

Case Study 2: Assessing large projects thoroughly and qualitatively

Geraint A. Wiggins

April 30, 2008

1 Introduction

There is a perennial problem in higher education of how to mark large-scale items of work fairly, when these work items are essentially qualitative—that is to say, there are not clear right and wrong answers. Also problematic is a general tendency to shy away from giving marks at the extreme of a marking scale, which penalises students strongly in a system where percentages are added at unit level to compute a final mark.

In Computing, we have a further problem, which is that staff are not accustomed to marking essay-type questions, in general, and therefore it is even more difficult for them to achieve effective marking of student projects. In Computing, the final year project is the centre around which the rest of assessment orbits, and so it is particularly important that it is appropriately and consistently marked.

2 The Problem

Until recently, there were broadly two ways that computing projects were marked, throughout the UK. The first is by detailed breakdown of marks, into percentage units, for each part of a task, which are then summarised by addition. The second is by holistic, unstructured marking.

The first approach does tend to cover the whole gamut of marks. But it often fails to capture specific, higher-level concepts, such as synthesis. These, however, can be anticipated, and given their own explicit batch of marks. A bigger problem is with the naïve linearity of addition—useful in supermarkets, but not necessarily helpful when something quite simple might indicate a big difference between two low-standard projects, but that same thing would be trivial or even irrelevant in distinguishing two high-standard efforts.

The second approach is weak on detail, and it is very hard to maintain consistency between markers and between projects. When a holistic mark is produced, there may not be explicit reasoning about the different aspects that contribute to it, and therefore there may be considerable variance between projects of the same standard, unless the marker is very careful and very skilled. Since it is human to err, we should not assume that this is a safe way of marking. And experience clearly shows that holistic markers do tend to shy away from extremes, raising the question of “what are the extremes for?”.

A further problem with the marking of work with numbers is that there is often a discrepancy between the mark given by different markers for the same level of achievement: this is why grade related criteria are an aid to standardisation.

3 The Solution

To solve this problem, we use a qualitative marking sheet, based on the learning outcomes of the project module, and on a reasoning system devised by Prof. Qiang Shen (formerly at Edinburgh, now at Wales, Aberystwyth). Prof. Shen is expert in automating qualitative reasoning systems such as this.

The system works by first identifying the distinguishing criteria in the assessment, and grouping them into three: basic, advanced and exceptional. The basic criteria are essentially the learning outcomes, and

the requirement is that to pass with honours, each of these should be at least “adequate”. The rest of the scheme is then aimed at distinguishing the levels of achievement of those who pass. The advanced criteria are focused around second class level, and the exceptional ones around 1st class, though the boundaries are deliberately (and desirably) fuzzy.

Each criterion is separately graded on a scale of inadequate, adequate, average, good, and excellent. These grades are written on a sheet, and the grades alone are then used to read off the mark range of the project: we have two versions, one with 10-mark ranges, and one with 5-mark ranges, and the former is attached in appendix. Once projects have been grouped by mark range, they can be placed in order within that range (again by reference to the mark sheets), and awarded a number accordingly.

Experience shows that the vast majority of students develop a profile which is relatively high on basic criteria, but which then tails off across the others. The algorithm mentioned above (see appendix) captures this nicely, but it is acknowledged that non-standard profiles may need individual attention: for example, a project may be inadequate in one basic aspect, but excellent in all the advanced ones, and in such a case, individual judgement must be used. Intervention is entirely appropriate in a context where one is allocating 16% of a students’ final mark for one (albeit large) item of work—an entirely automated system is likely to disadvantage some students.

Our projects are blind-double marked. Using this scheme has reduced disagreement between blind-double markers significantly, in Edinburgh, at City, and here. External examiners have commented favourably on the precision and level of the scheme, from the point of view both of process and of standards. The scheme and the accompanying instructions, are given in appendix.

CIS320: THIRD YEAR PROJECT ASSESSMENT

COMPUTER SCIENCE

NOTES FOR GUIDANCE

Your ultimate aim is to allocate an overall numerical mark on the College-wide marking scale:

$$\text{Fail} < 35\% \leq \text{Pass} < 40\% \leq \text{III} < 50\% \leq \text{II.2} < 60\% \leq \text{II.1} < 70\% \leq \text{I}$$

To arrive at a mark, follow the guidelines given below. These refer to the criteria listed in the assessment report form.

- 0-24: Bad Fail** The project is inadequate in four or five basic criteria.
- 25-34: Fail** The project is inadequate in two or three of the basic criteria.
- 35-39: Pass** The project is inadequate in exactly one of the basic criteria.
- 40-49: III** The project is adequate on each of the basic criteria.
- 50-59: II.2** The project is at least average on each of the basic criteria and is average on most of the additional criteria.
- 60-69: II.1** The project is at least good on each of the basic criteria and is at least average and sometimes good or excellent on each of the additional criteria.
- 70-79: Low I** The project is good or excellent on each of the basic and additional criteria.
- 80-89: Mid I** The project is good or excellent on each of the basic and additional criteria and also has elements of the exceptional criteria.
- 90-100: High I** The project is good or excellent on each of the basic, additional and exceptional criteria.

Some projects will not fit neatly into any category, eg strong on additional criteria, but weak on a basic one. In this case you are asked to trade one criterion off against another as best you can, bearing in mind that failure on a basic criterion is a serious fault.

Markers should assess the project on the basis of evidence provided both in the written dissertation and in oral presentation. If you are aware of any mitigating factors which should be taken into account, please do not compensate for them in your assessment, but mention them in the appropriate section in your report and indicate the degree of compensation you feel would be appropriate. If you feel that the dissertation does not do justice to the work carried out by the student, please make this clear in your report together with an explanation. Supervisors should also note the extent to which the student was self-directed or else required close supervision, and confirm their satisfaction (under the heading “explanation of the project” with the extent to which the student is able to demonstrate detailed understanding of the software.

CIS320: PROJECT ASSESSMENT REPORT

Computer Science, Computing & Information Systems, Internet Computing

Student Name: _____

Project Title: _____

Marker: _____

Mark Sheet completed by: Supervisor Second Marker

Number of Supervisions Attended (Supervisor sheet only):

Mark: _____ %

INSTRUCTIONS

Read the guidance notes and grade the dissertation under each of the following criteria. Comments overleaf.

BASIC CRITERIA	<i>inadequate</i>	<i>adequate</i>	<i>average</i>	<i>good</i>	<i>excellent</i>	N/A
	1	2	3	4		
Deadlines Met						
Explanation of Overall Project	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Explanation of the Software	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quality of the Software	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Completion of the Project	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quality of the Dissertation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ADVANCED CRITERIA	<i>inadequate</i>	<i>adequate</i>	<i>average</i>	<i>good</i>	<i>excellent</i>	N/A
Discussion of Area and Background	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Critical Evaluation of own work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Justification of Design Decisions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
EXCEPTIONAL CRITERIA	<i>inadequate</i>	<i>adequate</i>	<i>average</i>	<i>good</i>	<i>excellent</i>	N/A
Evidence of Ingenuity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Critical Evaluation of Previous Work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>