'*It's not what gets taught , or how well it may be taught, but who is doing the teaching* ': Can student evaluations ever deliver a fair assessment on teaching excellence in higher education?

Dr Amanda French Birmingham City University

Foregrounding the importance of 'teaching excellence' to students in the naming of the TEF is deliberate, just as in its choice of matrixes the TEF makes an unproblematic causal link between teaching excellence, student outcomes and earning. This foregrounding works because for decades now, UK education policy has been characterised by Ball's (2003) persuasive concept of 'performativity' in teaching, which has increasingly required:

> …the re-invention of [teaching] professionals themselves as units of resource whose performance and productivity must constantly be audited so that it can be enhanced' (Shore & Wright 1999, 559).

In order to audit teaching in higher education a commodified concept of 'student satisfaction' has been operationalised through student evaluations of one type of another and used in a number of ways in the current global higher education environment. For example, in addition to their much vaunted promise to deliver valuable 'big data' to inform and improve provision and facilities for students in universities, national student evaluations, like other education effectiveness matrixes, are increasingly used for commercial purposes, such as the ranking of HEIs in various national and international league tables such as the World University Rankings as well as the English Teaching Excellence Framework (TEF) (Gunn, 2018, McCaig, 2018, McGettigan, 2013).

This widespread use of student-generated data signals the extent to which student evaluations are now regarded by governments and higher education management as an authoritative source of information on all aspects of HE provision (Tomlinson, 2015; Scullion and Molesworth, 2016). Internationally, influential national student evaluations of higher education include the American 'National Survey of Student Engagement' and Australia's 'Course Experience Questionnaire'. The use of these national student evaluations in higher education is linked to the increasing attention paid to 'student voice' and 'student experience', especially in the UK, since the introduction of tuition fees. The inexorable rise of student evaluations, often carried out on a national scale, is usually justified precisely because, it is claimed they express the 'student voice' and assess the 'student experience'. However, how one defines 'student voice' (Robinson and Taylor, 2013) and how the 'student experience' (Sabri,2011) can be evaluated is problematic and highly contested.

The ubiquity of these national student evaluations reflects a tight policy focus on student engagement and outcomes expressed through a concomitant commodification and veneration of student feedback by governments and university management. A manifestation of the Global Education Reform Movement (GERM), student evaluations are driven by its three main policy principles: accountability, standards and decentralisation. Indeed, it is easy to track how over the last 30 years, in England and Wales particularly, educational data mining in the UK has been increasingly influenced by countless GERM accountability and measurement policy initiatives, culminating to date in the Teaching Excellence Framework (TEF), a key plank of the 2017 Higher Education Reform Act (HERA) (O'Leary and French, 2017). As in any neoliberal model of education, TEF positions higher education students as consumers: consequently their opinion of the 'product' they

have been 'sold', namely their degree, and the people responsible for delivering it 'at the chalk face', their teachers, require careful scrutiny (by government, by parents and by students themselves) in order that they can be judged to see if they and it have been 'value for money' (VFM) (Naidoo, Shankar, and Veer 2011).

While the TEF has only been in place since 2017 it is clear that the UK government is interested to see the extent to which prospective students already regard TEF ratings as positive or negative market indicators. The survey carried out annually by the Universities and Colleges Admissions service (UCAS, 2018) for prospective students included for the first time a number of questions about the TEF. Out of the half a million applicants who had applied by the 15 January deadline, 85,000 responded to the new TEF questions, a response rate of more than 15 per cent. Out of that number 17.1% stated that they 'had heard of the TEF' prior to making their applications before UCAS's January deadline. 58% of that group stated that the TEF awards were 'important' or 'extremely important' when deciding which HEI to apply to, and that they were more likely to apply to an institution with a TEF Gold rating. Similarly, three out of every five respondents who did not know what the TEF was, said they would have found the awards 'important', or 'extremely important', had they known about them.

This suggests that, as more students become aware of the TEF, albeit as rather vaguely defined proxy for teaching quality, it will function in exactly the way that Government hoped it would, as a market indicator. It is perhaps predictable that the UCAS report makes no attempt to find out if those potential applicants who had heard of the TEF actually knew anything about how it had been out together or how the ratings had been achieved. A proxy, after all, works because it does not invite or require close scrutiny of its self, rather it stands for/stands in for something else, usually in order for the something else to be measured or compared. The TEF it is clearly being positioned unproblematically by UCAS in the study above as a proxy for assessing the quality of teaching prospective students can expect to receive at any given HEI. Moreover, of interest to the focus of this chapter, the use of student evaluations, is the fact that within the TEF itself the NSS is being used as one of the supporting sources of information about teaching excellence that give the TEF validity in the eyes of students.

