



Applying Rubrics Assessment in Undergraduate Computer Science Education

**Margot Schuhmacher
Selby Markham**

Monash University
Caulfield East, AUSTRALIA

margot@mail.csse.monash.edu.au

ABSTRACT

In a tertiary environment assessment of students' work can be difficult and time consuming for lecturers and their assistants. Teaching staff must provide students with instruction and guidance to enable them to complete required assessment tasks. In order to enhance assessment consistency, marking guides for staff use are generated. In spite of well-developed approaches by teaching staff there are growing concerns over marking reliability, standards and the interpretation of these for grades. Students are often unclear about what is expected of them in assessment situations. One approach to dealing with these issues, a rubrics assessment method, is presented here in the context of an undergraduate computer science unit. The method being reported provides clearer guides for students in their preparation of the assessable component, as well as making the assessment for the lecturer and assistant teaching staff simpler, less time consuming, and more consistent between the staff. Furthermore, the positive feedback approach used in this process assists the student in the understanding of his or her own learning.

Keywords: Scoring rubrics, criterion reference assessment, marking reliability.

1. INTRODUCTION

In a tertiary environment, lecturers typically devise their own framework and content for an educational unit. A key component of the teaching framework is to assess the students' learning of the content being taught, resulting in a mark or grade awarded to the student. Setting assignments and exams, and developing marking guides for these usually achieve assessment of learning objectives. An important aspect of the assessment procedure is to ensure the marking of the assessable components are reliable and follow a standard. As pointed out by O'Donovan, Price and Rust (2001), there is now a growing concern for marking reliability and standards, calling for public accountability.

Experienced lecturers can often view assessable student tasks such as assignments and easily sort them into grades. They have a mental picture of what meets the assignment requirements resulting in a pass grade, what is clearly an exceptional piece of work deserving a distinction or higher distinction, and what constitutes a fail. But how do we convey this concept to students, marking assistants and newer inexperienced lecturers, and how do we account for standard and reliable marking?

Primary or secondary school teacher training include courses on how to teach, set assessments, award marks and grades, and an exposure to criterion-referenced assessment (Black and Dockrell



1980,1984;Price and Rust 1999;Turner 1980). Criterion-referenced assessment is an assessment where an individual's performance is compared to a specific learning objective or performance standard and not to the performance of other students. Criteria are guidelines, rules, characteristics, or dimensions that are used to judge the quality of student performance. Scoring rubrics (Moskal 2000) are based on criteria and define what the criteria mean and how they are used.

In Australian universities, tertiary lecturers may have formal training in educational delivery issues, and attend course like those offered to staff at Monash University (Celts 2002). For example, Monash's Higher Education Development Unit (HEDU) provides a Graduate Certificate in Higher Education that includes instruction on how to set assignments and tests, and how to mark and grade these, including the use of criterion-referenced assessment (Celts 2002). This is a required course for all new academic staff, but other staff may not be aware of the possibilities in using a descriptive assessment such as criterion-referenced assessment or scoring rubrics. This issue is exacerbated by the fact that the use of such an assessment method can be quite complex and requires high staff and student involvement for it to be of value (O'Donovan *et al.* 2001). Through mentoring and experience, staff then develop their own assessment methods and grading of students.

In this paper, we describe the introduction of scoring rubrics for practical assignments and grading in Information Technology Project Management (CSE2203) at Monash University. This is a compulsory undergraduate unit in the Bachelor of Computing comprising some 300 students over two teaching semesters. Section 2 gives an insight into CSE2203 and section 3 describes how the assessment method was introduced in this unit. Positive student and staff evaluations are reported in section 4, with shortcomings highlighted. We conclude with further work proposed in this area in 2002.

2. UNIT BACKGROUND

Information Technology Project Management (CSE2203) introduces the fundamental principles, tools and techniques of software project management. The conceptual material presented in lectures is reinforced by practical application within the context of a software development project.

Students work in project teams with roles allocated to each group member. The project is defined against a set process model. Project definition, estimation, and tracking and reporting techniques presented in lectures are employed during the course of the project. Real-life problems are injected into the project in the form of changes to user requirements, budget and time-lines. Emphasis is placed on the ability to provide up-to-date management information on the actual state of the project against established milestones; reports are requested on an ad-hoc basis. A project review phase is used to analyze and report on project estimates against actual time, cost and resource expenditure.

Following principles of the Team Software Process (Humphrey 2000), the project process incorporates two cycles. Assessable components include a project plan, design documents, test plan, product prototype, and project reflection report for the first cycle, and technical review, customer presentation, product manager presentation and final postmortem in the second cycle.

