



King's Research Portal

DOI:

[10.1007/s10734-018-0300-z](https://doi.org/10.1007/s10734-018-0300-z)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Spence, C. W. (2018). 'Judgement' versus 'metrics' in higher education management. *Higher Education*.
<https://doi.org/10.1007/s10734-018-0300-z>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



'Judgement' versus 'metrics' in higher education management

Crawford Spence¹ 

Published online: 06 August 2018
© The Author(s) 2018

Abstract

This article argues that universities currently privilege an instrumental *ethos of measurement* in the management of academic work. Such an ethos has deleterious consequences, both for knowledge production and knowledge transfer to students. Specifically, evidence points towards the production of increasingly well-crafted and ever more numerous research outputs that are useful in permitting universities to posture as world class institutions but that ultimately are of questionable social value. Additionally, the ever more granular management of teaching and pedagogy in universities is implicated in the sacrifice of broad and deep intellectual enquiry in favour of ostensibly more economically relevant skills that prepare graduates for the travails of the labour market. In both cases, metric fetishization serves to undermine nobler, socially minded visions of what a university should be. For such visions to flourish, it is imperative that universities take steps that explicitly privilege a collegial *ethos of judgement* over a managerialist *ethos of measurement*.

Keywords Measurement · Metrics · Judgement · Teaching · Research · Excellence · Managerialism

Introduction

In this interpretive essay, it will be argued that dominant modes of managing teaching and research in Higher Education Institutions produce significantly dysfunctional effects. These effects stem not from the management per se of teaching and research but from the specific metricised approach that has become prevalent in both cases. Through an obsession with measurement over judgement and 'evaluative forms of governance' (Ferlie et al. 2008) that are increasingly granular and instrumental in orientation (Ekman et al. 2018), teaching and research in UK Higher Education Institutions (HEIs) are essentially being managed to the point at which Humboldtian intellectual ideals of broad and critical thinking are undermined. The UK is focused in here as an extreme case or, perhaps, as a harbinger of more widespread

✉ Crawford Spence
Crawford.spence@kcl.ac.uk

¹ King's College London, 30 Aldwych, London WC2B 4BG, UK

marketisation to come elsewhere. In a period of 20 years, the UK has witnessed the transformation of a previously elite higher education system that was overwhelmingly state funded into a mass higher education system in which funding has been pushed onto students (Parker 2018). On one hand, pursuing research excellence in the current policy climate in the UK leaves us with vast quantities of articles and other outputs that are well-crafted but often lacking in social meaning. On the other hand, the new agenda for teaching excellence in a marketised environment is encouraging institutions to take all sorts of intellectual short cuts in order to improve statistics on student satisfaction and employability. Overall, the pursuit of excellence, as currently conceived by UK HEIs, engenders mediocrity. HEIs would be much better advised to stick to Humboldtian ideals of cultivating critical minds rather than the instrumentalism of student satisfaction scores or nurturing a culture of ‘big hits’ in ‘top journals’ for academics.

Specifically, the management of teaching and research in HEIs is both desirable and necessary, but it needs to be guided by a principle of privileging judgement over measurement. Privileging the latter over the former is currently prevented in HEIs and denotes a lack of confidence in academic expertise, mistrust of the professoriate or, perhaps, a more straightforward strategic decision to enter into the competitive world of league tables and ranking schemes (Espeland and Sauder 2007). Either way, the limits of measurement need to be recognised: ‘measurement may provide us with distorted knowledge—knowledge that seems solid but is actually deceptive’ (Muller 2018, p. 3). Metrics can help, it will be argued below, provided that they do not replace judgement based on experience. As a profession based upon expert knowledge (Abbott 1988), the accumulation of which takes many years, academia thrives upon decisions being taken on the basis of relational expertise (Barley 1996). Relational expertise denotes the embodiment of tacit, situational knowledge over time. For example, lawyers tend to outperform lay people in civil cases, not primarily because of their greater mastery of legal doctrine but because of their ability to read the cues in the courtroom, their sense of when a particular argument is likely to be more persuasive than another and their knowledge of how to deal with other relevant actors in the legal environment (Sandefur 2015). It is this relational expertise that underpins aspects of professional judgement in many professional jurisdictions, including academia where historically peer review has been the key mechanism for various aspects of academic work, from the consecration of PhDs through to the way in which research outputs are reviewed and evaluated. Yet, dominant ways of measuring individual and organisational performance in the academic world increasingly marginalise the role played by relational expertise in deference to a growing constellation of crude performance measures that are geared towards facilitating external accountability. This reflects a suppression of collegial social control by a more market managerialist orientation (Martin et al. 2015) through which the very notion of academia as a profession is undermined.

The essay proceeds as follows. The following section outlines a number of problems that emanate from privileging measurement over judgement in the domain of research. These problems are illustrated by reference to ongoing debates over the role of journal rankings in academic performance measurement systems and emerging concerns over the instrumental accounting of research impact, the latter being increasingly institutionalised in response to the UK’s Research Excellence Framework (REF). A subsequent section then looks at how the privileging of measurement over judgement in the management of teaching leads to counter-productive outcomes for both students and society. These outcomes are illustrated by reference to literature on the increasing salience of employability criteria in academic programmes and consideration of the way in which the Teaching Excellence Framework (TEF) focuses on

narrowly conceived outcomes for students. The essay then concludes by offering some suggestions for a management system guided primarily by judgement rather than measurement.