Students, in addition to the national student surveys described above, are also exhorted throughout their time at university, to complete endless in-house pre-module, mid-module evaluations and overall course evaluations. A recent UCU report exploring the impact and implications of the TEF for lecturers in UK HEIs (O'Leary, Cui and French, 2018,) reported that a significant number of teaching staff in UK HEIs had experienced an increase in what they called 'monitoring mechanisms and accountability procedures' linked to student feedback of one kind or another, since the implementation of the TEF (although they did not always directly link them to the TEF). These included an increase in standardised in-house module and programme evaluations, which undoubtedly facilitated a corresponding rise in the use of learning analytics to assess provision.

In-house module evaluations in UK universities, like their US equivalents, Student Evaluations of Teaching (SETs), are, like their large national counterparts necessarily rather blunt instruments that do little to capture the complexity of teaching and learning interaction. Hornstein (2017) makes the important point that SETs were intended, when first introduced in the 1970's, to help improve and shape the quality of teaching. However, they subsequently became the 'primary indicator for

summative evaluation, that is, to "sum up" overall performance to decide about promotion, and tenure'. In other words, SETs have been allowed to become the dominant indicator of teaching competence in US higher education. They are designed to elicit quantitative and qualitative information, the former usually through 1–5 Likert scale or multiple choice questions, supplemented by student comments through an open text box, which provides the latter. Despite their limitations all student evaluations, to a greater or lesser extent are invested with a rather specious authority which endows the 'student voice' and the 'student experience' with authenticity, validity and reliability, when, as discussed below they are arguably quite suspect. Indeed, the fragility of this assumed authority can be seen in the extent to which most contemporary research on student evaluation focusses on the need to improve student survey design methods (Richardson, 2005; Sulis, Porcu and Capursi, 2019). Nonetheless, the assumption of a 'scientific' or empirical veneer for student evaluations, both national and in-house, is necessary because they are so frequently mobilised by senior HEI management and governments, not only to justify specific internal strategic plans and processes (such as promotion, redundancy and the development or running down of different departments), but to respond to broad policy edicts like the TEF.

In addition to the difficulties in actually measuring the quality of teaching discussed elsewhere in this collection, there is a body of research exploring how student perceptions of teachers' teaching excellence, or otherwise, in higher education play out very differently depending on the gender, age and social class lecturers doing the teaching. This chapter argues that these differences make it difficult to ensure that students' assessment of higher education teaching are fair and/or consistent with regard to the teaching they are experiencing across different courses, disciplines and institutions. In particular, this issue of fairness and consistency has implications for teaching staff in higher education who deviate from the stereotype of the white, straight, middle-class male university lecturer. Indeed, it appears likely that student assumptions and prejudices about gender, race, disability and sexuality, expressed through their evaluation of the teaching they receive in higher education, can constitute a form of discrimination facing many lecturers.

Moreover this is a form of discrimination which largely goes unrecognised but is extremely potentially damaging when such evaluations are mobilised for evaluative purposes, as they are in the TEF as the National Student Survey (NSS) is one of the matrixes used in the formulation of the TEF rating for each English HEI. Many studies in the USA have shown that SETS are biased against black and female lecturers (Abrami, d'Apollonia, & Rosenfield, 2007; Benton & Cashin, 2014). Unfortunately, very little work has been done to see if the same could be said about national and in house evaluations in UK HEIs. Despite their shortcomings, the material generated from the NSS and in-house student feedback is currently used to inform, not only internal quality processes but contributes to the 'additional contextual information' provided by HEIs as part of an institution's TEF submission. (They will also be one of the primary sources of information for the forthcoming subject-based TEFs). In short the consequences of failure to score well on student evaluations of teaching quality, remain high.

In the light of these concerns this chapter, therefore, seeks to critically interrogate the rather doubtful claims often made about student evaluations, in particular, the claim that they are a reliable arbiter of 'value for money' (VFM) in higher education, due to the essentially experiential and subjective nature of student feedback on their teaching experiences at university, which rarely gets

acknowledged even when research has explored the extent to which it is affected by other factors such as the gender or race of the lecturer (Richardson 2005; Darwin 2016, Benton and Cashin 2014). This is not to say that student-generated data could not be harnessed more effectively to engage students in meaningful dialogue about the quality of their various experiences in higher education. This is something that will be explored in the final section of this chapter. However, Tomlinson's (2015) study into student perceptions of themselves as consumers highlights how HEIs mostly fail to understand how, in a highly marketised higher education sector, they actually make sense of students' complex relationships with their university (and the staff, professional and academic who work in them). Moreover, there is evidence to suggest that in line with wider neoliberal conceptions of individualism, competition and the VFM agenda in higher education, that HEI's have been more interested in casting students in the role of informed and rational consumers of a 'university experience', who have exercised choice over the private investment they have made in their own education by signing up for a degree. This, of course, is arguably more straightforward than asking them to grapple with the complexities of teaching and learning; the vagaries of the labour market for graduate employment across different disciplines and the inherent inequalities of the higher education sector in its current heirarchised form.

