Introduction

The paper introduces Rust’s ‘bete noir’ – “the ubiquitous use of numbers in the assessment of students’ work........ It does not seem to matter where you look around the world, academics seem to be entrenched in systems that (mis)use numbers to judge and record students’ assessment.”

Seven arguments are raised against traditional practices in the use of numbers as the basis of assessment judgements and reporting.

(1) **We Need to Stop Using Percentages Because They Do Not Tell Us Anything Useful**

This is because two students can be awarded the same score (say 55%) for significantly different performance on the same assessment task - in which case the number obscures the difference between the two. In addition, the fact that a marker chose not to give the work a score of 54% would appear to suggest an ability to distinguish the quality of the work to a precision of one percentage point - a distinction of one hundredth.

(2) **We Need to Scrap Percentages Because They Are Not Scalable**

This is because a piece of work that scores 90% is not exactly one and a half times better than a piece of work that scores 60%. In which case the scale is not truly quantitative, so any standard arithmetic operations that treat those grades as if they are is illegitimate (Dalziel, 1998).

(3) **We Must Stop Combining Scores As If They Were Just Numbers**

This is because numbers are symbolic representations of a range of different judgements rather than just numbers that can be added, combined and moved about. For example, adding the mark for the report of a laboratory practical to the mark from an exam obscures the fact that they represent different types of learning outcomes; treating them simply as numbers obscures these underlying differences. Also, simplistically adding and aggregating numbers while
ignoring differences in range and mean deviation makes the resulting outcomes statistically unsound.

(4) **We Must Not Force Our Assessment to Fit a ‘Normal’ Distribution Curve**
This is because (as Bloom pointed out) “The normal curve is a distribution most appropriate to chance and random activity. Education is a purposeful activity and we seek to have students learn what we would teach. Therefore, if we are effective, the distribution of grades will be anything but a normal curve. In fact, a normal curve is evidence of our failure to teach” (Bloom et al, 1971).

In addition, norm referenced assessment fosters competition rather than collaboration among students.

(5) **We Must Stop Assessment Judgements From Being Distorted by Erroneous Other Factors**
This is because assessment judgements should reflect learning achievements or outcomes rather than extraneous factors such as attendance or late penalties (see Assessment Brief 17 for Sadler’s detailed discussion of this issue).

(6) **We Must Stop Assessment Judgements From Being Distorted by Assessment Task, Subject Discipline and Institutional Rules**
This is because UK research has shown that “students are more likely to score highly on coursework (essays, reports, presentations, etc.) than in examinations, and also in more numerate disciplines, such as mathematics and engineering, compared with disciplines such as history or sociology”. For example, graduating students of mathematics are four times more likely to get a first (the top grade) compared to students in history – partly because of the systemic unfairness in the ways that marks are used to derive the degree classification. You can score 100% on a math task; in the UK, good work in history would be lucky to score 75%. But there will be no concession to this difference in the way they are treated by the university’s system; arithmetically, each of these results will treated in exactly the same way.

(7) **We Must Stop Assessment Practices Having a Negative Effect on Learning**
This is because a focus on marks can mean that “…students become more interested in the mark and less interested in the subject over the course of their studies” (Newstead, 2002).

Feedback (provided without marks) has been found more likely to result in improved future work than feedback provided with marks attached.

**SoTLA (the Scholarship of Teaching and Learning AND Assessment)**
Rust concludes that the persistence of some unsound use of numbers in university assessment practice despite counter-arguments supported by a weight of research evidence constitutes a strong case for the Scholarship of Teaching and Learning movement, to extend its agenda, to include the Scholarship of Assessment. “SoTL needs to become SoTLA as having assessment in the name would act as a constant reminder of the central importance of assessment in the teaching and learning process, and the need to develop the scholarship of assessment within the academy. And only when we have a critical mass of faculty who understand the scholarship of assessment do I think we have any chance of bringing about significant change and a stop to our unscholarly practices.”

---

This series of briefs on assessment topics has been prepared by the UQ Teaching and Educational Development Institute (TEDI) for UQ teaching academics. “UQ ASSESSMENT BRIEFS” of journal articles, book chapters, reviews, websites, reports etc are distributed to Faculty and School Teaching and Learning Chairs in a form designed to encourage wider distribution.
Please send feedback, requests or suggestions for future “UQ ASSESSMENT BRIEFS” to:

Dr Clair Hughes (clair.hughes@uq.edu.au)
Senior Lecturer in Higher Education
Teaching and Educational Development Institute (TEDI)
The University of Queensland, St Lucia Qld 4072
Telephone +61 7 336 52456
Facsimile +61 7 3365 1966