Integrity and quality in universities: accountability, excellence and success

ONORA O'NEILL

Diversity in universities

In the backs of our minds most of us treasure images and ideals of what a university should be. We may to have in mind the universities of medieval, renaissance or Enlightenment Europe; of collegiate Oxbridge; of the Humboldtian University aspiring to *Lernfreiheit und Lehrfreiheit*; of Cardinal Newman's idea of a teaching university; of the American liberal arts colleges; of the great civic universities; of contemporary globally significant research universities with splendid graduate and professional schools, their sights firmly set on innovation and impact beyond their walls.[1]

Universities are now hugely diverse, not only in size, funding and governance, but in other more substantive respects. They teach and do research in different areas, to differing standards, and in differing proportions. They differ in the proportion of their activity that is laboratory based; in the proportion of their students who are residential; in the proportion who are mature (in a bureaucratic sense!); in the proportion who study whatever counts as 'full time';[2] in the proportion who work while studying; indeed in the proportion who work while studying what counts as 'full time'; in the proportion of their budgets devoted to research; in the proportion of their students who complete their courses; in the extent to which they deploy distance learning; in the academic standards attained by their applicants and graduates; and in the subsequent success – or otherwise – of their graduates. All of this is without touching on the murkier worlds of corporate universities[3] and franchised campuses, let alone the flourishing and surprisingly overt market in fake university diplomas and credentials.[4]

There is corresponding diversity in the modes of governance used in universities. Governing boards may be controlled by states or cities, by Churches, by self-perpetuating trustees, by the body of academics, or now even by companies. Funding may be supplied by taxpayers, by student fees, by research contracts, by charitable endowment or alumni giving – or by a mix of these. Diversity and complexity are evident in all directions,

Given the diversity of institutions and of their aims and activities, it is hard to say anything systematic about university governance, and its success or otherwise in securing excellence in universities. So I shall concentrate on university *accountability*[5] and *excellence*, with a brief preliminary explanation for this choice of focus.

Governance, taken in the large sense, comprises the totality of systems by which institutions – for present purposes, universities – organise and control their activities. Accountability organises ways of monitoring the standards to which universities and their component institutions, staff and students carry out the tasks that are assigned to them,

---

[1] For thoughtful reflections on the extent to which we can still take a common view of what universities are or should aspire to be, see Stefan Collini, *What are Universities For?* (London: Penguin Books, 2012).

[2] Apparently a dwindling number of hours per week in US and UK universities. For evidence on the US, see Richard Arum and Josipa Roksa, *Academically Adrift: Limited Learning on College Campuses* (University of Chicago, 2010); and for the UK, see successive HEPI reports, beginning with Bahram Bekhradnia, Carolyn Whitnall and Tom Sastry, *The academic experience of students in English universities* (Higher Education Policy Institute, October 2006) and followed up in subsequent years.

[3] Hamburger University, the McDonalds training institution, has a number of 'campuses', but the joint provision model under the auspices of a university is more common – for example, Harrods has set up a degree course for its staff in conjunction with Anglia Ruskin University. See http://business.anglia.ac.uk/harrods.cfm

[4] See www.diplomaxpress.com. The website is evidently committed to truth in advertising, and offers 'TRUE authentic quality fake diplomas, fake degrees, and fake transcripts' and promises the more energetic – or perhaps the more gullible! – options to 'choose your own grades, GPA, classes, major, and more!'

[5] For more complete discussion of this approach to accountability, see Onora O'Neill, 'Gerechtigkeit, Vertrauen und Zurechenbarkeit', in Otto Neumaier, Clemens Sedmak and Michael Zichy (eds), *Gerechtigkeit: Auf der Suche nach einem Gleichgewicht* (Frankfurt: Ontos Verlag, 2005), 33–55.

and seeks to hold them to account for the standard to which they do so. It combines retrospective and disciplinary aspects of governance, dealing both with recording and incentivising compliance and standards – and with detecting and penalising failures. Some aspects of university accountability are similar to those in other large organisations: there is nothing very distinctive about securing financial accountability in universities.