Meanwhile students, via these various surveys and evaluations have endless opportunities to award plaudits or express disappointment on everything they experience whilst at university, from the price and availability of the parking on campus and the quality of the food in the canteens, to how well they feel they have been prepared for the world of work and taught their subject. A corollary of the 'the customer knows best' ethos, this aspect of marketisation in HE encourages the illusion of a kind of democratisation of accountability to students-as-customers (perhaps best encapsulated by the vacuous sloganeering of, '*you said, we did*'). However, whilst accepting that student surveys can usefully highlight levels of student discontent regarding, say, the library's inadequate resources or poorly ventilated classrooms, the legitimacy of students' perceptions about the quality of teaching they experience in HE should be treated with caution for as outlined above.

Student evaluations about the quality and efficacy of the teaching they received whilst at university, as opposed to the question of whether they liked or disliked their tutors or enjoyed some teaching sessions more than others are inherently problematic. To take them seriously in their current form, as a way of understanding how to improve higher education teaching assumes, with very little empirical evidence, that students possess a critical understanding of how and why they are taught in certain ways (Nast, 1999). Moreover, students' conceptions about teaching are essentially subjective, which, to be fair, is not surprising as they can only be based on individuals' usually limited experience of higher education teaching (after all most of them pass through HE only once as undergraduates).

Packaging student voice and experience through neatly delineated educational effectiveness evaluations like the NSS therefore, belies the actual complexity involved in teaching and learning interactions. As Wood (2017) states

> …teaching [is] emergent, multifaceted and contextually based. It refutes notions of 'best practice' and argues that any attempt to capture 'excellent practice' is to reduce the holistic

> nature of the processes that bring teaching, learning, curriculum and assessment together. (pg. 40)

As suggested above however, it appears that governments find it easier to rely on often superficial student evaluations of teaching rather than spend time and money exploring it with teachers and students given that identifying and measuring indices of teaching excellence or quality are so notoriously difficult to determine (O'Leary and Wood, 2019). Nor do they acknowledge or explore any wider social and cultural factors that might affect students' perceptions of teaching such as racism, sexism and homophobia, issues that will be explored in the next section of this chapter. For these reasons it can be argued that if such perceptions remain unexamined, the credence given to students' evaluation of their teaching in higher education by government and management will inevitably have unpredictable and unfair consequences for many HE staff engaged in teaching.

For example a number of US studies have revealed that low scores in student evaluations can be used to initiate disciplinary action against higher education teachers often affecting any future chances of promotion (Davis, 2009; Centra and Gaubatz, 2000; Young, Rush, and Shaw, 2009). A recent report in the UK, commissioned by the Union for Colleges and Universities (UCU) in (2018) *Understanding, recognising and rewarding teaching quality in higher education: an exploration of the impact and implications of the Teaching Excellence Framework* (O'Leary, Cui, and French) was the first to explore the views of large numbers of staff working in HEIs about the TEF and its focus on teaching excellence. Unfortunately, many of its respondents reported similar punitive responses by management to low NSS and in-house student evaluations, such as courses being put into 'special measures' which included staff being given targets to achieve higher NSS scores next time round. This kind of knee-jerk response to poor student evaluations, if allowed to go unchallenged, conjures up dystopian image of a performativity/audit culture where higher education teaching staff are principally viewed as units of resource whose academic labour is only measured in terms of the economic needs of their employers (How much do you cost to employ? How many students can you teach?) and their customers, the students (What am I paying for and am I getting my money's worth?)

All of which is not to say that it is not important for HEIs to monitor and quality assure their teaching provision. Issues such as the transparency and consistency of assessment processes, the relevance and currency of curriculum content and efficacy and accessibility of teachers are live issues, they do matter to teaching staff and students should be able to comment on and ultimately inform. However, as discussed above, because tuition fees position students as 'consumers' or 'customers' universities a tendency to frame such pedagogic issues as 'customer satisfaction issues' . This has resulted in an institutional nervousness about managing student complaints about assessment, curricular and teaching, especially when they are couched in terms of the students asserting that they did not get the grade they 'wanted' or 'expected' or 'had paid for'. However, despite the well-researched complexity of teaching and learning interactions and outcomes there is a tendency for poor grades to be blamed on poor teaching.