2.1 PREVIOUS UNIT ASSESSMENT

Teaching staff consists of the unit leader (lecturer) and four to six tutors to assess the practical components of the unit. The previous assessment structure included a marking guide for staff to help ensure consistency in assessment. Table 1 gives an example of the format used and it can be seen that fractions of marks were allocated for each item. It was constructed with descriptive comments such as 'award up to ½ for this item' or 'award up to 1 mark for this item'. Tutors then used their own judgement as to the qualitative value of the item. The example marking guide (table 1) was for the project definition documents (project plan, quality plan and test plan), which are the first submission students produce in cycle one of the project. A final mark for the submission was then calculated by summing the three components. The assessment was returned to the students with the grade they achieved of type HD (high distinction), D (distinction), C (credit), P (pass), NP (near pass) or N (fail) based on the final mark achieved out of the available six marks.

It can be seen that this was a fairly opaque process from the point of view of the student, and experience in using it showed that some tutors had difficulty with it. The decision was made that an assessment

Component	Project Plan	Quality Plan	Test Plan
Descriptive Item	Award up to ½ mark for each of the following: Introduction, project organization, managerial process, technical process, schedules and budgets, risk management plan	Allocate up to 1 mark for each of the following: Coding and design, standards, requirements validation	Allocate up to 1 mark for: Test cases for the whole system
Available Marks	3	2	1

Table 1. Example Marking Guide

method was needed that worked towards the following aims:

1. Achieve more uniform assessment amongst staff
2. Make assessment simpler for staff
3. Provide a clearer guide to students for the assessment criteria with the description of the submission requirements
4. Give feedback to students on their learning and achievements.

It was decided to use the scoring rubrics assessment method with its implementation described in the next section.

3. SCORING RUBRICS

Scoring rubrics are a descriptive scoring scheme to guide assessment of students' work when a quality judgement is required (Moskal 2000). Similarly criterion-referenced assessment tells us how well students are performing on specific goals or standards. Scoring rubrics are commonly used in college and public schools, and give valuable feedback for students and parents (Moskal 2000, Chicago Public Schools 2001). A database for assessment rubrics can be found at Kathy Schrock's Guide for Educators (Schrock 2001), with further sample assessment rubrics for courses, class participation, research papers, group presentations, instructional units, and lesson plans available on the Web and in educational literature (Price and Rust 1999, Rubric 2001, Turner 1980). The common criteria assessment grids presented by O'Donovan (2001) are similar to the scoring rubric descriptive described in this paper.

The rubric organizes and clarifies the scoring criteria so that two teachers who apply the rubric to a student's work will generally arrive at the same score

(Chicago Public Schools 2001, Moskal 2000). Using a rubric is common for writing samples (Moskal 2000) and reduces assessment time and stress for written papers (Colorado School of Mines 2001). They are not however limited to writing samples, and can be applied to evaluate group activities, projects, oral presentations, english, mathematics, and science (Moskal 2000).

3.1 DEVELOPING A SCORING RUBRIC

To develop a scoring rubric you must first identify the qualities that need to be displayed in a student's work to demonstrate proficient performance (Moskal 2000). In CSE2203 'exceptional' was used for the proficient criteria. An example descriptive for an 'exceptional' test plan was defined as: 'summary includes page number, date of test and comment for test. Tests cases for the system address the total system functionality and are correctly formed'. Further examples for the project and quality plan, design documents and presentation style are given in table 2, with an example product rubric given in table 3. Once the exceptional criterion is defined, the lowest criteria for performance are defined, demonstrated by the examples in table 2 and 3, which were termed an 'amateur' score. An amateur score is the equivalent of a fail grade in university education. For example the descriptive of 'amateur' for a product assessment was: 'limited functionality implemented, and/or poor design, which is not easy for the user to navigate. Product not designed to meet all functionality. Poor or no validation'. However, students awarded a score of 'amateur' rather than a 'fail' grade receive positive feedback, encouraging them to investigate and understand where the performance was lacking.

