. 2). They also note that Hagar, Gonczi, & Athanasou (1994) argue that CRA only assesses trivial, atomistic tasks and that Wenger (1998) argues that assessment criteria cannot convey every possible criteria. Regardless of these problems, Dunn et al.'s assumption, as with the authors already discussed, is that CRA is worthwhile and that the problem is in the difficulties with implementing it, which, it is assumed, can somehow be overcome.

Carlson et al. (2000) did a study where lecturers developed their own marking grids. Some lecturers did make use of exemplars of work, either from past years or by work they wrote themselves. Students did seem to appreciate these. However, limitations were noted, particularly the difficulty of deviating from the sample provided; as one student commented, "I found it difficult to deviate from the sample. It was so good and I found myself writing the same sentences" (from Carlson et al., 2000, p. 112). Students did appreciate that they would not be marked on a bell curve. Again, though, there were complaints that the criteria were useless as they were too vague.

Sadler (2005) has noted that CRA has not "delivered on its aspirations". Once more, problems are acknowledged, but it is assumed that procedures can be found for it to work. Sadler stresses the use of exemplars to show what is required for different standards, though he reports no study showing the feasibility or use of exemplars.

## ***The Present Study***

We have seen, in the results of the present study, that many of the problems referred to in the literature are present. It is true that there were markers in the current study who appreciated the marking grid. Given the other problems observed, however, alternatives, such as group discussions before marking or showing markers examples of work already marked by an experienced marker, might be just as appreciated. Giving students exemplars of work along with a CRA marking grid, on the other hand, is problematic in terms of feasibility and the problem that students may be unlikely to deviate from a '*model*' answer.

Particularly disturbing was the observation that there was a stifling of creativity and a working to individual criteria, while missing the whole. Sadler (2005, p. 178) states that "students deserve to know the criteria by which judgments will be made about the quality of their work". This simple

statement is perhaps what proponents of CRA have in mind when they support CRA despite the problems. There is a certain sense in which Sadler's statement is correct. If, for example, it is not obvious from the assessment task that certain things will be considered, then these should be spelt out. If, for example, the markers expect that a certain published article should be discussed, and if, given the material students have been given, that would not be obvious, then the students should be told. However, it seems that spelling out standards for all criteria is counterproductive. Does it really help to say that a logical argument will receive a higher grade than one that gives a disjointed argument, or that an assignment with poor spelling and grammar will receive a relatively low mark, or that when asked to write a project plan to avoid pitfalls in a previous case study, that you will receive more marks if you can identify the past problems and propose ways of solving them? Don't we want to graduate students who can recognize *for themselves* what is wanted by some assignment, just as an employer wants workers who can work out *for themselves* what would lead to a good piece of work? Also, don't we want students to strive for giving the best solution they can, not simply aim for a less than top grade? As one student remarked to us at an end of year gathering "*I only looked at the left hand side of the marking grid. Why would you look at the rest of the sheet?*" Perhaps implementing CRA is taking the wrong tack. Perhaps instead of CRA for each unit, students need to have exposure to a specific unit where they develop skills in determining what would count as good solutions to any given assignment, along with learning generic skills such as making logical arguments and communicating effectively. Students and graduates would then be more self-reliant, more self-directed, and not have to ask the question: What *exactly* do you want me to do?

## Conclusion

The present study stemmed from the endorsement and promotion of an assessment policy by a university. The teaching team of a final year IT Project Management team developed and implemented CRA in their unit. The results were disappointing; students were confused about the criteria, felt their creativity was stifled, and staff felt the standards had dropped. A more critical review of the literature suggested that many problems with CRA are known, but that most authors keep wedded to the notion that theoretically CRA should be good and that we must persist with somehow educating lecturers to implement CRA in a better manner. While prior research suggests that CRA's failures are due to implementation errors, our work suggests that the real problem is that the assumptions behind CRA are flawed. We suggest instead that it would be more appropriate to educate students on skills that apply to all subjects and assignments, whether they be university or work assignments. Students and graduates need to develop a sense of independence and confidence in what they do; being able to determine what would constitute a good solution to a given task is an important part of what they need to learn.

## References

- Abbiss, J. & Hay, I. (1992). Criterion-referenced assessment and the teaching of geography in New Zealand universities, *New Zealand Journal of Geography*, 94, 2-5.
- Barrie S.C., Brew A. & McCulloch M. (1999). Qualitatively different conceptions of assessment criteria. Paper presented at AARE - NZARE Conference, Melbourne, Australia.
- Box, I. (2004). Object-oriented analysis, criterion referencing and Bloom. *Proceedings of the 6<sup>th</sup> Australasian Computing Education Conference*, Dunedin, New Zealand.
- Brooker, R. & Smith, D. (1996). Assessing tertiary students in an education faculty: rhetoric and reality. *Higher Education Research & Development*, 15, 163-175.

- Carlson, T., MacDonald, D., Gorely, T., Hanrahan, S., & Burgess-Limerick, R. (2000). Implementing criterion-referenced assessment within a multi-disciplinary university department. *Higher Education Research & Development*, 19, 103-116.
- Csikszentmihalyi, M. (1994). The Domain of Creativity. In D. Feldman, M. Csikszentmihalyi & H. Gardner (Eds.), *Changing the World: A Framework for the Study of Creativity*. Westport CT: Praeger.
- Dunn, L., Parry, S., & Morgan, C. (2002). Seeking quality in criterion referenced assessment. *Learning Communities and Assessment Cultures Conference*, Northumbria, England.
- Eccleston, K. (2001). 'I know a 2:1 when I see it': Understanding criteria for degree classifications in franchised university programmes. *Journal of Further and Higher Education*, 25, 301-313.
- Hagar, P., Gonczi, A., & Athanasou, J. (1994). General issues about the assessment of competence. *Assessment & Evaluation in Higher Education*, 19, 3-15.
- Hay, I. (1995). Communicating geographics: development and application of a communication and instruction manual in the geography discipline of an Australian university, *Journal of Geography in Higher Education*, 19, 159-175.
- Isaacs, G. (2001). *Assessment for Learning*. Brisbane, Australia: The Teaching and Educational Development Institute, University of Queensland.
- Neil, D. T., Wadley, D. A. & Phinn, S. R. (1999). A generic framework for criterion-referenced assessment of undergraduate essays. *Journal of Geography in Higher Education*, 23, 303-325.
- O'Donovan, B., Price, M., & Rust, C. (2001). The student experience of criterion-referenced assessment (Through the introduction of a common criteria assessment grid). *Innovations in Education and Teaching*. 38, 74-85.
- Rust, C., Price, M. & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28, 147-164.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30, 175-194.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge, England: Cambridge University Press.

## Biographies



**Dr. Tony Jewels** is a lecturer within the School of Information Systems at Queensland University of Technology, Brisbane, coordinating IT project management units. An IT professional for over 30 years, he has worked in a variety of roles on many leading edge projects throughout the world. His practical experience combined with academic qualifications in business management and information technology provides students at both undergraduate and postgraduate level with real world perspectives of life within the IT sector.



**Dr. Marilyn Ford** is an associate professor in the School of Information and Communication Technology and a member of the Applied Cognitive Neuroscience Research Centre at Griffith University. She has publications in the areas of reasoning, education, and sentence perception and production.



**Wendy Jones** is a practicing project manager working currently within the higher education sector on curriculum development and grant applications. She tutors in IT project management at Queensland University of Technology, Brisbane, and has had an extensive professional career in project management within the IT sector.