Managing research production lines

Although knowledge production clearly requires some degree of co-ordination and organisation, problems arise when an essentially diffuse and creative process becomes subject to the disciplining power of numbers. Therefore, it will be argued below that the limits of effectively managing research lie somewhere within the metricisation of knowledge production. Once metrics start to appear, dysfunctional behaviour often ensues, and intelligent judgement can easily be thwarted. This section will attempt to locate broadly where the roots of the current problems lie. The problems are indeed significant. While knowledge production in the modern university is burgeoning in quantitative terms, qualitative assessments reveal an increasingly hollow state of affairs.

In one respect, the notion that the modern university is actively encouraging intellectual vapidness confounds what one witnesses on the surface. There has been a steady proliferation of research outputs in many countries in recent decades (Alvesson et al. 2017). Successive research assessment exercises and increasing rankings of schools, subject areas and institutions have been incredibly successful in stimulating research activity to present levels, seemingly showing that research is consistently improving (Marginson 2014). By various accounts, academics are now busier and working faster than ever (Berg and Seeber 2016). However, it does not necessarily follow that all of this research production has something interesting to say. Indeed, Alvesson et al. (2017, p. 4), writing about the social sciences, suggest that we are currently witnessing ‘a proliferation of meaningless research of no value to society and of only modest value to its authors—apart from in the context of securing employment and promotion’. The social value/productivity tension is evident rather starkly in the hard sciences as well, with Nobel laureates such as Peter Higgs suggesting that current productivity pressures are anathema to the pursuit of major scientific discoveries (Aitkenhead 2013).

Alvesson et al.’s (2017) contention is by no means unique in this respect (see also Berg and Seeber 2016; Butler and Spoelstra 2014; Tourish and Willmott 2015). It is that research has become a game for many social scientists, whose work is meaningful in some senses but not in others. For example, producing research can be intellectually stimulating for the researcher as well as being a means to the end of career advancement—so it is *meaningful to the ego* (Alvesson et al. 2017). Producing research might also be *meaningful to the micro-tribe* to which the researcher belongs or seeks access to, thereby creating a sense of belongingness and group identity. However, it is only rarely that the vast quantity of research outputs produced in social science is *meaningful to society*, in that it deals with pressing social issues or proffers solutions or relevant critiques of current institutional arrangements.

On an individual level, concerned researchers might find moral wayfinders in *cris de coeurs* such as Alvesson et al.’s (2017 and Berg and Seeber’s (2016). The latter argues for a culture of slowness in the academy, while Sennett’s (2008) lionisation of craftsmanship—doing good work for its own sake rather than for some instrumental end—also calls for a focus on work that is somehow insulated from the surrounding zeitgeist of meaningless arguments, mindless productivity and ephemeral goods and services. However, to focus on individual strategies for coping in such contexts would be to ignore the structural forces at play. In other words, the

management systems and imperatives are evident in HEIs which encourage researchers to transform themselves into well-oiled, fully charged 4×4^1 publishing machines.

Journal rankings and managerialism

Many of the above concerns are embedded within the ongoing debate about the role of journal rankings and research performance measurement. However, it should be noted from the outset that it is hard to find any voices that call for the complete abolition of all forms of measurement. Even the harsher critics of the current ‘regime of excellence’ recognise either that ‘some performance pressure and measurement is not necessarily bad’ (Alvesson et al. 2017, p. 21) or that there has been a ‘perceived shift...from a ‘system of patronage’ that relied on personal bias and favouritism to a more meritocratic approach that is based on journal publications’ (Butler and Spoelstra 2014, p. 542).

The problem, as a number of scholars have pointed out, comes with the fetishization of journal rankings and the suspension of analysis that often accompanies both these and other ranking schemes, such as citation counts and H-scores Tourish and Willmott (2015); for example, note how the *Association of Business Schools Academic Journal Guide* (the Guide, hereafter) is used in the UK, and increasingly elsewhere, by business school deans to inform decisions about the allocation of teaching and administrative duties and is the main reference point when it comes to promotion, retention, appointment and probation deliberations. This is a problem, Tourish and Willmott (2015) argue, because the emphasis in academic publishing then shifts from the quality of ideas to a concern with publication destination. According to commentators, these dynamics have resulted in ‘a great stream of publications that are both uninteresting and unread’ (Muller 2018, p. 79) or, more worryingly, in ‘a social and intellectual closure [and] a narrowing of practices and expectations among academics’ (Parker 2018, p. 93).

Hence, academics often congratulate each other on the basis of ‘big hits’ in ‘top journals’ rather than discussing the content of each other’s work. For example, exchanges in the corridors of the university increasingly consist of academics talking about other academics having ‘hit X journal Y times this year’, either out of envy, admiration or most likely both. Discussions of the ideas contained in the articles in question are rarer, implying that the content of the work is less important than a simple assessment of how difficult it was to have it published. Listening to such posturing should really prompt feelings of ambivalence and despondency at the lack of intellectual engagement. The latter can induce a feeling of being out of sync with the surrounding rules of the game (Dirk and Gelderblom 2017). Sadly, the reality is that the rules of the publishing game have been internalised so much that to suggest politely that the conversation turns to the ideas contained within said articles would most likely expose the proposer as naïve and somehow missing the point of the entire academic enterprise.