	Exceptional	Admirable	Acceptable	Amateur
Project & Quality Plans	All areas of plan addressed correctly including deliverables, project scope, functional specification, team roles, goals & measures, risk management plan, configuration management, schedule with tasks & estimates. The quality plan includes well-formed code, document and design review standards, and a review checklist.	As per exceptional with up to 2 areas not addressed satisfactorily. e.g. No plan for addressing risks, no measures of goals; checklist not satisfactory.	As per exceptional with up to 4 areas not addressed satisfactorily, OR, 2 areas not addressed.	Project plan requirements Not understood. Few areas completed satisfactorily.
Test Plan	Summary includes page number and comment for test. Test cases for system addresses the total system functionality and are correctly formed.	Summary page as per example only, with test as per 'Exceptional'	Insufficient tests but correctly formed.	Test cases not presented correctly e.g. Testing for buttons rather than business functionality
Design Docs	High-level design, class charts, use-cases and screen design all addressed	As per 'Exceptional' with 1 area deficient eg. Insufficient use-cases	As per 'Exceptional' with 2 areas deficient	Design documents unsatisfactorily presented
Presentation	All documentation presented professionally and to a standard with footers, identification and version control. Document control evident. Guide Part I Completed.	As per 'Exceptional' with one area omitted	As per 'Exceptional' with two areas omitted	Unsatisfactory presentation. Does not understand professional documentation

Table 2. Example project definition assessment rubric

	Exceptional	Admirable	Acceptable	Amateur
Product Demo	Meets all functionality in the design. All functionality fully implemented. Exceptional design with HCI principles in mind. Validation (error checking) implemented. Setup/exe file	As per 'Exceptional' with up to 2 areas of functionality not fully implemented, but includes validation	As per 'Exceptional' with some functionality not fully implemented. i.e. some areas are hard coded. Limited validation	Limited functionality implemented, and/or poor design, which is not easy for the user to navigate. Product not designed to meet all functionality. Poor or no validation.
Change Request	Functionality designed and fully implemented in product design. Integrated neatly into existing product.	Functionality designed and partially integrated into product.	Functionality designed into product, but not implemented.	Not included (No bonus marks)

Table 3. Example product assessment rubric

	Exceptional	Admirable	Acceptable	Amateur
Project & Quality Plans	?			
Test Plan		?		
Design Docs				?
Presentation			?	
Overall Grade: Admirable (Distinction)				

Table 4. Example assessment award matrix

	Exceptional	Admirable	Acceptable	Amateur
Project & Quality Plans	1.5	1.2	0.9	0.6
Test Plan	1.5	1.2	0.9	0.6
Design Docs	1.5	1.2	0.9	0.6
Presentation	0.5	0.4	0.3	0.2

Table 5. Example assessment matrix value

Exceptional is in the range 4.1 - 5.0 this relates to a HD
Admirable is in the range 3.1 - 4.0 this relates to a C or D
Acceptable is in the range 2.1 – 3.0 this relates to a P or NP
Amateur is 2 or less this relates to an N

Table 6. Final score tally

After the top and bottom end of the rubric are defined, a middle level of performance can be determined resulting in a three score level. More than three score levels can be defined by creating greater distinctions between the levels.

In CSE2203, a four score level of ‘exceptional’, ‘admirable’, ‘acceptable’ and ‘amateur’ was chosen. These would relate to the grades equivalent to: high distinction (HD), distinction (D) to a credit (C), pass (P) to a near pass (NP), and fail (N). Other examples of score level definitions are:

- ◆ power, proficient, intermediate, novice,
- ◆ distinguished, very good, satisfactory, unsatisfactory,
- ◆ distinguished, proficient, intermediate, novice,
- ◆ exceptional, effective, acceptable, unsatisfactory;
- ◆ expert, practitioner, apprentice, novice, and
- ◆ excellent, satisfactory, unsatisfactory (Rubric 2001).

3.2 USING A SCORING RUBRIC

The four score level rubric described in section 3.1 was used in CSE2203 in 2001 for all student practical submissions for the assessable components. Tutors used the four score level rubrics descriptives (table 2, table 3) for each submission component and ticked the appropriate score achieved in the assessment award matrix (table 4). The rubric score then relates to a matrix value mark as demonstrated in table 5, and a final rubric score and grade are derived from the standard defined in table 6.

To demonstrate the process, an example of a student assessment is described for a project definition. The descriptive used is the project definition assessment rubric (table 2). This student example assessment received ‘exceptional’ for the project and quality plans, ‘admirable’ for the test plan, ‘amateur’ for design documents, and ‘acceptable’ for presentation (table 4). Using the assessment matrix (table 5) the marks are tallied as: $1.5 + 1.2 + 0.6 + 0.3 = 3.6$.