But other aspects of university accountability are highly distinctive, and of particular interest. How should universities be held to account for the quality and integrity of their teaching and research? A century ago, securing quality and integrity would have been seen mainly as matters for individual academics, for professional bodies (particularly in certain subjects), for academic departments and senates, generally operating under a degree of state or Church oversight. However, liberal ideals of university autonomy and academic freedom were already widely accepted, and this has not changed during the 20th century.[6] Changes in the views of university accountability across the last 30 years do not, I think, signal any general rejection of liberal views of university autonomy. In many university systems, academics still control significant aspects of student access, determine and deliver the syllabus, examine students and award credentials. They also control the conduct of research, admission to research training, and the publication of research reports. All of these activities would be compromised without the constant contribution of individual scholars and scientists.

Yet, since the late 20th century, we have seen huge changes in the forms of accountability to which universities are subject. These changes constitute a rather clumsy attempt to achieve accountability for the greatly increased public revenues supporting teaching and research while maintaining respect for academic freedom and university autonomy.

The new accountabilities: quality control

These different forms of public accountability are based on quite controversial innovations in quality control. It may seem that universities cannot be made accountable for the quality of what they do unless those who hold them to account can determine what they do and produce. If that were the case, public accountability would indeed undermine and corrode academic freedom. The results might be highly damaging.

The currently received view, however, is that it is possible for external bodies to hold universities to account for the quality of their teaching and research without compromising academic freedom and integrity. This is typically done by looking at rather *abstract aspects of university performance* that, it is supposed, can be objectively measured and recorded, while leaving universities and academics a large degree of control of the *content of the syllabus and choice of research topics*. A central characteristic of these approaches to accountability is that they purport simply to measure what universities and academics choose to do. This supposedly leaves universities and academics free to make academic choices, while providing objective evidence of their success – or lack of success.[7]

Some of the abstract characteristics typically measured and recorded in order to secure accountability, without undermining academic freedom, are genuinely quantitative – staff/student ratios, laboratory and library provision, numbers of students, numbers of students completing courses of study, numbers of overseas students recruited. Yet even in these cases it is often hard to be sure that the metrics used give accurate, let alone comparable, measures. For example, it may make a large difference whether a university counts numbers of employed staff or numbers of full-time equivalents, and the calibration of what counts as *full time* is likely to vary in ways that reflect employment law and local needs. Even these genuinely quantitative measures usually create problems, and they ignore many substantive aspects of teaching and research that affect the quality of what is done.

However, other widely used approaches to quality assessment purport to measure, yet they lack genuine units of account, and they ignore much that matters but is not readily counted or measured, let alone compared. For example, some metrics tally the number of students who drop out,[8] or who get less good degrees,[9] or who are in employment a certain time after graduation.[10] All of us know how unreliable and incomplete the evidence for these *ostensibly* numerical measures can be, and the real difficulty of telling what is going well and what less well. For example, is it a good or a bad sign if a university that admits students with adequate but not excellent preparation, then graduates a high proportion of those students? Are they admirably making more good bricks with less straw, or are they short-changing their students and society at large by awarding credentials to students of limited achievement?

The same is true of the many research metrics devised in recent years. Research productivity measured by numbers of publications has risen hugely – but metrics for research quality remain controversial. Increasing productivity has little value unless quality is maintained or improved. Yet many metrics for research quality measure quantity rather than quality. Where research metrics are closely based on rigorous peer-reviewed publication and journal rankings, measures of productivity may have some objectivity, but there are widespread worries that while some metrics are adequate indicators for some sorts of work, they may not offer reliable or valid measures of quality for others.

The complexity of the situation is increased when universities and academics respond rationally to the fact that aspects of their performance are being measured, and to the knowledge that their scores may affect their funding and future, so modify what they do. For example, if rates of

---

[6] There are of course still sporadic demands even in liberal societies that universities provide specific sorts of instruction, or that universities do, or do not, conduct research in certain areas.

[7] Of course there are many complaints that the use of these measures of quality changes distorts or damages what universities and academics do.

[8] Is dropping out just failure to sit exams? Or failure to attend? Or is it formal withdrawal?

[9] Comparisons are particularly hard in this area – particularly if some universities permit students to extend their time of study and others do not.

