Assessing the quality of higher education

Written submission to the BIS Select Committee, from the London Mathematical Society

This submission relates to mathematics teaching in higher education.

1. Summary.

- The London Mathematical Society welcomes government recognition of the importance of teaching in higher education. However we have serious concerns about the procedures currently being adopted to give effect to this recognition.
- HEFCE's longstanding efforts on "quality assessment" do not seem to be coordinated with BIS's proposed "Teaching Excellence Framework" (TEF).
- We recognise that quality needs to be suitably monitored in an increasingly varied HE sector. However, we see considerable dangers in the possible hasty adoption of an inappropriate, tightly specified TEF, which is then used to assess the quality of teaching and to allocate 'rewards'.
- We explain below, why the kind of Framework that could deliver genuine improvement is one which avoids inappropriate 'metrics' (such as short-term measurable 'outputs', or student 'satisfaction', or degree classes), which improves the interface between teaching and administration, and which impinges on individual academics mainly by offering criticism, support, and opportunities for thoughtful improvement – allowing a varied, steadily improving provision of mathematics teaching in higher education.
- Any effective TEF will need to be developed in concert with the academic community, rather than being externally imposed. Hence the design of such a framework must have serious input from those with significant experience of teaching Mathematics in higher education. (The London Mathematical Society would like to offer its expertise here.) It must also leave scope for individual academics to exercise their academic judgement.
- The successful development of such a framework will require an extended timeframe.

Outline

2. Preamble The Conservative Party 2015 election manifesto included a commitment to:

> "introduce a framework to recognise universities offering the highest teaching quality".

It also declared boldly that: "We will lead the world in maths and science." We therefore feel justified in focusing this submission on the issues raised from the specific perspective of the *mathematical sciences*.

The manifesto left unexplained how "teaching quality" was to be assessed. It was equally silent on the kind of "framework" proposed, how this might help one decide which universities deserve to be "recognised", and what form this "recognition" might take.

Some might argue that such details can always be sorted out later, and that HEFCE had been working on this for many months. However, there seems to be an unhelpful divide between HEFCE's work on 'quality assessment' and BIS's totally separate work on a TEF. The delay in publishing the BIS Green Paper on the proposed TEF (till November/December?) makes it even more difficult to fill in the apparent gaps.

We suggest below that any approach to 'teaching excellence' should start by taking into account our collective experience of assessing "teaching quality" in schools. That experience highlights the need for any TEF to simplify HEFCE's focus on "outcomes" and on "the student academic experience", by

(a) making the crucial distinction between *short-term* and *long-term* outcomes, and
(b) recognising the fact that, in many respects, the value of a student's academic experience only becomes apparent (even to the student) retrospectively – often long after graduation.

The framework which has been used for assessing teacher quality in schools has proved problematic, and counter-productive (e.g. in spawning an all-pervading culture of "levels" and "sublevels" for assessing student progress, to which the system is now addicted – despite Ministerial edicts seeking to outlaw their use).

Assessing quality in higher education presents significantly greater difficulties than at school level (e.g. finding suitable metrics, identifying qualified assessors, etc.).

The recent HEFCE discussion document *The future of quality assessment in higher education* pre-echoed the second of the quoted commitments from the Conservative manifesto when it declared that:

> "The UK higher education system […] has a worldwide reputation for excellence.
> We want to maintain and enhance that reputation."

This sounds admirable. However, the published summary of responses to the HEFCE discussion document fails to recognise how easily this reputation could be damaged. The respondent "stakeholders" were largely administrators (for whom the question was *how* to implement procedures in a congenial way), rather than academics (who might be able to explain some of the dangers).

Hence, our reluctant conclusion is that the matter needs to be subjected to much more careful scrutiny and analysis *by practitioners* before we adopt some well-meaning, but ill-conceived TEF – which may appear to deliver on the less important manifesto promise, but which may fail the more important challenge of preserving and enhancing the reputation of "maths and science" and other HE disciplines.

3. Some remarks on assessing teaching quality in schools

Effective mathematics teaching comes in many different forms. We indicate some of the known consequences of trying to devise a uniform framework to assess the quality of teaching.

3.1. At school level, attempts to assess teaching quality have focused on crude metrics (KS2 SAT results, GCSE results, Ofsted measures of progress, etc.). This unfortunate focus has systematically distorted the educational process – forcing teachers and managers to concentrate on ways of improving *what is measured*, rather than on what pupils need to learn *for their subsequent*

progression. In short, the quality assessment process in schools has obliged everyone to demonstrate "success" *in the short term*, while penalising those who operate for the long-term benefit of their pupils.

The result has been that officials, teachers and examiners have all worked together to deliver what was demanded – namely "improved outcomes" as measured by national assessments. Their concerted efforts have indeed led to 'improvements' – sometimes dramatically so. For example:

- the percentage of pupils achieving "Level 4" at KS2 in mathematics increased from 45% in 1995 to 86% in 2014, while the percentage achieving "Level 5" or above more than tripled (from 13% in 1995 to 42% in 2014);
- the numbers achieving grades A/A* in GCSE mathematics more than doubled between 2000 and 2015 (with the "improvement" at A level being even more dramatic).

These remarkable outcomes reflect a huge investment of effort and money – by pupils, by teachers, and by HM Treasury. And they may at first sight appear to validate the underlying policy (and its associated accountability framework of league tables, quality assessment, etc.).

Until, that is, one notes the contrast between these *internal* measures and **international assessments** – such as performance in TIMSS (Year 9), or in PISA – which **indicate stagnation.**

3.2 The evidence alluded to in 3.1 reveals the fundamental tension between *short-term* and *long-term* outcomes, or improvement.