Such a narrow VFM agenda can easily end up treating teaching staff as if they were all the same sort of widgets whose continued employment relies on them successfully meeting a series of allegedly neutral performance indicators such as how many of your students achieve 'good' degrees' and how much are they earning 5 years after they graduate, (both TEF benchmarks), as well as student evaluations which ask then to rate more overtly subjective questions such as 'Staff are good at explaining things' and 'Staff have made the subject interesting' (both NSS questions). Such a working environment, hedged about, as it so often is, with short-term fixed contracts and performance-related pay, encourages compliance and reduces the willingness of teaching staff in higher education to challenge negative comments about their management of students and/or their delivery of their subject. Over time, student evaluations, like other education effectiveness matrixes surveys increasingly have, therefore, the potential to be used by management as tools to insist on more centralised control of teaching processes and practices. This can and has resulted in the forced removal of unpopular programmes, assessment models and delivery styles as well as the marginalisation of non-compliant teaching staff; all of which ultimately limits academic freedom and professional autonomy within the higher education sector. Simultaneously such centralised ~~also~~ control inhibits the development of innovative and creative learning which has to be to the detriment of students, irrespective of their discipline.

In antithesis to the VFM agenda this chapter argues that an evaluation of the quality of teaching and the learning experience that it produces are mutually constituted through the relationship between the lecturer and the student(s) involved. However, the idea that all such teaching and learning relationships are equally constituted in the same way (and be evaluated/understood using the same instrument of measurement), is problematic. In order to begin to rethink it one can evoke Stuart Hall's (1997) 'articulation of certain kinds of differences' to account for how any mutual constitution of the quality of teaching and learning will inevitably mediated by certain kinds of inequalities. Acknowledging, therefore, that difference and inequalities will inevitably play a part in any evaluative process is a more productive way of thinking about how more informed indices of teaching quality might be more usefully understood and operationalised in higher education. This approach, however, requires an exploration of how existing racialised patterns and sexualised and gendered patterns reoccur, and sustain inequalities currently in the UK higher education sector.

To try and frame its critical interrogation of student evaluations this chapter treats higher education teaching staff as 'embodied subjects' (Shilling, 2005) working in highly contested political spaces where they are positioned in distinct ways as providers of teaching to students. Examples of research that help explore how and why students evaluations discriminate against certain groups of teaching staff in higher education will be explored and discussed, in the next section, using the concept of intersectionality as an over-arching theory in recognition of the disadvantage created by membership of other groups in society such as disability, LGBTQ+, age or religion. Intersectionality is a useful concept because it insists on how location (time and space) work together with personal factors such as, race, class, and sexuality and gender, to constitute lecturers as embodied subjects within certain spaces, such as higher education. An intersectional approach to how and why students perceive and judge higher education teaching staff differently, acknowledges that there is more than one type of disadvantage – and advantage – which exist, often simultaneously.

In this way, intersectionality helps to explore how lecturers and students are all always co-located in multiple ways within instances of teaching and learning allowing for critical interrogation of how student generated data, plays out differently for different higher education lecturers. Specifically, it helps identity how the different categories to which higher education lecturers belong reflect the different power relations that characterise their teaching and learning interactions, meaning that teaching staff and students are never located in a singular power position defined by each other's race, gender, sexuality etc. As a way of trying to understand why student evaluations are so potentially discriminatory this chapter will also draw on feminist and Critical Race Theory (CRT), along with queer and disabled critical theories to interrogate the values and processes underpinning educational effectiveness' matrixes based on student evaluations. In doing so it points to a way to possibly rehabilitate student evaluations through deconstructing and challenging the normative and discriminatory values that they often reveal but do not acknowledge.

In the USA there is plenty of evidence to suggest that students' evaluation of the teaching they experience in higher education is more likely to reflect their biases, prejudices towards those that teach them (Abrami, d'Apollonia, & Rosenfield, 2007; Benton & Cashin, 2014 2010; Nast, 1999), as well as any misconceptions about teaching per se that they may harbour (Steward and Phelps, 2000). For example, there is a tradition of research going back several decades, that suggests that US higher education students perceive, evaluate, and treat female lecturers quite differently than they do their male counterparts. This suggest that students persistently reproduce gendered responses, which reflect wider social and cultural sexual stereotyping rather than informed responses to their teaching and learning experiences when asked to evaluate their teachers' performance (Basow, 1995; Centra & Gaubatz, 2000; Feldman, 1992; Young, Rush, & Shaw, 2009). Research conducted at the University of California, for example, found that female staff consistently received lower scores on student evaluations of teaching (Boring, Ottoboni and Stark 2016). Likewise, a study in Canada found that female tutors were more likely to be judged harshly than male tutors in student evaluations when they did not give higher marks. Indeed, students receiving lower marks often made reference to negative female gender stereotypes with regard to the female teachers who had marked their work (Sinclair and Zunda 2000).