[10] Employment statistics depend on the quality of alumni and student records, and are seldom up to date.

completion are treated as an important criterion for funding higher education, universities will clearly make efforts to ensure that fewer students fail or flunk: the obvious move is to ensure that more pass their exams. Of course, this can be done creditably by improving teaching and motivating students – but there are other less desirable and cheaper ways of improving scores, for example by lowering pass marks, or making courses and examinations easier. There is sadly quite a lot of empirical evidence that academics and students are tacitly colluding in adopting a less demanding view of study: doing so may suit both parties if students want a credential more than an education and academics want less teaching so that they are free to do more research.[11]

Once aspects of academic performance are deployed for purposes of accountability, behavioural effects such as these are very likely. Indeed, from the point of view of the public funders who hold universities to account, changing behaviour is the aim. Systems of accountability are *meant* to create incentives for those held to account to do better. However those incentives are sometimes perverse: academics and students may be 'gaming the system', seeking to deliver better scores on the performance indicators, even if they cannot produce a better performance.

From metrics to rankings

All of these problems are exacerbated when scores on various metrics are combined to create league tables. This art form is meant to provide a simple view of the *relative* quality of universities, or of university departments, and is now done on a global scale. But *any* way of combining scores on these questionable indicators to create rankings and league tables involves many contestable assumptions in addition to those already made by choosing specific metrics.[12]

None of this daunts those who seek to hold universities to account by constructing rankings. In the last decade, two global ranking systems have emerged: the Shanghai Jiaotung University academic ranking of world universities (ARWU),[13] and the *Times Higher Education* world university rankings.[14] As is well known, these league tables have not ranked many European universities in the top 50 global universities, apart from some in the UK, where there have been demanding quality assessments systems both for research and for teaching for some decades. The EU is now considering developing a more differentiated ranking that will, it is supposed, rank different aspects of universities separately, rather than providing a single composite ranking. Presumably the hope is that the distinctive excellences of many European institutions will then be duly acknowledged.[15] This proposed new ranking goes by the imaginative name *U-Multirank*.[16] Needless to say, *U-Multirank* has received cogent criticism, in particular in a 2010 report of the *League of European Research Universities*.[17]

My own view is that while the proponents of *U-Multirank* evidently hope to devise a metric that acknowledges the diversity of European universities by ranking different aspects of universities separately, the outcome (if it is funded) will disappoint. It will reproduce the very failings that are said to mar the present league tables. For anybody who thinks it advantageous will be able to aggregate the separate scores to create a unitary league table, just as the aggregated scores of current league tables are now commonly disaggregated by the public relations departments of universities in order to publicise the more favourable aspects of their scores. Once comparative measures of university performance are compiled, it is easy to combine them in various ways to create rankings, and once that is done it is easy and tempting for institutional leaders and others to claim that carefully selected rankings should be viewed as the central and objective measures of institutional quality.

Excellence and success

Metrics and the league tables created out of them are supposed to provide objective measures of the quality of universities, which can be used in the first place to compare and to rank, but also to penalise and reward. When connected to funding decisions, they provide potent measures of accountability. Yet league tables are not, in my view, useful ways of judging university excellence. The very diversity of universities, and the fact that ranking is a high stakes affair that matters *all too much* to university administrators, and indeed to academics and students, paradoxically ensures that the league tables will *not* offer good ways of holding universities to account: they hold universities to account for *achieving or appearing to achieve* some *comparative* success. But that success is not always evidence of excellence, and excellence is not always reflected in rankings in league tables.

Aristotle's *Nichomachean Ethics* begins with the famous words

> Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good; and for this reason the good has rightly been declared to be that at which all things aim.[18]

In the following chapters Aristotle investigates the goods at which we aim, and argues that they are not unitary. In Chapter 6 he concludes that 'good is said in many senses': there are many *aretai* or *excellences*, but there is no

---

[11] See Arum and Roksa, note 2.

[12] See the discussion of the use of school rankings in Harvey Goldstein and Beth Foley, *Measuring Success: League tables in the public sector* (British Academy Policy Centre report, 2012), available at www.britac.ac.uk/policy/Measuring-success.cfm

[13] Shanghai Jiaotung University http://www.arwu.org/index.jsp#–

[14] http://www.timeshighereducation.co.uk/world-university-rankings/

[15] For example, the fact that there are excellent institutions that concentrate just on teaching (the French *grandes écoles*) or that only do research would not then lead to a poorer ranking.

