Marking and Grading

Dr Dengsheng Zhang¹

1. Norm-referencing versus Criterion-referencing Assessment

Norm and criterion referenced assessment are two distinctly different methods of awarding grades that express quite different values about teaching, learning and student achievement [1].

Norm referenced assessment, or ‘grading on the curve’ as it is commonly known, places groups of students into predetermined bands of achievements. Students compete for limited numbers of grades within these bands which range between fail and excellence.

The essential characteristic of norm-referencing is that students are awarded their grades on the basis of their ranking within a particular cohort. Norm-referencing involves fitting a ranked list of students’ ‘raw scores’ to a pre-determined distribution for awarding grades. Usually, grades are spread to fit a ‘bell curve’ (a ‘normal distribution’ in statistical terminology), either by qualitative, informal rough-reckoning or by statistical techniques of varying complexity. For large student cohorts, statistical moderation processes are used to adjust or standardise student scores to fit a normal distribution. This adjustment is necessary when comparability of scores across different subjects is required.

In contrast, criterion-referencing, as the name implies, involves determining a student’s grade by comparing his or her achievements with clearly stated criteria for learning outcomes and clearly stated standards for particular levels of performance. Unlike norm-referencing, there is no pre-determined grade distribution to be generated and a student’s grades is in no way influenced by the performance of others. Theoretically, all students within a particular cohort could receive very high (or very low) grades depending solely on the levels of individuals’ performances against the established criteria and standards. The goal of criterion-referencing is to report student achievement against objective reference points that are independent of the cohort being assessed. Criterion-referencing can lead to simple pass-fail grading schema, such as in determining fitness-to-practice in professional fields. Criterion-referencing can also lead to reporting student achievement or progress on a series of key criteria rather than as a single grade or percentage.

Which of these methods is preferable? Norm-referencing assessment emphasizes comparability and objectivity, it is an efficient marking regime which is particularly suitable for assessing large cohort of students in modern universities. Criterion-referencing assessment emphasizes on quality and flexibility, while it is good for students, it is difficult to implement due to it’s complexity and time consuming. As the result, most students’ grades in universities are determined on a mix of both methods, even though there may not be an explicit policy to do so. In fact, the two

¹ This work is done under the Graduate Certificate in Higher Education Course
methods are somewhat interdependent, more so than the brief explanations above might suggest.

Logically, norm-referencing must rely on some initial criterion-referencing, since students’ ‘raw’ scores must presumably be determined in the first instance by assessors who have some objective criteria in mind. Criterion-referencing, on the other hand, appears more educationally defensible. But criterion-referencing may be very difficult, if not impossible, to implement in a pure form in many disciplines. It is not always possible to be entirely objective and to comprehensively articulate criteria for learning outcomes: some subjectivity in setting and interpreting levels of achievement is inevitable in higher education. This being the case, sometimes the best we can hope for is to compare individuals’ achievements relative to their peers.

Norm-referencing, on its own — and if strictly and narrowly implemented — is undoubtedly unfair. With norm-referencing, a student’s grade depends – to some extent at least – not only on his or her level of achievement, but also on the achievement of other students. This might lead to obvious inequities if applied without thought to any other considerations. For example, a student who fails in one year may well have passed in other years! The potential for unfairness of this kind is most likely in smaller student cohorts, where norm-referencing may force a spread of grades and exaggerate differences in achievement. Alternatively, norm-referencing might artificially compress the range of difference that actually exists [2-5].

In my teaching of computer models for business decisions, I have combined both norm-referencing and criterion-referencing methods in assignment assessment. 70% of the assessment is norm-referencing marked and 30% of it is criterion-referencing marked. The criteria of marking the project assignment are listed in the following.