While a general consensus exists that gender plays a role in how students perceive and interact with their lecturers it is difficult to work out how this is also affected by variations in several mediating factors such as disciplinary fields/ teaching styles and subject material. While it is inevitably problematic to isolate gender as a source for student bias after face to face teaching (as there are so many other variations that might affect a student's perception of the learning experience), one study sought to see if it could be discerned when students were taught online where it was possible to manipulated the gender of the online tutors teaching them. MacNell, Driscoll and Hunt's (2014) USA study created two assistant instructors in an online class who each operated under two different gender identities. In their evaluations students consistently rated the assumed male instructor significantly more positively than the assumed female counterpart, regardless of their actual gender. Interestingly, the degree of gender bias evinced by the students varied depending on other factors such as student's gender, their age and the area and discipline they were working in, factors that in themselves warrant further investigation as they too could be gendered or affected by other variables. The effect of gender of student evaluations of teaching is also apparent in the choice of language deployed in responses to

open-text questions by students in teaching evaluations. In MacNeill's et al (2014) students used language such as "*brilliant, awesome and knowledgeable*", about the 'male' online instructor but when the same teaching experience was thought to come from a woman, 'she' was categorised as "*bossy and annoying*"'.

In studies focusing on the reactions of students to black teaching staff researchers have found similar prejudices surfacing in student evaluations of teaching. Research conducted in the UK (Bell and Brooks 2018) found that by cross-referencing NSS scores with statistics on staff demographics from the Higher Education Statistics Agency (HESA), they could show that the ethnicity of lecturers had a significant impact on NSS scores, with black, Asian and minority ethnic (BAME) lecturers consistently being scored lower than their white counterparts. This work echoed research in the USA, which found that black and minority ethnic teachers tended to be evaluated more harshly on websites such as RateMyProfessor.com (Close Subtirelu, 2015). The chapter 'Rapport and Relationships: The Student Perspective on Teaching Excellenc' (Lawrence, Hunt, Shaw, and Synmioe), in this collection also raises the question of how different groups of students, in this instance, widening participation may have different expectations and assumptions about teaching, which again are not acknowledged in current student evaluations.

Ahmed (2004) has argued that academic authority does not seem to 'stick' to female or black academics in the same way as it does to their white counterparts, rendering them at a distinct disadvantage when it comes to asking students to measure their quality and effectiveness as higher education lecturers through matrixes like NSS which contribute to institutional TEF rating. Judgements about what a higher education lecturers look like, sound like, act like can be linked to the students have of ethnic minorities generally. Puwar (2004) likewise makes the point that

> […] some bodies through their history are seen as natural occupants of specific spaces, some bodies have the right to belong in certain locations, while others are marked out as trespasser's, and therefore out of place (p.51).

Trotman (2009) examines the impact of what she calls the'' imposter phenomenon' to describe how black women have been doubly victimised by scholarly neglect and racist assumptions which often leaves them feeling exposed and vulnerable in higher education. Leach (2007) succinctly terms the position of lesbian, gay bisexual and transgender workers as "space invaders", which perhaps goes some way to explaining why students often view them and their academic labour in more negative terms than their white and heterosexual colleagues. One also needs to take into consideration how lecturers who do not fit the norm feel themselves about their position in higher education. Research by Wright, Thompson and Channer, (2007), focuses on British universities and examines the experiences and challenges faced by Black women in academia, which often centre around their acute awareness that they occupy a space traditionally reserved for their white middle class counterparts. Finally, Rollock |(2019), Arday (2015) and Bhopal (2016) have all done more recent valuable research on how BAME academics have consistently been subject to racial discrimination and increased precarity in higher education by management.

In the light of the above research findings, not surprisingly perhaps, many staff in higher education are increasingly fearful of student evaluations, especially as the TEF encourages such an unhelpful

focus on the 'performance of teaching' (Ball, 2012) rather than any meaningful understanding of the quality of the teaching and learning interactions that students are experiencing with different teachers. As Ball writes:

> […]Within the rigours and disciplines of performativity we are required to spend increasing amounts of our time in making ourselves accountable, reporting on what we do rather than doing it. There are new sets of skills to be acquired here: skills of presentation and of inflation, making the most of ourselves, making a spectacle of ourselves. (2012, 30)

These new skills of 'presentation' of enacting the performance of excellent teaching are clearly more difficult to achieve when the 'actors' enacting them do not, as discussed previously, 'look the part' to begin with. Without being able to critically investigate at the different ways in which different groups of staff across higher education are perceived we are missing important information on how students' evaluation of their experience in HE plays out differently for those various groups.