[16] The U-Multirank project was initiated and funded by the European Commission (DG Education and Culture), and is not yet complete. See http://www.u-multirank.eu/

[17] *University Rankings: Diversity Excellence and the European Initiative*, League of European Research Universities, Advice paper June 2010. It includes the following caustic remark: '... another expensive tentacle of the audit culture? Is there evidence that there is a lack of "transparency" about HEIs in Europe that inhibits either potential students or potential collaborators in making sensible choices that is sufficient to justify creation of a costly and time-consuming enterprise?' http://www.leru.org/files/publications/LERU_AP3_2010_Ranking.pdf. And see also *The Modernisation of Higher Education in Europe* (27th Report of European Union Committee, House of Lords, March 2012, HL Paper 275).

[18] Aristotle, *Nichomachean Ethics*, trans. D. Ross, I I, 1094a.

overarching excellence to which all others are subordinated.[19] However, if excellence is not unitary, compiling measurements (of variably quality and comparability) into league tables raises distinctive and difficult questions.

Excellence is surely a noble aim for universities – as for other institutions and activities. However, since there are many excellences, and since those of universities vary with the activities they undertake, it may be hard to measure how good a university is, or to determine how much better it is than another university. Once we acknowledge the plurality of excellences that universities may seek, we can no longer imagine that those who seek excellence can simply aim to do better than other universities, although that can (but need not) be one of the results of striving. Where standards are low, even the most successful may be less than excellent; where they are high, even those of great excellence may not be the most successful.

A good reason for taking the Aristotelian conception of excellence seriously is that we are not then compelled to see the pursuit of excellence as a *zero sum game*: we can imagine, indeed encourage, a world in which all universities do excellent teaching and research. By contrast, we cannot even imagine a world in which all universities are equally successful in teaching or research, since success, unlike excellence, is a *positional good*.

Conclusion: Limits of extra mural accountability

These rather depressing reflections on current fashions for university accountability are not an argument against measuring achievement and success. There are often good reasons to do so. But if there are good reasons to do so, I suggest that it would be better to measure only what can be measured with reasonable accuracy – not necessarily with precision – and to refrain from measuring matters that can be manipulated or massaged by those who are to be held to account.

For some time it has struck me as surprising that we learn so little about universities from the league tables, and that we seldom see scores on various useful measures that are not open to manipulation. I have come to suspect that this may be because universities and academics – and perhaps the public at large – prefer *not* to have accurate information. Such information might after all show up realities that many would prefer to cloak. It might show up real differences in quality.

It is noticeable that educational achievements that can be measured with reasonable accuracy are seldom included in rankings and ratings. For example, it would be useful to know how hard the students at a given university work – but this is not generally done (we know in the UK – but not from the league tables – that students doing certain degrees, such as medicine at Cambridge, or at Imperial, work about *three* times the number of hours per week of the average British student). It can be useful to know how competently students speak and write the language of instruction – both at registration and at graduation: and this can be done; but is seldom done. (This struck me because we do it at the University of Sharjah, of which I am a Trustee, which is a university in the Gulf with English as the medium of instruction: TOEFL scores, *administered independently of the university* provide us with a reasonably reliable measure.) It can be useful to know how many pages of written work a student produces in a year or semester, and how many of these pages receive detailed comment and feedback from instructors: this is highly variable between universities, yet is wholly ignored by standard metrics of university excellence. It looks as if the enthusiasts for metrics and quality assessment may be reluctant to measure matters that are educationally revealing. Similar points can be made about research metrics, where counting the number of outputs (e.g. publications, or specifically peer-reviewed publications per annum) at least provides a measure of diligence. However, these metrics are respected only to the extent that they shadow serious, and time-consuming, academic judgement – for example the judgements that go into the evaluation of grant proposals and peer review for publication.

It is, I believe, still far from evident that the complex extra mural loops of accountability that have been constructed in recent decades achieve their supposed objectives. Many do not measure university excellence in convincing ways; some divert academic and institutional time and resource in ways that detract from excellence. At their worst, they create perverse incentives. Even when they do not do so, they divert attention from excellence to comparative success defined in narrow ways. Are these the best way of holding universities to account that we can imagine or devise?

Baroness O'Neill is Honorary Professor emeritus of Ethics and Political Philosophy, Faculty of Philosophy, University of Cambridge; and past President of the British Academy.