A comparison of norm-referencing and criterion-referencing methods for determining student grades in higher education

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 (such as in senior secondary education), 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 (such as when subject scores are added to create an aggregate ENTER score for making university selection decisions).

Norm-referencing is based on the assumption that a roughly similar range of human performance can be expected for any student group. There is a strong culture of norm-referencing in higher education. It is evident in many commonplace practices, such as the expectation that the mean of a cohort’s results should be a fixed percentage year-in year-out (often this occurs when comparability across subjects is needed for the award of prizes, for instance), or the policy of awarding first class honours sparingly to a set number of students, and so on.

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? Mostly, students’ grades in universities are decided on a mix of both methods, even though there may not be an explicit policy to do so. In fact, the two 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.

Excerpt from James, R., McInnis, C. and Devlin, M. (2002) Assessing Learning in Australian Universities. This section was prepared by Richard James.
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.

Criterion-referencing is worth aspiring towards. Criterion-referencing requires giving thought to expected learning outcomes, it is transparent for students, and the grades derived should be defensible in reasonably objective terms – students should be able to trace their grades to the specifics of their performance on set tasks. Criterion-referencing lays an important framework for student engagement with the learning process and its outcomes.

Recognising, however, that some degree of subjectivity is inevitable in higher education, it is also worthwhile to monitor grade distributions – in other words, to use a modest process of norm-referencing to watch the outcomes of a predominantly criterion-referenced grading model. In doing so, if it is believed too many students are receiving low grades, or too many students are receiving high grades, or the distribution is in some way oddly spread, then this might suggest something is amiss and the assessment process needs looking at. There may be, for instance, a problem with the overall degree of difficulty of the assessment tasks (for example, not enough challenging examination questions, or too few, or assignment tasks that fail to discriminate between students with differing levels of knowledge and skills). There might also be inconsistencies in the way different assessors are judging student work.

Best practice in grading in higher education involves striking a balance between criterion-referencing and norm-referencing. This balance should be strongly oriented towards criterion-referencing as the primary and dominant principle.

In summary:

1) begin with clear statements of expected learning outcomes and levels of achievement;

2) communicate these statements to students (they should be written so they make sense to students);

3) measure student achievement as objectively as possible against these statements, and compute results and grades transparently on this basis; and

4) keep an eye on the spread of grades or scores that are emerging to be alert to anything amiss in assessment tasks and assessor interpretations.