Alvesson et al. (2017) take this further by suggesting that many academics do not really read in a thorough way anymore. Rather, they engage in vacuum cleaning literature reviews in a symbolic attempt to demonstrate to reviewers that they have identified the key arguments in a particular stream of literature, all with the end of identifying an all-important ‘gap’ in said research. This privileges formulaic research and a ‘drab uniformity’ (Tourish and Willmott 2015, p. 41) that makes only incremental contributions to existing debates, rather than offering

¹ 4×4 denotes the academic equivalent of an all-terrain vehicle that produces the maximum possible score in a UK REF assessment—4 published outputs all rated in the top category of 4*.

ground-breaking or challenging ideas. New journals in challenging areas or that encourage alternative methodologies will struggle to thrive in such an environment, as researchers are increasingly encouraged to publish only in well-regarded outlets that already have established impact factors.

Current measurement regimes breed conservatism, the main beneficiaries of which are those engaging in US-style positivist research methodologies for US journals (Tourish and Willmott 2015). Many academics buy into this to some extent, partially if not wholeheartedly, even if doing so is at odds with their own deeper intellectual proclivities. This reflects a wider state of affairs in the social sciences, in which individuals are producing work that is publishable and yields them a return on investment in career terms (Alvesson 2012) rather than that which excites them or is socially worthwhile.

Measuring impact

A notable feature of the REF in the UK is the new focus on impact, which accounted for 20% of overall REF scores in the 2014 exercise. Governmental concern with the social and economic impacts of academic research has a long history, but in recent times, it was catalysed primarily by the Worry Report in 2006. Impact will rise to 25% in REF 2021. Arguably it will account for 29% of the total if the additional environment-related impact criteria are also included. Impact is defined as ‘all kinds of social, economic and cultural benefits and impacts beyond academia’ (Funding Councils 2011, para 11.a). At the operational level of university research committees, the definition is no more precise. Trying to unpick what it actually means is often met with the rather nebulous answer of ‘reach and significance’.

The definition of impact is therefore wide ranging, perhaps implying that it might elude easy measurement. Power (2015) documents the way in which a tentative and somewhat arbitrary infrastructure for evaluating research impact was developed within one Russell Group institution in the lead up to REF 2014. The measurement of impact is now widespread in UK universities, although its specific form is not isomorphic as is the measurement of research output quality via journal ranking lists. Indeed, Power (2015, p. 44) suggests that impact measurement itself is not merely drawn from the performance measurement zeitgeist but ‘represents a *radicalisation* of prior managerial trends in higher education by institutionalising the demonstrable use value of research as a new norm of academic performance evaluation’ (*emphasis added*).

Recognising the difficulties of measuring in quantitative terms the impact of social research, universities present evidence of their ‘reach and significance’ via Impact Case Studies (ICS). There is little in the way of guidelines about how to go about constructing an ICS, which possibly explains the wide ranging and somewhat volatile REF scores attached to this particular area in the 2014 exercise. Nevertheless, a number of practices have become common within universities in order putatively to facilitate the capture of, and offer evidence for, the reach and significance of research. One of these practices is ‘solicited testimony’, which involves not merely discovering impact but, in many ways, its active construction via asking ‘friends in high places’ to write testimonials explaining how a particular research project was central to a policy initiative or a change in organisational behaviour. According to Alvesson et al. (2017, p. 128), this has ‘encouraged a tissue of fabrications, hypes, and lies where the most spurious connections were made in order to tick the boxes and claim relevance for dubious academic work’.

The construction of ICS started out as an ad hoc process, with researchers authoring their own narratives. However, over time, an impact regime emerged which produced more stability and standardisation in the production of ICS. In Power's (2015) institutional case study, this involved the use of journalists to do the eventual writing, a practice which has been taken up by other institutions as well. Many institutions are in the process of hiring 'impact officers' to prepare for REF 2021. Essentially, impact experts are becoming embedded into universities in an attempt to provide regularity, predictability and monitoring capacity to the whole process.

As with journal rankings, the development of this calculative infrastructure produces problematic consequences. While in one sense, the exhortation to do something beyond merely bagging a 'big hit in a top journal' opens up the possibility for more socially meaningful research; the somewhat contrived means of evidencing impact encourages a more instrumental and conservative approach to the research process. Focusing on the inappropriately measurable and faux evidential once again discourages risky endeavours whose consequences are diffuse, multiple and perhaps simultaneously positive and negative in favour of the neat and the tidy. We are, once again, not a million miles away from the behaviour of Alvesson et al.'s (2017) meaningless-modern-day-gap-spotting-academic.

Summary

Knowledge production in the modern university is increasingly devoid of passion and vibrancy. Although not lacking rigour or refinement, academic research is ever more driven by a combination of career advancement and inter-university competition rather than free-spirited intellectual enquiry. In many ways, this is the product of the perverse management of academic research, both in terms of research outputs and its wider socio-economic impact. There is something very akin to a bubble economy in knowledge production, and many academics appear to be riding the boom until the bubble bursts. To avoid the traumatic effects that a system-wide crash would engender, action that reduces the bubble and brings knowledge production into line with socially meaningful criteria could be taken now. This would probably require a significant reduction in the quantity of research outputs and a reorientation towards quality. It would also require a relaxation, but not elimination, of performance metrics and an erosion of other calculative infrastructure. We will discuss suggestions for a more enlightened research management regime more fully in the [concluding section](#), after considering similar metric-induced problems with the management of teaching in the modern university.

Managing the student experience and measuring teaching excellence

The approaches above to the management of knowledge production are essentially institutional responses to the REF. Similar dysfunctional responses to the TEF are starting to be discerned within UK universities, even though the origins of the TEF were laced with quite different ideological imperatives than those underpinning the REF. The REF can be traced back to its original incarnation as the Research Assessment Exercise (RAE) in 1986, which emerged as a mechanism to distribute public funding to universities on the basis of assessed research quality (Barker 2007). Performing well in the RAE/REF historically had significant economic consequences. The economic stakes have reduced dramatically in recent years as universities are increasingly funded by means other than the public purse. Nevertheless, an institutional obsession with doing well in the REF remains, pointing towards the symbolic

power that ranking schemes come to exert over academic institutions (Espeland and Sauder 2007, 2016; Pusser and Marginson 2013).

Underlying the TEF is a different rationale. The TEF is ostensibly being put in place by the UK government in order to ensure that teaching is no longer the poor cousin to research within universities. However, it is also explicitly part of a conservative agenda to treat students as consumers at the heart of, what it sees as, a market for educational services (Collini 2016). Key to this is ensuring that students have a satisfactory experience (as measured by the National Student Survey, upon which the TEF heavily relies) at university, that they see their degree through to completion (Higher Education Statistics Agency data on drop-out rates here are used) and that they come away employable and possessing the skills that are demanded by UK plc.

One key motivation underpinning the TEF is accountability, specifically in the context of informing students so that they can make better educational choices. Arguing against accountability or transparency seems anti-progressive. However, Best (2005, p. 105) argues that surrounding any discourse of accountability or transparency are always very specific political intentions, more often than not tied to the ‘perverse, speculative logics of financial markets’ (see also Strathern 2000). At the macro level, for example, transparency is often advanced as a fig leaf for neo-liberal structural adjustment. In the context of the current political zeitgeist, calls for the greater accountability and transparency of educational institutions can be traced back to new public management initiatives and are closely tied to consumerist understandings of education (Muller 2018). As such, greater accountability and transparency in this context are likely to encourage a much narrower student experience rather than higher standards, quality, choice, etc. Indeed, the parallel with similar initiatives in the healthcare sector is instructive. According to Muller (2018), a number of studies have shown that surgeons become less willing to operate on cardiac patients following the introduction of publicly available metrics. If we want our educators to take risks and challenge students, public accountability measures spawned by the present political and intellectual zeitgeist are likely to make things worse.

Employability

Governmental emphasis on employability was evident from the government white paper of 2016:

‘This government is focused on strengthening the education system...to ensure that once and for all we address the gap in skills at technical and higher technical levels that affect the nation’s productivity’ (BIS 2016, p. 10).

The increasing focus on employability in HEIs, which many studies presumed to be a rather anodyne and uncontroversial phenomenon (Silva et al. 2016), is nevertheless particularly problematic when looking at the way in which universities are attempting to manage and measure it. Indeed, Frankham (2017) goes as far to say that the approach to boosting employability that is emerging within universities appears not to prepare students for the workplace. For example, from the outset, the skills that employers want are highly situated and contextual and, as such, cannot be fully taught in universities.

Another thing that universities are now doing is to lay on increasing numbers of career events or practical skill sessions. In many UK HEIs, any proposed new bachelors or master’s degree will not even be considered at committee level until a stakeholder consultation has been undertaken to ensure positive employability outcomes for prospective students. This appears to

encourage not merely all sorts of empty verbiage in course proposal forms but has actual substantive implications such as lecture slots being given over to a session on ‘professional skills’ or ‘career development’. Research does show that these career-focused interventions are correlated with better graduate employment outcomes (Taylor and Hooley 2014). However, more critical studies show how careers services are implicated in the production and reproduction of social inequality, with particular institutions tending to funnel students towards particular jobs (i.e., elite students pushed towards more elite jobs, less elite students towards less elite jobs, etc. Binder et al. 2015). Therefore, HEIs pursuing ‘employability’ for their graduates might not be a welfare-neutral or welfare-enhancing enterprise when viewed from a wider socio-economic vantage point.

Focusing on employability metrics can be problematic in other ways too. In business schools, for example, an increasing portion of MBA programmes are delivered, not by academics or external practitioners, but by in-house careers teams (Hunt et al. 2017). This is driven largely by international MBA rankings such as those run by The Economist and The Financial Times, both of which are heavily weighted towards future career outcomes. Some students enjoy these sessions, but others see them for the perfunctory box ticking exercise that they are. In response to this, career teams are constantly looking for new ways to jazz them up. Lezaun and Muniesa (2017) suggest that there is an increasing tendency towards experiential curricula in business schools, which effectively constitute a ‘subrealist shield’ that insulates students from challenges that await them in the real world. For example, initiatives from Warwick Business School’s MBA programme include rebranding a ‘coaching and negotiations’ module as ‘courageous conversations’ or hiring comedians to come in and instruct students in the art of scripting impromptu stand-up routines as part of cultivating a more dynamic persona. We are a long way here from traditional, intellectual conceptions of what a university education is as the focus on such programmes is on how to develop an entrepreneurial, career-focused sense of self rather than a skillset that might actually be useful for solving organisational and societal problems (McDonald 2017; Parker 2018). Indeed, we might actually be quite far from employability as well, as much of this corresponds more to infotainment than anything else.

It is not just in the UK or in relation to MBA programmes that concerns have been raised about the dilution of academic standards. Arum and Roksa (2011) highlight similar trends in the USA, where HEIs are characterised as *academically adrift* while being socially alive, active and attentive. In a longitudinal study of multiple US colleges, Arum and Roksa (2011) demonstrate, *inter alia*, that 50% of students have no single class which required more than 20 pages of writing over the course of the semester, and 36% of students study alone less than 5 hours per week. Essentially, Arum and Roksa (2011) argue that student experiences on campus are decreasingly about academic attainment and increasingly about socialisation, psychological well-being and extra-curricular activity. In other words, university is failing to sufficiently enhance the academic skills of students between enrolment and graduation. In a follow-up study looking at the same students postgraduation, Arum and Roksa (2014) show that the result of this negligence of cognitive formation is that graduates spend prolonged periods adrift, effectively postpone adulthood as they struggle to find high-quality graduate jobs and often have to depend on their parents for financial support and delay commitments related to marriage, children and home ownership until later than they otherwise would. In other words, academically adrift HEIs are producing graduates who lack the critical thinking skills that are actually required to enter the graduate labour market and make informed life choices.

This crisis of cognitive formation has elsewhere been explained in terms of a gradual displacement of knowledge from both society and the curriculum (Wheelahan 2010). As a result of neo-liberal policy reforms and the advent of the knowledge society HEIs have, for a number of decades now, been under pressure to ensure that they produce graduates who act in the national economic interest. This leads to a focus on generic marketable skills rather than specific theoretical knowledge. A fundamental contradiction is engendered by this in that the knowledge society does not seem to value knowledge, as the latter has become ‘relativized, regarded as ephemeral, transient and ultimately unreliable, even though its role has never been more important’ (Wheelahan 2010, p. 105). Placing great emphasis on employment outcome metrics in curriculum design is not merely a reflection of deeper cultural trends, but a radicalization of a new ‘vocationalism’ that can only exacerbate the intellectual dilution of the student experience.

Focusing minds on graduate employability in this way was widely criticised as the TEF proposals made their way through the House of Lords (the UK’s secondary chamber of government), and it has received even greater criticism since (Shattock 2018). The approach to tabulating employment success—focusing on destinations 6 months postgraduation—also creates perverse incentives for institutions. Postgraduation student gap years and peripatetic career paths are now seemingly under threat for no valid intellectual reason other than those institutions—conscious of the metrics—who are working harder and harder to shepherd their final year students into the voracious jaws of the graduate labour force.

More generally, in lionising ‘anticipated returns from the labour market...as the ultimate measure of success’ (Collini 2018, p. 38), the ‘student’ herself risks becoming objectified as a docile subject. The obsession, rational for the most part, with obtaining either an upper second- or first-class honours degree, places greater pressure upon academics to ensure that a clear path to these classifications is laid out for students and that there are no tricky obstructions in the way. According to Frankham (2017), this leads to a situation whereby students are increasingly spoon-fed curricula; the examination of which is more predictable and controllable. Indeed, it is a curious paradox that, just at the moment when students appear less intellectually engaged than ever, grade inflation is an increasingly widespread phenomenon (Bachan 2015). With teaching staff increasingly under pressure to ‘teach to the metrics’, grades rise just as critical enquiry falls by the wayside (Muller 2018, p. 71). It is hard to see how such a situation can effectively contribute to better societal outcomes in terms of a dynamic and skilled labour force for a modern knowledge economy. Indeed, as Arum and Roksa (2014) show, focusing overtly on employability concerns and investing heavily in careers services, while not simultaneously ensuring rigorous and challenging academic standards in the classroom, will actually lead to adverse employability outcomes for graduates.

Teaching excellence

‘Excellence’ is a laudable ideal, one basically impossible to argue against. As with any vague concept, it essentially eludes measurement, yet the current institutional context encourages, indeed forces, institutions to enumerate and tabulate in the name of accountability. Measuring the simple when the outcome is complex is a characteristic flaw of many metricised environments (Muller 2018). Measurement of what purports to be excellence requires taking the complex and making it discrete. This is achieved by its fragmentation into a series of concepts such as employability, skills or student satisfaction. To the extent that these concepts ignore the complexities of socio-economic context, as they tend to do, focusing on them will more likely

succeed in cultivating a culture of mediocrity than a culture of excellence (Gourlay and Stevenson 2017). Instead—consistent with this article’s emphasis upon professional judgement—Gourlay and Stevenson (2017) suggest that excellence be conceived as ‘relational’, meaning that it is highly context specific and impossible to capture in student satisfaction scores. The latter can be boosted through a series of crowd pleasing initiatives—giving clear indication of answers to the exam, showing DVDs, telling good jokes—which may be anathema to challenging students intellectually. Indeed, there is a substantial literature showing that student evaluations of teacher effectiveness and student learning are somewhat decoupled from each other, implying that the former are a spurious basis for measuring teaching excellence (see, for example, Boring 2017; Uttl et al. 2017)

In this context, the elevation of the National Student Survey (NSS) to such a level of importance that it might be used to determine fee levels that universities can set seems perverse at best and tyrannical at worst. The NSS is a low-cost, crude way of measuring teaching quality, although it does not actually measure quality at all so much as student reactions to their experience at university (of which teaching is a part). Survey scores are influenced by a whole host of non-teaching-related issues, including how institutions ‘sell’ the survey to students, what political issues are holding the attention of students’ unions at the time of the census, whether there is a pension dispute ongoing, and so on. This illustrates a problem with metric led approaches to performance assessment—they often miss the point of what they are supposed to be measuring (Muller 2018).

It should be noted, however, that the TEF claims to *combine* judgement with metrics and is, therefore, not metric-led. Fifteen-page narrative reports are provided by institutions to supplement the metrics. This role for narrative submissions needs to be qualified along two dimensions. First, as Shattock (2018) has astutely observed, each of the six core metric ratings is assigned a number of ‘plus’ or ‘minus’ flags. For example, only institutions with at least three ‘plus’ flags and no ‘minus’ flags are eligible for a gold award. Effectively, this weighs outcomes heavily in favour of metrics rather than narratives. Second, even if the metrics are departed from some limited extent, this non-metricised space is still not filled with any actual assessment of teaching quality, only stylized discussions thereof.

To be clear, none of this criticism is tantamount to a suggestion that teaching quality should not be assessed in some way. On the contrary, just as Tourish and Willmott (2015) argue that peer review is a better mechanism for assessing research quality than journal rankings, Collini (2016) suggests that the largely metric-based system that makes up the TEF be replaced with a system of inspections, much as Ofsted does with primary and secondary schools in the UK. This would permit judgement, not measurement. Such a system would no doubt have its flaws—and would most likely be more expensive for the government²—but it would obviate many of the deleterious consequences outlined above by permitting a direct evaluation of the object putatively at the heart of the whole exercise, namely teaching.

The TEF argues Collini (2016) will succeed in producing the following: a cadre of experts whose job it is to administer the TEF, a greater role for business in shaping university curricula, more ludicrous posturing and self-aggrandizing by universities, more league tables and, more gaming of the system. However, what it is unlikely to produce is better teaching. Similar views are expressed by Shattock (2018, p. 22), who predicts that ‘some of the best minds in institutions will be devoted to ‘gaming’ the data to ensure that their institutions are positioned

² In the short term this might be true, but a full accounting of the money that institutions will spend on developing internal TEF infrastructures would quite probably show it as a misallocation of valuable resources.

to protect their brand'. Indeed, others have suggested that universities are already spending more time and money on preparing for the next TEF exercise than they are in trying actually to improve teaching (Race 2017). Of course, the two are ostensibly equivalent, but there seems to be an increasingly widespread acceptance among academics that they are not the same thing, much as institutional responses to the REF are widely seen as being interested in something other than actual research excellence.

Summary

Many of the previous problems identified with the REF—stimulation of a bubble economy of research of dubious social import; expansion of performance measurement regimes at the university level which produce dysfunctional behaviour, etc.—are therefore being replicated with the TEF. This is interesting precisely because of the different origins of both assessment exercises. The dysfunctional nature of both therefore must be explained by something beyond political orientation, as each was initiated by different governments and for different reasons. The problems lie with explicit programmes of consumerisation that privilege measurement over judgement³. Narrow focus on measurement gives universities the illusion that they are managing and controlling something more effectively, but doing so produces consequences that run counter to the spirit of the overall exercise.

Towards judgement over measurement

It has been argued that the management of both research and teaching in modern HEIs is distorted by an overarching ethos that privileges measurement over judgement. From the evaluation of research impact through assessments of teaching quality and student experience, priority is given to those aspects which can be relatively easily subjected to quantification and, by extension, whose progress can be charted over time. Such measurement regimes invariably produce dysfunctional behaviour as many academics start to game the system and fudge the numbers by looking for quick wins on easy metrics. Early career scholars are particularly susceptible to this behaviour becoming part of their scientific habitus. Academics suffer in this environment, with laudable notions of 'collegiality', 'standing in society', 'guild identity' and the 'dignity of learning' all starting to sound rather anachronistic (Collini 2018). However, the main losers are ultimately the students, who leave university increasingly ill prepared for the real world, and society, which fails to get what it needs from its public institutions. The failure to provide students with what they arguably really need is somewhat paradoxical, as we find ourselves in the midst of a consumerist, marketised zeitgeist that privileges the 'student experience' above all else. But the customer is not always right, or necessarily fully aware of what is in his or her best interests. As such, 'consumer satisfaction is not a worthy aim for colleges and universities' (Arum and Roksa 2014: 2561).

The solution to this measurement-induced dysfunction is not necessarily less management of teaching and research. This essay is not an argument *against* management per se. Rather, it is a call for more enlightened, more reflective and more holistic approaches to managing

³ It should be noted, however, that metricisation is not exclusively a market-based phenomenon, as studies looking at performance measurement in various areas of the public sector have shown (see, for example, Hood 2006).

teaching and research. The issue is not ultimately metrics *or* judgement, but ensuring that the former inform—and are ultimately subservient to—the latter. Pollock et al. (2018) note that one of the dominant themes in the sociology of evaluation is that of ‘reactive conformance’, in which organisations are enslaved to the measure. However, Pollock et al. (2018) suggest that the multiplicity of ranking schema to which organisations are now subject actually creates a space for organisational agency such that organisations can pick and choose ranking schema to which they submit themselves and engage in a more active process of ‘reflexive transformation’. In the context of HEI, there is possibly room for reflexive transformation *vis-à-vis* research, although organisational agency is more circumscribed by the emerging institutional context when it comes to teaching.

In the context of the REF, the problems are not the REF itself—although it is far from a perfect system (Marginson 2014)—which proceeds on the basis of peer review, is staffed by eminent academics rather than managers and is open to significant amounts of professional interpretation and judgement. Rather, the main problem is the subindustry of performance measurement that HEIs have established in response to REF imperatives, such as the reification of hugely contentious journal ranking schema or the elaboration of desperate and spurious approaches to demonstrating research impact. Even though some publication outlets are clearly of higher quality than others, any system that encourages academics to focus more on *where* they publish than *what* they publish is intellectually deleterious. The REF itself does not do this, but universities do. The removal of journal lists from internal REF management processes, hiring decisions and promotion panels, or at the very least approaching such lists with extreme scepticism, would go a long way towards the stimulation of intellectual discussion between academic colleagues rather than the muscular, frenetic competition that is promoted by the present system. As Muller (2018, p. 80) states:

‘What, you might ask, is the alternative to tallying up the number of publications, the times they were cited, and the reach of the journals in which articles are published: The answer is professional judgement.’

As painstakingly obvious and blunt as this might sound, the best way to evaluate the importance of a research output might be to actually read it. Of course, peer review systems are imperfect and subject to political dynamics of their own. Not all judgemental outcomes of peer review processes are good ones, and even higher quality journals can exhibit haphazard reviewing or editing behaviour at times. However, such a system does at least open up the possibility that good judgements be arrived at by reflexive individuals whose intellectual dispositions have not been fully colonised by neo-liberal policy mantras. The same cannot be said of systems that crudely privilege numbers over anything else. Metrics and ranking schema are not neutral but tend to reproduce dominant norms (Pusser and Marginson 2013). This is not to say that numbers are not important, on the contrary. Metrics on journal outputs and citations can help inform judgements but cannot be used mechanically as a substitute for this. Numbers can help management, but management by numbers reduces *homo academicus* to an automaton geared towards the routine production of kilometres of sophisticated yet solipsistic verbiage.

As regards the TEF, Collini’s (2016) suggestion for an inspection system rather than a metricised submission regime would also make a lot of sense, although commentators suggest that such an approach would be more burdensome and costly than pre-existing metrics (Shattock 2018). In some instances, measurement *per se* here is not the problem so much as the reification of rather consumerist measures such as NSS or Postgraduate Taught Experience

Survey (PTES) scores. More sophisticated measures could be developed, such as the Collegiate Learning Assessment (CLA) used in the Arum and Roksa (2011, 2014) studies, which measure critical thinking and complex reasoning⁴. Of course, the current institutional environment in the UK affords institutions very little agency in this respect. More radical ways of circumventing such regimes would involve perhaps HEIs transforming themselves into private entities. In many ways, there might be certain logic to this, as the government currently expects them to behave as if they were subject to the laws of the market but continues to subject them to ever expanding evaluative forms of governance (Ferlie et al. 2008). However, in order to salvage what is left of a public service ethos, English HEIs would be better advised to defend the notion of university education as a public good (Marginson 2011) and more aggressively seek to shape the institutional environment in ways that create greater room for the flourishing of actual teaching excellence.

The solution to the better managing of teaching (and research) is essentially twofold: relaxation of measurement in order to permit space for academic judgement, on the one hand; and, on the other hand, where metrics are used, to ensure that the metrics chosen are sophisticated enough to capture what actually matters most for students, cognitive ability. Current approaches to managing teaching and research focus on what is *easily* measured, *easily* managed and *easily* reported to external stakeholders. This is useful when calculating crude return on investment measures, but doing so avoids the more profound soul-searching that is an essential pre-requisite for the production of socially *meaningful* knowledge and cognitively adept graduates. It has been argued here that such outcomes will be dependent upon the explicit privileging of an *ethos of judgement*—which is reflected in processes of collegial social control (Martin et al. 2015) over an *ethos of measurement*—which is associated more closely with principles of market managerialism.

Acknowledgements Thanks to Mahmoud Ezzamel, Ewan Ferlie, Simon Marginson, Andrea Mennicken and Rita Samiolo for detailed comments on earlier drafts of this essay. The ideas in this paper were fermented during the time spent doing an MBA in Higher Education Management at University College London's Institute of Education between 2016 and 2018.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. Chicago, IL, US: University of Chicago Press.
- Aitkenhead, R. (2013). Peter Higgs: I wouldn't be productive enough for today's academic system The Guardian, Fri 6th December. <https://www.theguardian.com/science/2013/dec/06/peter-higgs-boson-academic-system>.
- Alvesson, M. (2012). Do we have something to say? From re-search to roi-search and back again. *Organization*, 20(1), 79–90.
- Alvesson, M., Gabriel, Y., & Paulsen, R. (2017). *Return to meaning: A social science with something to say*. Oxford: Oxford University Press.

⁴ Indeed, Arum and Roksa (2014) even suggest that student graduations might be granted on the basis of achieving adequate CLA scores rather than on simply accumulating sufficient credit hours.

- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.
- Arum, R., & Roksa, J. (2014). *Aspiring adults adrift: Tentative transitions of college graduates*. Chicago: University of Chicago Press Kindle edition.
- Bachan, R. (2015). Grade inflation in UK higher education. *Studies in Higher Education*, 42(8), 1580–1600.
- Barker, K. (2007). The UK research assessment exercise: the evolution of a national research evaluation system. *Research Evaluation*, 16(1), 3012.
- Barley, S. T. (1996). Technicians in the workplace: ethnographic evidence for bringing work into organizational studies. *Administrative Science Quarterly*, 41(3), 404–441.
- Berg, M., & Seeber, B. K. (2016). *The slow professor: Challenging the culture of speed in the academy*. Toronto: University of Toronto Press.
- Best, J. (2005). *The limits of transparency: ambiguity and the history of international finance*. Ithaca: Cornell University Press.
- Binder, A. J., Davis, D. B., & Bloom, N. (2015). Career Funnelling: how elite students learn to define and desire ‘prestigious’ jobs. *Sociology of Education*, 89(1), 20–39.
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27–41.
- Butler, N., & Spoelstra, S. (2014). The regime of excellence and the erosion of ethos in critical management studies. *British Journal of Management*, 25, 538–550.
- Collini, S. (2016). Who are the spongers now? *London Review of Books*, 38(2), 33–37.
- Collini, S. (2018). Diary: The marketisation doctrine. *London Review of Books*, 40(9), 38–39.
- Department for Business, Innovation and Skills (2016). Success as a knowledge economy: teaching excellence, social mobility and student choice, White Paper. www.gov.uk/government/publications. Accessed 30 June 2018.
- Dirk, W. P., & Gelderblom, D. (2017). Higher education policy change and the hysteresis effect: Bourdieusian analysis of transformation at the site of a post-apartheid university. *Higher Education*, 74(2), 341–355.
- Ekman, M., Lindgren, M., & Packendorff, J. (2018). Universities need leadership, academics need management: discursive tensions and voids in the deregulation of Swedish higher education legislation. *Higher Education*, 75(2), 299–321.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
- Espeland, W. N., & Sauder, M. (2016). *Engines of anxiety: Academic rankings, reputation, and accountability*. Russell Sage Foundation.
- Ferlie, E., Musselin, C., & Andresani, G. (2008). The steering of higher education systems: a public management perspective. *Higher Education*, 56, 325–348.
- Frankham, J. (2017). Employability and higher education: the follies of the ‘productivity challenge’ in the Teaching Excellence Framework. *Journal of Education Policy*, 32(5), 628–641.
- Funding Councils. (2011). *Decisions on assessing research impact*. Bristol: HEFCE.
- Gourlay, L., & Stevenson, J. (2017). Teaching excellence in higher education: critical perspectives. *Teaching in Higher Education*, 22(4), 391–395.
- Hood, C. (2006). Gaming in targetworld: the targets approach to managing British public services. *Public Administration Review*, 66(4), 515–520.
- Hunt, J. M., Langowitz, N., Rollag, K., & Hebert-Maccaro, K. (2017). Helping students make progress in their careers: an attribute analysis of effective vs ineffective student development plans. *The International Journal of Management Education*, 15(3), 397–408.
- Lezaun, J., & Muniesa, F. (2017). Twilight in the leadership playground: subrealism and the training of the business self. *Journal of Cultural Economy*, 10(3), 265–279.
- Marginson, S. (2011). Higher education and public good. *Higher Education Quarterly*, 65(4), 411–433.
- Marginson, S. (2014). UK research is getting better all the time – or is it? Guardian, 23rd December.
- Martin, G. P., Armstrong, N., Aveling, E.-L., Herbert, G., & Dixon-Woods, M. (2015). Professionalism redundant, reshaped, or reinvigorated? Realizing the “third logic” in contemporary health care. *Journal of Health and Social Behavior*, 56(3), 378–397.
- McDonald, D. (2017). *The Golden Passport: Harvard Business School, the limits of capitalism and the moral failure of the MBA elite*. New York: Harper Collins.
- Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.
- Parker, M. (2018). *Shut down the business school: What's wrong with management education*. London: Pluto Press.
- Pollock, N., D’Adderio, L., Williams, R., & Leforestier, L. (2018). Conforming or transforming? How organizations respond to multiple rankings. *Accounting, Organizations and Society*, 64, 55–68.
- Power, M. K. (2015). How accounting begins: object formation and the accretion of infrastructure. *Accounting, Organizations and Society*, 47, 43–55.
- Pusser, B., & Marginson, S. (2013). University rankings in critical perspective. *The Journal of Higher Education*, 84(4), 544–568.

- Race, P. (2017). The Teaching Excellence Framework: yet more competition – and on the wrong things! *Compass: Journal of Learning and Teaching*, 10(2).
- Sandefur, R. L. (2015). Elements of professional expertise: understanding relational and substantive expertise through lawyers' impact. *American Sociological Review*, 80(5), 909–933.
- Sennett, R. (2008). *The Craftsman*. New Haven, Conn.: Yale University Press.
- Shattock, M. (2018). Better informing the market? The Teaching Excellence Framework (TEF) in British higher education. *International Higher Education*, 92, 21–22.
- Silva, P., Lopes, B., Costa, M., Seabra, D., Melo, A. I., Brito, E., & Dias, G. P. (2016). Stairway to employment? Internships in higher education. *Higher Education*, 72(6), 703–721.
- Strathern, M. (2000). The tyranny of transparency. *British Educational Research Journal*, 26(3), 309–321.
- Taylor, A. R., & Hooley, T. (2014). Evaluating the impact of career management skills module and internship programme within a university business school. *British Journal of Guidance and Counselling*, 42(5), 487–499.
- Tourish, D., & Willmott, H. (2015). In defiance of folly: journal rankings, mindless measures and the ABS guide. *Critical Perspectives on Accounting*, 26, 37–46.
- Uttl, B., White, C. A., & Wong-Gonzalez, D. (2017). Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42.
- Warry Report. (2006). *Increasing the economic impact of research councils: Advice to the director general of science and innovation from the research council economic impact group*. London: BIS.
- Wheeler, L. (2010). *Why knowledge matters in curriculum: a social realist argument*. Abingdon: Routledge.