The tallied mark is then matched to a score or grade from the final score tally descriptive in table 6. For this particular example (with a mark of 3.6), the student received an overall ‘admirable’ score or ‘distinction’ grade, which is then included in the assessment award in table 4. The assessment award matrix (table 4) is returned to the student who can cross-reference the assessment rubric (table 2) to see where they demonstrated or lacked understanding of the requirements.

Feedback received from tutors and students at the end of the semester 1 survey showed it would be valuable for students to have the assessment rubric guide prior to submitting their assessment component. In semester 2, students were supplied with the assessment rubric guides (e.g. tables 2 and 3) as part of the assignment requirement, giving them further knowledge of the requirements, as well as the type of work needed to achieve a particular grade.

3.3 REFINING THE DEVELOPMENT OF A SCORING RUBRIC

In CSE2203, scoring rubrics were initiated using the process previously described. Since the introduction of this process, further refinement has been included each semester. The experience showed that the descriptives themselves could be subjective, and needed to be further clarified with examples, along with assignment exemplars where possible. Including tutors in this process was very useful. The ‘exceptional’ descriptive for a test plan in the project definition assessment (see table 2) has now been expanded to Summary includes page number, date of test, stage passed (prototype, not implemented, complete) and comment for test. Test cases for the system address the total system functionality – matches use cases, and functional specifications from the project plan. Test for negative cases/errors included. Tests are correctly formed – name, input data, expected result.’

4. EVALUATION

The unit lecturer conducts a survey (both student and staff) at the completion of each semester for evaluation purposes. The student survey is tailored to the unit and follows a customer satisfaction model to student evaluation of teacher performance (Guolla 1999, Postema and Markham 2001).

From the lecturers' perspective:

- ◆ It took more time up front to prepare the assessment guides
- ◆ It was easier and quicker to mark the assignments
- ◆ It was easier to clarify student queries on the assessment awarded by tutors (sometimes students query the grades they received and a simple reference to the rubric score criteria clarifies this)
- ◆ Greater consistency of marking between tutors was apparent with the more uniform approach.

4.1 STAFF EVALUATION

Tutor evaluation of the rubric assessment method is shown in table 7 for semester 1 and semester 2 respectively. There were 4 tutors in semester 1 (three were experienced and one was a novice), and six in semester 2. The semester 2 tutors consisted of three who had tutored in semester 1 and three new to the unit, but with prior tutoring experience.

Three questions were included in the survey and were ranked on a 1-5 Likert scale, where a high score indicates a favourable response. Question one asked about the ease of assessing student assignments using this method where 1 corresponds to 'hard' and 5 is 'easy'. The mean response was 4.5 and 4.0 for the respective semesters, indicating the method was easy for tutors to use. The next question asked for

an evaluation of the method compared to other units they tutored, where a response of 5 would indicate it was 'less difficult'. The mean for this question was 5.0 and 4.0 for the 2 semesters respectively. This is a very good result with a strong indicator that the method is easier for tutors to use than methods deployed in other units they tutor in. Finally an evaluation of the time to mark compared to other units was included, where 5 would indicate 'less time'. The response was 4.5 and 3.0 respectively for the two semesters. A very positive response in semester 1, reduced to no change in semester 2. A qualitative analysis of comments made by tutors indicated they spent more time on assessment in this unit than in other units they teach, due to the fact that this unit comprises many more assessable components (particularly documentation), requiring a lot of reading time.

The survey also included a provision for comments on the educational value of the assessment method. The feedback in semester 1 was very positive with comments on 'its clarity and organization, practicality, not needing to rely on minute dissection and allocation of marks, and gives students feedback to improve their work'. A negative comment was that students were unaware of where they would lose marks if they just followed the assignment specification. This was addressed in semester 2, by providing students with the assessment rubrics as part of their assignment specifications. The only comment in semester 2 was that the method was good, but too general, with a request for more specific detail of mark allocation. This comment would indicate that further thought into the rubric construction with examples would be valuable. As O'Donovan (2001) point out, the formation of descriptives requires a team effort. Terminology needs to be clear and examples need to be provided to students and tutors.

Criteria	Semester 1	Semester 2
Ease of assessing student assignments	4.5	4.0
Ease compared to other units	5.0	4.0
Time compared to other units	4.5	3.0

Table 7. Tutor ranking of assessment method

Tutors need to clarify and discuss the descriptives with the students, rather than just make them an available document (Black and Dockrell 1984, O'Donovan 2001).