This chapter does not seek reject the usefulness of student evaluations in any circumstances rather it is asking for a more nuanced approach to their use, both in terms of how they are designed and how they are used to inform policy and practice in higher education. In the first instance that there a case for saying that universities need to invest more time in informing and preparing students for student evaluations. Currently there is no evidence to show that HEIs are working to help students think critically about teaching quality prior to being asked to judge their lecturers. (In Athena Swan/ Stonewall for example there is the expectation/ recommendation that senior staff involved in recruitments /conferment to be engaged in unconscious bias training – in such requirement attends student engagement in NSS and other in-house valuations (as discussed in Kate's Thomas Preface to this collection).

In terms of using student evaluations as a fair means of assessing quality of teaching universities need to acknowledge the impact of staff diversity on students' judgements. This is especially significant as the higher education teaching workforce is not very diverse at all. Indeed until the last century women, BAME, disabled and LTBG+ people have politically, historically and socially have been excluded and marginalised both within and outside of the traditionally very privileged space that universities occupy in English society. Therefore, there needs to be active consideration, and more importantly, more research into how students' own educational background, as well as factors such as their gender and ethnicity do affect their perceptions and judgement on lecturers' teaching. Such research could form part of TEF's 'lessons learned exercise' as refinements to the process are made for future years. There is an opportunity to develop a TEF that addresses myriad forms of potentially 'othered' staff identities for staff as well as students.

In conclusion, this chapter has argued that the TEF's use of both national and in-house evaluative data from students' needs to be more attuned to the limitations and shortcomings of student evaluations as a way of measuring and comparing 'teaching excellence'. It has argued for a more nuanced understanding of how students experience their teaching and learning interactions with lecturers, that is fluid and allow for complexity and a recognition that teaching and learning is necessarily experienced very differently across the very diverse institutions disciplines and courses that characterise higher education provision. These difference moreover need, if they are to be

meaningful, to be judged on their own terms, and in relation to similar forms of provision. In particular there needs to be a sector-wide commitment to challenging the extent to which, as the research cited in the chapter suggests, student conceptualisations of teaching quality are often unconsciously and consciously entrenched white/male/heteronormative/able-bodied lecturer archetypes which create a powerful lens through which they view teaching competence and confidence.

Likewise, the wider social impact of homophobia, disablism sexism and racism need to be acknowledged at a structural, institutional level, with staff and students, as part of their everyday teaching and learning interactions. However, although diversity initiatives can reveal 'the gap between words, images and deeds' (Ahmed, 2007, p.607), there is a need to ensure that they lead to action (Pilkington, 2013). For example, there could be a commitment by Government to experiment with more sophisticated, sensitive evaluative instruments and institutional strategies that could help critique and explore the diversity of learning processes and pedagogic approaches in contemporary higher education which recognise the centrality and complexity of the emotions and practical skills involved in learning and teaching. This goes beyond a commitment from managers to support all staff who take curricular and pedagogical risks, but to also recognise that some staff are more vulnerable and/or precariously positioned within the Academy, irrespective of how they teach.

The current focus in the TEF on increasingly standardised performance outputs, often far removed from the demands of everyday teaching, for teaching staff does nothing to encourage a productive climate of cooperation and collaboration around teaching and teaching development. Instead, to date, the TEF appears to have further legitimised an already established tendency by the executive towards micro-management and endless target-setting. Furthermore, set against a higher education environment beset with threats about reduced provision, cuts in funding and resultant employment insecurity, teaching staff feel under pressure to meet targets and comply with institutional edicts around curriculum and assessment. Consequently, as Holley *et al.* (2006) found in their research about widening participation students, that

> . . . [lecturer] strategies for responding to students are not governed by student need – but by management and government targets – and in the end, the students are further unsupported and silenced. (Holley *et al*. 2006, p. 4)

This then is the final reason that we need to rethink student evaluations and, in particular, the role they play in the TEF and in-house institutional policy-making. It is that it is questionable that they actually, in their current form, do allow a diversity of student opinions and experiences about the teaching that they receive in higher education, both negative and positive, to be heard and acted on in any meaningful way. By engaging in a more multifaceted, nuanced dialogue with students, with teaching staff will be more above to share and explore existing effective teaching practices and develop new more inclusive and creative ways of learning in higher education. Essentially, "actions speak louder than words" and until new approaches to student evaluations become part of a comprehensive strategy for equality, diversity and inclusivity owned and driven by policy-makers and senior managers across higher education nothing will change (Husbands, 2019).

References

Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry & J. C. Smart (Eds.), The scholarship of teaching and learning in higher education: An evidence-based perspective (pp. 385–445). Dordrecht, The Netherlands: Springer.

Ahmed, S. (2007). A phenomenology of whiteness. Feminist Theory, 8(2), 149–168.).

Alexander, C. and Arday, J. (eds) (2015) Aiming Higher: Race, Inequality and Diversity in the Academy, London: Runnymede

Ball, S.J. (2012) The making of a neoliberal academic, Research in secondary teacher education, 2, (1) pp. 29–31.