**Problem proposal and specifications: (9 marks)**

1. The case must be formulated from the company they have worked, attach email from employer for evidence. (1 marks)
2. The problem specification must consist of the required components and have a reasonable length (8 marks)
   2.1. Background or introduction. This is the part where the type of operation is occurring in the company or department, what problem or issue they are facing and what decisions or solutions they are seeking. This part also gives clues to applying any business model. (1 mark)
   2.2. Specifications of the problem. This part specifies the problem in details. It provides information and parameter values which are necessary for applying a specific business model. (4 marks)
   2.3. Data. This part provides or summarizes the data needed for the application of the model. In some cases, the data may have been given in the above specification process, but normally, summarize the data in tabular or list form. (1 mark)
   2.4. Assignment questions. This part presents the questions to be answered in the proposed problem. The assigned questions should be meaningful and helpful to the decision maker. There should be at least three questions to be answered. (1 mark)
   2.5. The proposal should be reasonable length, normally one and half A4 pages with 12-point font size. (1 mark)
Solutions: (21 marks)
3. Solutions must be within 10 A4 pages (2 marks)
4. Solutions must be readable and logically clear (2 marks)
5. The solution must answer at least 3 non-repetitive questions. The 3 questions does not include the computer solution (6 marks)
6. There must be a model formulation process to the solutions, that is, the mathematical formulas or process of the model must be presented (3 marks)
7. There must be a computer solution with clear Excel formulation procedures (3 marks)
8. There must be a sensitivity analysis to analyze possible alternatives or resource constraints, sensitivity report must be presented (3 marks)
9. There must be interpretations of the results and make the appropriate decisions based on the results (2 marks)

The six levels of Bloom’s taxonomy are roughly grouped into three pairs and match against three passing grades, thus: knowledge and comprehension for pass grade (1~5); application and analysis for credit or distinction grade (5~7), and; synthesis and evaluation for high distinction (8~9).

The following is the comparison of the marking results on this project by the marker and myself.

Table 1. The result of 6 assignments marked by the marker

<table>
<thead>
<tr>
<th>Assessment items</th>
<th>Assignment papers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Specifications</td>
<td>7.5</td>
</tr>
<tr>
<td>Solutions</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>26.5</td>
</tr>
</tbody>
</table>

Table 2. The result of 6 assignments marked by myself

<table>
<thead>
<tr>
<th>Assessment items</th>
<th>Assignment papers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Specifications</td>
<td>8</td>
</tr>
<tr>
<td>Solutions</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
</tr>
</tbody>
</table>

The two results have a significant difference, on average, the difference is over 8 (26%) marks for each paper, and notably there is no significant difference between the 6 total marks given by the marker. While it is clear from examining the 6 papers that the marker is not following the last three criteria, the main reason is that the marker is not doing a good job, because she is new to this subject and has no background on it. In three cases, the proposals are too simple and the required three assignment questions are essentially a single one, I comment that the number of customers and resource costs being considered to make more sensible decisions. But the marker has ignored them. She only follows the rules literally because there is no criterion indicates that the proposal has to be to certain standard and the questions have to be
significant. These need to be pointed out in future. In two cases, there are a few errors in the specifications which cause confusion, but the marker fails to find them. In three cases, there is no sensible interpretation of the solutions and decision making, but no marks are deducted by the marker. This project presents a challenge to both the students and the marker. To mark this project, the marker has to have a significant understanding of all the business models taught in this subject and spend significant amount of time in judging the solutions. It especially needs significant judgement to determine whether the proposal is a good one or a poor one. Due to the marker’s poor judgement in marking this project, the lecturer has to give her marked samples and require her to recheck this project for all the marked assignments. The criteria for this type of questions can only be verified by tutor who has strong background in the topics.

2. Group Assessment

Group work in university study has become more and more important due to the demanding requirement of collaborative work in corporations. Many faculties like faculty of IT have required that group work be part of the assessment of students’ learning skills. In the last FIT survey, students were asked about their team work experience in their learning, the specific question reads: “The course helps me to develop my ability to work as a team member”. GSCIT scored the poorest in this question. The school reacted by calling on an academic staff meeting to address this very particular issue. During the meeting, staff exchanged their experience on supervising group work and there was a hot discussion. The two very issues raised in Webb’s article [6] were heavily argued: group assessment emphasizing on learning process versus group assessment emphasizing on learning outcome. The few staff from industry stress very much on learning outcome, while other staff favor more on learning process.

Many staff raised the difficulty of coordinating/monitoring group working and allocating scores to members. Some staff have been using web log to record students’ involvement and to allocate scores to members accordingly. Some staff suggested incorporating project assignments in certain units in addition to the third year project to encourage more group work in students’ learning. The current criteria for assessing third year project have also been reviewed. As the result of the review, the following decisions have been made:

- A 20% marks are set up for a new criterion: Work as a team
- A 10% Presentation is separated out from the final report as a new criterion
- The previous 40% intermediate reports is given less weight to 30%
- The final product/system is weighted down to 40% from 60%
- Group work will be brought into other units within the courses
- The project coordinators are to find out how to measure the 20% team work

It is a clear decision to shift from emphasizing on learning outcome to more emphasizing on learning process.
Reference: