The latest TEF Assessment Framework, automated analysis of data, and some Scottish anxiety *by Vicky Gunn*

I had the good fortune to be in Rio for the Paralympics this September. My step- daughter was competing in an endurance road race. For her, the most important thing was improving on previous race times, but she'd hoped to get a medal as well (even though this was not predicted by her 'metrics'). At the end of September, the Westminster Government through HEFCE, published the TEF2 Technical Specification[i] and I found, to my astonishment that the original differentiating

phrases (meets expectations, excellent, outstanding) were to be replaced with medals: Bronze, Silver, Gold. All of this got me thinking about the Teaching Excellence Framework, built like British Cycling on the idea that we can differentiate excellence for competitive purposes and this is a good in itself. I find this comparison deeply troubling. I, like many involved in quality and teaching development in Scottish Higher Education, have invested several years of my professional life to fostering cultures of enhancement. Indeed, the distance travelled to improvement in teaching provision has been a mantra within the totality of Scottish higher education's stakeholders (academic, government, student bodies alike). In the Quality Enhancement Framework (QEF), we have been more interested in seeing all Scottish institutions getting 'personal bests' (hence demonstrating continuous improvement from within their own context), rather than doing better than all the others (final outcome measure).

However, now that we have the first set of TEF indicative metrics, I (like a cycling coach) am assailed with a few doubts about the laudable concentration of raising the quality of the whole Scottish sector. This is an aim of the QEF. This resulted in engaged participants of a quality system which steadfastly refused the divisiveness associated with differentiated institutional quality review outcomes. Yet, if we individually enter it, the TEF will now demand this of us. Should Scotland then change its QEF substantially with its aspirational collectivism to be consigned to being a phantasm of a previous era? How long can such a discourse last in the face of *going for gold*? Should I, as an institutional Head of Learning and Teaching, now focus on competing with HEIs, so my institution is seen as outstanding in comparison to all the others and place the sector's aspirational culture in a box marked 'soppy idealism'? To put it in British Cycling's inelegant but superlatively economic phrasing: how will my small specialist institution *medal* when facing larger, wealthier institutions?

From the outset, there is one aspect of the QEF upon which the TEF seems to have shone a challenging light: our avoidance of the general shift to a metrics data 'collection' which can be subjected to automated analysis. The reality is that the TEF has ushered in a paradigm shift for quality enhancement assessors in Scotland and we have to identify how to mediate its perverse influence. We need to find an adequate way to move forward and will have to do that without a research base from which to make decisions. I realised this at an early meeting between what was then BIS and the HEA's Pro Vice- Chancellors Group. Those of us used to making enhancement judgements offered quite different criteria to those familiar with outcomes-based subject review for judging the items outlined in the initial technical consultation as 'criteria' (a phrase which has been maintained for TEF2). Just for clarity, as we pointed out at that meeting, what are stated as criteria were actually what most quality folks (especially anyone who has written programme specifications) would know as assessment descriptors. Criteria would differentiate what it is necessary to demonstrate at each 'medal' level, descriptors are the over-arching objects from which degrees of differentiation are made. Robust criteria provide the reassurance of criterion-referenced

assessment (supposedly more objective) which all the disciplines, even the creative subjects, have been persuaded to take on by quality assurers and academic developers, rather than the norm-referenced systems many of us (outside the STEM subjects) had used previously. But it seems in the new automated model that such differential criteria are unnecessary. This is an important distinction because differential criteria requiring professional qualitative judgements of panel members and assessors can be open to criticism and might push the whole system into expensive appeals. If so, the data-set metrics, rather than any broader documentation accompanying them, are likely to win the day. Qualitative judgements (the locus of academic expertise on many teaching quality and excellence panels) which engage with what we thought might be a generalised provider submission will get lost in litigation paranoia.