Ball, S.J. (2001) 'Performativities and fabrications in the education economy: towards the performative society'. In D. Gleeson and C. Husbands (eds.) The performing school: managing teaching and learning in a performance culture, 3- 13. London: RoutledgeFalmer.

Ball, S. (2003) The teacher's soul and the terrors of performativity, Journal of Education Policy, 18 (2) pp.215-228.

Basow, S. A. (1995). Student evaluations of college professors: When gender matters. Journal of Educational Psychology, 87, 656–665.

Bell, A. and Brooks, C. (2016). Is There a Magic Link between Research Activity, Professional Teaching Qualifications and Student Satisfaction? Social Science Research Network June 2016). Accessed on December 12th 2019 at : http://ssrn.com/abstract=2712412

Benton, S. L., & Cashin, W. E. (2014). Student ratings of instruction in college and university courses. In Michael B. Paulsen (Ed.), Higher Education: Handbook of Theory & Research, 29, pp. 279-326. Dordrecht, The Netherlands: Springer

Bolat, E. and O'Sullivan, H., (2017) Radicalising the marketing of higher education: learning from student-generated social media data. Journal of Marketing Management, 33 (9-10), pp. 742-763.

Bhopal, K. (2016) The experiences of Black and minority ethnic academics: a comparative study of the unequal academy. London and New York: Routledge.

Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. ScienceOpen Research 7 (January 2016). Accessed on November 16th 2019 at SOR: https://www.scienceopen.com/document?vid=818d8ec0-5908-47d8-86b4-5dc38f04b23e

Butler, J. (2004) Undoing gender. New York & London: Routledge.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? Journal of Higher Education, 71, pp. 17–33.

Close Subtirelu, N., 2015. '"She does have an accent but…": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com'. Language and Society 44 (1), pp. 35-62.

Collini, S. (2011) From Robbins to McKinsey, London Review of Books, 33 (16) pp. 9–14.

Crenshaw, K. W. (2011) 'Twenty Years of Critical Race Theory: Looking back to Move Forward', Connecticut Law Review, 43(5), pp. 1253–1354

Darwin, S. (2016) The emergence of contesting motives for student feedback-based evaluation in Australian higher education, Higher Education Research & Development, (35) 3, p. 419-432.

Department for Education (DfE). 2017. Teaching Excellence and Student Outcomes Framework Specification. Birmingham: Department for Education (DfE). Accessed December 2018. https://assets.publishing.service.gov.uk/government/ uploads/system/uploads/attachment_data/file/650179/Teaching_Excellence_and_Student_Outcomes_Framework_ specification.pdf.

Delgado, R. and Stefancic, J. (2001) Critical Race Theory: An Introduction. New York: New York University Press.

Feldman, K. A. (1992). College students' views of male and female college teachers: Evidence from the social laboratory and experiments – Part 1. Research in Higher Education, 33, 317–375.

Gillborn, D. (2006) 'Critical race theory beyond North America: Toward a trans-Atlantic dialogue on racism and antiracism in educational theory and praxis', in Dixson, A. D. and Rosseau, C. K. (eds) Critical race theory: All God's children got a song. New York, NY: Routledge, pp. 241–265.

Gee, J. (1999) 'New people in new worlds: networks, the new capitalism and schools'. In B. Cope and M. Kalantzis (eds.) Multiliteracies: literacy learning and the design of social futures, pp. 43–68. London: Routledge.

Gunn, A. 2018. "Metrics and Methodologies for Measuring Teaching Quality in Higher Education: Developing the Teaching Excellence Framework (TEF)." Educational Review 70 (2): 129–48. doi:10.1080/00131911.2017.1410106.

Hall, S. (1997). Representation: cultural representations and signifying practices. London: Sage.

Holley, D., Sinfield, S. and Burns, T., 2006. "It was horrid, very very horrid": a student perspective on coming to an inner city university in the UK. Social Responsibility Journal, 2 (1), 36-41.

Hornstein, H. (2017) Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance, Cogent Education, (4) 1 (published online)

Leach, T. (2007), "Space Invaders: Race, Gender and Bodies Out of Place", Equal Opportunities International, Vol. 26 No. 5, pp. 507-514.

McCaig, C., ed. 2018. "Continuity and Discontinuity on the Road to Risk and Exit: Stages of Marketisation in Comparative Policy Analysis." Chap 4 in The Marketisation of English Higher Education (Great Debates in Higher Education). Bingley: Emerald Publishing Limited.

McGettigan, A. 2013. The Great University Gamble: Money, Markets and the Future of Higher Education. London: Pluto Press.