4.2 STUDENT EVALUATION

The unit survey had a low response rate of 38 out of 100 students in semester 1, from an online survey. Semester 2 survey reverted to the traditional paper one, and received responses from 175 students out of a possible 200. Students were asked to rate the usefulness of the assessment method based on scores in the range 'exceptional' to 'amateur' with descriptives provided. The students' ranking of this proved to be positive with a mean value of 3.3 (an a 5-point scale) in semester 1 and 3.4 in semester 2. Compared to other types of assessment method, the rubrics score method was ranked as 3.2 and 3.4 for semester 1 and 2 respectively (see table 8), confirming that this type of assessment method is better than what they had experienced previously. As was already mentioned, students were supplied with the assessment guide in semester 2 at the time of the assignment requirements distribution. The average response for the value of this was rated as 3.6, confirming the importance of having assessment criteria up-front.

5. CONCLUSIONS AND FUTURE WORK

Assessing student submission components is not an easy task. Educators need to prepare marking guides that not only fairly evaluate the assessable

component, but additionally ensure consistency in marking between various assessors.

Marking standards, reliability and accountability are increasingly coming to the forefront. Students also need to receive positive and constructive feedback upon the return of their submissions, indicating where they have succeeded and/or need improvement to guide their learning process.

This paper has shown how an alternative assessment method based on scoring rubrics can be used in a computer science unit. The method requires more time upfront to prepare assessment guides and it is essential to distribute these to students prior to an assignment submission being due. Hence students have a clearer indication of the submission requirements. Positive feedback received from student survey evaluations on the assessment method along with the fact that teaching staff found the method useful, less time consuming, and easier to use than those used previously in this unit, or in other units, highlight the method's value. Greater consistency of marking was also apparent.

Some shortcomings have been identified which need to be addressed in future work. These include further thought into the construction of the rubric, including a more detailed breakdown of the components and more specific descriptives for these. The development of this will require input from tutors involved in the assessments, as well as discussions with students, and provision of examples to illustrate the grade requirements. A database of rubric scores is anticipated to eventuate from this study, which can be adapted to other computer science units in our school.

	Semester 1	Semester 2
Having guide was useful	N/A	3.6
Assessment method was useful	3.3	3.4
Compared to other assessment methods this is useful	3.2	3.4

Table 8. Student ranking of assessment method

REFERENCES

- Black, H.D. and Dockrell, W.B. (1980).** Diagnostic Assessment in Secondary Schools: A Teacher's Handbook. Scottish Council for Research in Education.
- Black, H.D. and Dockrell, W.B. (1984).** Criterion-referenced assessment in the classroom. Scottish Council for Research in Education.
- Celts (2002).** Centre for learning and teaching support. <http://www.celts.monash.edu.au>. Accessed March 2002.
- Chicago Public School. Instructional Intranet (2001).** Available online: <http://intranet.cps.k12.il.us/> Accessed October 2001.
- Coloroda School of Mines (2001).** Campus Writing Program. Available online: <http://www.mines.edu/Academic/lais/wc/wac/reducegrading.html>. Accessed October 2001.
- Guolla, M. (1999).** Assessing the teaching quality to student satisfaction relationship: Applied customer satisfaction research in the classroom, Journal of Marketing Theory and Practice, 7, 87-97.
- Humphrey, W.S. (2000).** Introduction to the Team Software Process. Addison Wesley.
- Moskal, B.M. (2000).** Scoring Rubrics: What, When and How? Practical Assessment, Research & Evaluation, 7(3). Available online: <http://ericae.net/pare/getvn.asp?v=7&n=3>. Accessed January 2001.
- O'Donovan, B. Price, M. and Rust, C. (2001).** The student experience of criterion-referenced assessment (through the introduction of a common criteria assessment grid). Innovations in Education and Training International; London; February 2001.
- Postema, M. and Markham, S. (2001).** Student Satisfaction: A method for exploring quality factors within computing education. Proceedings of the 14th Annual Conference of the NACCQ July 2001, Napier, New Zealand. Wickliffe Press Dunedin
- Price, M. and Rust, C. (1999).** The Experience of Introducing a Common Criteria Assessment Grid Across and Academic Department. Quality in Higher Education, Volume 5, No. 2, 133-44.
- Rubric (2001).** Available online: http://129.7.160.115/COURSE/INST_5931A/Rubric.html Accessed March 2001
- Schrock, K. (2001).** Guide for Educators. Discovery School. Available online: <http://school.discovery.com/schrockguide/assess.html>. Accessed November 2001
- Turner, Eileen (1980).** Criterion referenced assessment: An Annotated bibliography. Stirling Technical Reports in Education No. 1 University of Stirling