Crude metrics encourage short-term changes – in order to optimise scores on the immediately impending assessment. In mathematics, this is most easily achieved by teaching *rules* (rather than meaningful methods), and by encouraging the use of primitive, backward-looking techniques with which students are already familiar (rather than encouraging them to use new approaches, that may at first seem more complicated, but whose mastery is essential for subsequent progress).

For example, the *Numeracy strategy* (1999-2006) was not introduced to prepare primary pupils more effectively for secondary school mathematics: it was officially tasked to 'raise KS2 test results'. And it did precisely what was asked: national test scores rose significantly at the end of KS2. Moreover, the average score of England's Year 5 pupils on the international assessment TIMSS (2003) jumped dramatically – with this improvement being sustained (and further enhanced) in 2007 and 2011.

This (roughly 10%) improvement *in average scores* was greater than that for any other country, and might seem to justify the massive investment in time, effort and money. Yet there was a flaw. The improved average scores seem to have arisen from greater emphasis on short-term outcomes, and so concealed our failure to *teach for progression*: in particular, the improvement **had no long-term impact in secondary schools at KS3**: when the 2003 and 2007 Year 5 cohorts were tested 4 years later in Year 9 in TIMSS 2007 and 2011, their performance was no better than that of the 1999 and 2003 Year 9 cohorts.

Successful mathematics teaching is a long game. And it is all too easy to institute assessment regimes that undermine teachers' efforts to "invest in the future" – by forcing them to focus on short-term outcomes from their students.

4. Assessing undergraduate teaching. Current attempts to assess teaching in higher education are in some ways worse than those at school level.

4.1. School 'league tables' based on KS2 texts or on GCSE results have distorted what is taught – and have thereby undermined teachers' freedom to focus on their students' long-term progress. But they are at least based on external, "objective", metrics, which cannot be directly manipulated.

In contrast, undergraduate examinations are internally set; standards are externally monitored, but are supposed to reflect the varying nature of a desirably diverse set of institutions rather than be uniform. So it should be clear that the results, or outcomes (whether marks or degree classifications), cannot be used as a formal measure of quality. Yet many university rankings seek to "assess quality" by including the percentage of 1st and 2.1 degrees awarded – with the result that there are now (according to a recent media report) four times as many 1st class degrees as there were in 2000. This distortion would seem to be a direct consequence of ill-conceived metrics for "quality assessment", and should indicate dangers of a framework that offers rewards on the basis of outcomes that can be artificially 'tweaked'.

4.2. Exam marks vary from institution to institution; but they are at least based on something tangible: exam papers and scripts can in principle be scrutinised. In contrast, the "student satisfaction" measure from the *National Student Survey* seems to be entirely subjective – and perhaps meaningless. But because it "matters", institutions go to considerable lengths to 'influence' such metrics to their own perceived advantage. (Of course at all stages of a degree course feed back from students is important and useful for reflective improvement of teaching. Its uncritical use as a measure of teaching quality is the issue here.)

4.3. The effect has been

(a) to reduce a serious enterprise (whose reputation HEFCE claims it wishes to "maintain and enhance") to little more than a 'beauty parade', that encourages everyone to concentrate on superficial features,

and at the same time

(b) to undermine the idea of quality based on professionalism and trust.

It has also given rise to a burgeoning bureaucracy – an army of administrators – dedicated to the optimisation of meaningless metrics.

4.4. Such examples should lead BIS and HEFCE to seriously consider the advantages of a TEF which is designed to support and improve the quality of teaching, *without the bureaucratic hubris of pretending to be able to measure outcomes* (except in those cases where there is objective data that can be easily collected without distorting the educational process).

The truth would seem to be that we simply do not have suitable metrics for "assessing the quality of teaching in higher education", and would therefore be wise to analyse matters much more carefully before imposing a "framework" that overreaches itself, with negative consequences.

5. Conclusions.

5.1. Higher education is bound to challenge students academically. In mathematics, there is no escape from the difficulty of the material, from the care needed when presenting it, or from the

effort needed to master it. All students find this uncongenial at first – especially those from backgrounds where 'teacher accountability' may have led to spoon-feeding at GCSE, or at A level.

Experience consistently shows that the value of a course, or of a teacher, can only be assessed long-term – often looking back from later professional life, after graduation. Hence one should never misconstrue 'the student voice' – especially when expressed in the short term – as a measure of 'teacher quality'.

5.2. Given the evident lack of appropriate metrics, and the subjectivity of the 'consumer voice', the only effective TEF would seem to be one

- which is based on assessing a limited set of objective institutional data,
- which seeks to monitor teaching (with a relatively light touch), and
- which offers long-term support and encouragement for lecturers to refine, adapt, and improve what they do.

5.3. In focusing this support, innovation (including technology) can have a significant role to play. But the key interventions, which provide the greatest pay-off, are often relatively mundane: encouraging staff to reflect (collectively) on the material to be taught; making sure they understand the background of the students to be taught; providing opportunity for individual lecturers to assess and work on their own skills and limitations; to remove administrative constraints which interfere with the practical arrangements for the class; etc. And, in mathematics, the UK has a decent track-record in spontaneously cultivating such ways to improve teaching in higher education. For example:

- the annual series of undergraduate *UK Mathematics Teaching* conferences, which began in Nottingham back in 1976, where leading mathematicians worked together to explore potential improvements – leading to numerous positive developments in the 1980s and 1990s;
- the mathematics lecturers' conference each September, which provided a springboard for new appointees to think about the specific challenges of teaching mathematics in higher education.

5.4. In conclusion, we do not need some grandiose scheme of inappropriate metrics – imposing short-sighted conformity. Rather we advocate a TEF of the kind outlined in 5.2, which impinges on individual academics mainly by offering support, and by encouraging thoughtful improvement, leading to a richly varied, and steadily improving quality of mathematics teaching throughout higher education.