MacNell, L., Driscoll, A. & Hunt, A.N. (2015) 'What's in a Name': Exposing Gender Bias in Student Ratings of Teaching, Innovative Higher Education, 40 (4) pp 291–303

Molesworth, M., Scullion, R., & Nixon, E. (Eds.). (2011). The marketisation of higher education and the student consumer. Abingdon: Routledge.

Naidoo, R, Shankar, A & Veer, E 2011, 'The consumerist turn in higher education: Policy aspirations and outcomes', Journal of marketing management, 27 (11-12) pp. 1142-1162.

Nast, H. (1999) 'Sex', 'Race' and Multiculturalism: Critical consumption and the politics of course evaluations, Journal of Geography in Higher Education, 23 (1), pp. 102-115

O'Leary, M., V. Cui, and A. French. 2019. Understanding, Recognising and Rewarding Teaching Quality in Higher Education: An Exploration of the Impact and Implications of the Teaching Excellence Framework. Birmingham: University and Colleges Union. Accessed July 2019. https://www.ucu.org.uk/media/10092/Impact-of-TEF-report-Feb-2019/pdf/ ImpactofTEFreportFEb2019.

O'Leary, M., and A. French. 2017. Excellence in Higher Education, Challenges, Changes and the Teaching Excellence Framework. Bingley: Emerald Publishing Limited.

Pilkington, A. (2013) The interacting dynamics of institutional racism in higher education. Race, Ethnicity and Education. 16(2), pp. 225-245. 1361-3324.

Richardson, J. T. E. (2005) Instruments for obtaining student feedback: a review of the literature, Assessment & Evaluation in Higher Education, 30 (4) pp. 387-415

Sabri, D. (2011) What's wrong with 'the student experience'? Discourse: Studies in the Cultural Politics of Education, 32(5), pp. 657-667.

Shilling, C. (2005). The Body in Culture, Technology and Society. London, Thousand Oaks, New Delhi.: Sage Press

Scullion, R. & Molesworth, M. (2016) Normalisation of and resistance to consumer behaviour in higher education, Journal of Marketing for Higher Education, 26 (2) pp. 129-131

Shore, C. & S. Wright (1999) 'Audit culture and anthropology: neo-liberalism in British higher education'. The Journal of the Royal Anthropological Institute, 5(4) pp. 557–75

Su, F., and M. Wood. 2012. What Makes a Good University Lecturer: Students' Perceptions of Teaching Excellence. Journal of Applied Research in Higher Education, 4 (2) pp. 142–55.

Sulis, I., Porcu, M. & Capursi, V. (2019), On the Use of Student Evaluation of Teaching: A Longitudinal Analysis Combining Measurement Issues and Implications of the Exercise, Social Indicators Research, 142, (3), pp 1305–1331

Robinson, C. and Taylor, C. (2013) Student voice as a contested practice: Power and participation in two student voice projects, Improving Schools, 16 (1), pp. 32-46.

Rollock, N. 2019. Staying power: the career experiences and strategies of UK Black female Professors. Project Report. UCU, London. Accessed January 2020https://www.ucu.org.uk/media/10075/staying-power/pdf/ucu_rollock_february_2019.pdf

Tomlinson, M. (2015). Student perceptions of themselves as 'consumers' of higher education. British Journal of Sociology of Education, 38 (4) pp. 1-15.

Trotman, F.K. (2009). The imposter phenomenon among african american women in the U.S. institutions of higher education: Implications for counseling. Compelling Counseling. Interventions. Accessed December 2019 at http://counselingoutfitters.com/vistas/vistas09/Article_8_Trotman.pdf

UCAS (2018) Teaching Excellence and Student Outcomes Framework (TEF) and demand for full-time undergraduate higher education.

Wood, P. & O'Leary, M. (2019) 'Moving beyond teaching excellence: Developing a different narrative for England's higher education sector', International Journal of Comparative Education and Development, 21 (2).pp. 112-126.

Wood, P. (2017) 'From Teaching Excellence to Emergent Pedagogies: A Complex Process Alternative to Understanding the Role of Teaching in Higher Education.' In A. French and M. O'Leary (eds.), Teaching Excellence in Higher Education: Challenges, Changes and the Teaching Excellence Framework. Bingley: Emerald Publishing Ltd. pp 39-74.

Wright, C., Thompson, S. & Channer, Y. (2007) Out of Place: Black women academics in British universities, Women's History Review, 16 (2), pp. 145-162,

Young, S., Rush, L., & Shaw, D. (2009). Evaluating gender bias in ratings of university instructors' teaching effectiveness. International Journal of Scholarship of Teaching and Learning, 3, 1–14.

Sinclair, L., and Zunda, K., 2000. 'Motivated stereotyping of women: she's fine if she praised me but incompetent if she criticised me'. Personality and Social Psychology Bulletin 26 (11), 1329-1342.