However, if one explores the TEF assessors' decision-making process as now identified in the TEF2 technical specification, it is clear that actual qualitative decision-making will not operate in a manner recognisable to those of us who have sat on institutional reviews before. Indeed, the academic agency of the panel will be closely constrained from the outset. Positive and negative flags already resulting from an automated analysis of the metrics actually do the initial judgement-making for an assessor. What produces the flags is not an academic panel, it is actually the criteria of differentiation designed into the technical system.

Automated analysis seems to manage one big anxiety I had: how might unconscious bias function when competition with others rather than institutional 'personal bests' becomes the focus of qualitative judgement? There is clearly a tension between product orientation and development orientation and this will arguably play out in how assessors respond to the submissions they receive. The explicit or prescribed criteria that are agreed by panel members are only part of the process. What of the latent criteria we all bring from years of enhancement-led experience (enhancement latent criteria are qualitatively different to outcome criteria – for a start, enhancement criteria normally include judgements around the quality of potential actionable insights that an institution is deriving from its evaluation data. If I am used to looking for these to trade off against poor NSS scores, how will I function as a TEF assessor?) As we know from Sadler's work on qualitative judgement in assessment, explored in more depth for professional qualitative judgement by Wyatt-Smitt and Klenowski[ii], a judgement is made when two types of criteria (explicit and latent) meet the rules for use or non-use of those criteria (in what he refers to as trade offs). The flag system of the TEF seems to sort this, especially in the early stages of decision-making.

Of course, phrases such as *hypotheses* and *testing* sound scientifically robust, until you realise that assessors won't be testing an inductive hypothesis they have drawn from the data. Rather, they will be trying to manage the robustness of already identified 'flags' (which have emerged from automated benchmarking etc) and, where necessary, whether or not a provider's written submission accompanying the metrics adequately offers mitigation for out-of-step results. It will be at this stage that explicit/ latent criteria and trade off rules emerge. I hope HEFCE and/or the Scottish Funding Council invest in a relevant research project to explore this.

More pragmatically, the outcome of the TEF2 Consultation as represented in the TEF2 Technical Specification demonstrates that the Scottish HEIs' relatively uniform response to the items in the TEF has led to substantial concessions in terms of enabling the Scottish sector to take part (without overlaying English metrics systems entirely on how we report things such as multiple deprivation, progression and retention). All of this still leaves big

questions around the production of the metrics through algorithmic design and of their subsequent role in qualitative judgement by assessors. My current concerns are as follows:

- • How do translational algorithms work to enable the comparison of apples (the English three- year specialist undergraduate degree sector) and pears (the Scottish predominantly four-year, major-minor undergraduate degree sector), which the TEF, as now outlined, is actually attempting to achieve? Interestingly, translational algorithms operate in disability sports, as the diversity of extent of disability has to be factored into actual race times. This means that a very disabled cyclist with a slower actual time than other competitors can still win.
- • How can Scottish panel members ensure that the spoken culture of Scotland's sector (as noted in an over-arching statement of principles statement established by Vice-Principals Learning and Teaching in Scottish institutions when the TEF consultation first appeared) could be overlaid onto TEF assessment flags? What research could we do to explore how cultural perceptions, even slight ones, play out in both the design of the automated systems of analysis and the hypothesis testing to occur on panels?

Finally, and perhaps most substantially, are universities really so like athletes? Is that what we are reduced to in the mind of our governments? If so, we can anticipate metaphoric *doping* scandals, petty squabbles about algorithms, corruption (*fixing)* claims, coaches whose sole intention is to get their university to win whatever the work-life balance impact. Oh, no, wait a minute. Business as usual then?

[i] TEF Year 2 Technical Specification:
https://www.gov.uk/government/publications/teaching- excellence-framework-year-2-specification
[ii] Sadler, R (1985) The origins and functions of evaluative criteria *Educational Theory*, 35(3):285-97; Wyatt-Smith, C and Klenowski, V (2013) Explicit, latent and meta-criteria: types of criteria at play in professional judgement practice. *Assessment in Education: Principles, Policy & Practice*, 20(1): 35-52.

SRHE member Professor Vicky Gunn is Head of Learning and Teaching at the Glasgow School of Art. Follow her on Twitter @StacyGray45