

On some Limitations of Institutional Rationality

Ed Brandon

When you spend a fair time in an organisation I guess you are likely to notice that it often fails to practise what it preaches. One could call this a failure of rationality; one could call it other things. But when the organisation is a university, dedicated to expanding knowledge and understanding, or ‘unlocking the potential of the region’ as our own institution envisions, then it is perhaps fair to think that irrationality is at least one of the problems with this disjunction.

The general area I am concerned with is that of ‘maintaining standards’. I hope I may be excused documenting the prevalence of this broad issue in what is said by and about universities (and, of course, other educational institutions¹). I will focus on a couple of issues in this broad area: our assessment of students, and the conditions we place upon their entry.

One and a half issues concern the way we examine and use the results. [Martin Schade’s paper](#) made the point that what a lot of us do cannot be seen as a contribution to a student’s learning about their strengths or weaknesses – it is purely summative and in no way formative. But while that is an important point, my concern is rather different. It is something I picked up from my time in the field of Education. In general, we fail to standardise scores on tests and examinations in arriving at final results. The educational measurement literature assures us that it makes very little sense to average raw scores on different tests, yet this is precisely how class of degree is calculated.

To take a very elementary example from a reputable educational authority, the National Foundation for Educational Research (NFER) in the UK tells us that standardised scores are preferred in responsible testing for three reasons, the third of which is “so that scores from more than one test can be meaningfully compared or added together”²:

Standardised scores from most educational tests cover the same range from 70 to 140. Hence a pupil's standing in, say, mathematics and English can be compared directly using standardised scores. Similarly, should a teacher wish to add together scores from more than one test, for example in order to obtain a simple overall measure of attainment, they can be meaningfully combined if standardised scores are used, whereas it is not

¹ I concentrate on universities simply because I know them from the inside much better than schools. I think some, at least, of the issues are common to both.

² The other two are “in order to place test takers' scores on a readily understandable scale” and “so that an allowance can be made for the different ages of the pupils” – aims one might think pertinent even at tertiary level.

meaningful to add together raw scores from tests of different length or difficulty.

In occupational tests, the use of standardised scores enables the organisation to compare directly or add together sub-test scores or scores from different tests in a battery.³

In my experience, universities “add together ... scores from different” examinations without a care in the world. Transformations via percentages, or a GPA, do nothing to achieve the comparability that standardising yields.⁴

I suspect that the suggestion that standardisation of scores is desirable is liable to push us into what now seems to me something of a hornet’s nest: the contrast between criterion-referenced and norm-referenced tests. I am inclined to think that we like to believe our testing is criterion-referenced when it is not, or is a dubious amalgam. One extreme position, as Weil and Kroontje have it, is that a test of whether someone has reached a criterion should yield just two possible answers – “Criterion-based evaluation is best suited to mastery learning, but even there the system should only be bivariate. One either masters the material or one does not” (1977, 32). Indeed, it is pretty difficult to say what a 55% level mastery of epistemology might be. What I think actually happens, is that in supposing that some score or percentage represents the standard appropriate to some level of mastery, what we do is to imagine that it derives from a fictitious population of would-be testees – as Rowntree (1987) says, a kind of counterfactual norm-reference. I have to admit, however, that there is more work to be done on connecting the educational measurement literature with the procedures many of us continue to use.

I said one and a half issues since locally people use “standardise” in a different sense, that is, as a matter of ensuring consistency in the behaviour of different markers. The Caribbean Examinations Council, for instance, creates elaborate mark schemes, and gets senior examiners to work with a sample of scripts to make sure that everyone involved in marking a particular examination knows how to treat with the range of answers that will be given, and how to allocate marks consistently. When in a university one person does the marking for a course there might seem less scope for this sort of standardising, though (i) there is the question of a mark scheme, which we are now being asked to provide, and (ii) there is still a question of consistency over time. And there are a number of instances (where courses are taught at other institutions, for instance) where more than one person is in fact mainly responsible for marking a selection of scripts.

A similar failure to take assessment seriously can be approached via the question of matriculation and open entry. A very common argument when these questions are posed

³ From <http://www.nfer.ac.uk/research-areas/assessment/standardised-scores-and-percentile-ranks.cfm> (accessed May 23, 2008).

⁴ “A standard score is a dimensionless quantity derived by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation” (http://en.wikipedia.org/wiki/Standard_score, accessed May 23, 2008). As the NFER says, this procedure entails that the resulting score is “related to (a) the average score of all the test-takers, or (b) how spread out their scores are”.

is that it is only possible to maintain academic standards by requiring incoming students to have already acquired certain qualifications. This is obviously fallacious: the way to maintain standards is to be rigorous in testing for them, it matters not what a person has already done. Reliance on matriculation requirements thus seems to be an admission, via standard Gricean mechanism, that the university in fact does not rigorously examine its students.

There is of course a big issue of efficiency with respect to matriculation. It is largely a waste of time, effort, and other resources to enrol someone in a class in which they cannot succeed for lack of prerequisite knowledge or skill. But that doesn't seem to exhaust people's belief that students must have prior qualifications.

It would be helpful if one could explain these failures to act appropriately with respect to standards more illuminatingly than as mere irrationality. As [Steven Burwood](#) argued, traditional university education was a matter of the reproduction of a social élite, with a partiality for the reproduction of the academics themselves. When you were reproducing a social élite, the matriculation issue gave you one easy way of blocking entry to undesirables, while appealing only to academic and not patently social issues. Once inside, it didn't matter too much what a person did, so pretty rough and ready, sloppy means of measuring performance were acceptable. It seems to me that part of the explanation for the continuing sloppiness about examining and assessing students is a hang-over from those days. But as the numbers in tertiary education expand, there are pressures to require more accurate measurement.

References

Rowntree, D., 1987. *Assessing Students: How Shall We Know Them?* London: Taylor and Francis.

Weil, R.R. and W. Kroontje, 1977. [Grade inflation: causes and cures](#). *Journal of Agronomic Education* 6: 29-34.