

## The Teaching Excellence Framework (TEF): A snapshot review of relevant literature and current developments incorporating learning gain

Professor John G. Sharp<sup>1</sup> and Dr Emma Young<sup>2</sup>

<sup>1</sup> Lincoln Academy of Learning and Teaching (LALT), University of Lincoln

<sup>2</sup> Directorate of Learning, Teaching and Student Experience (LTSE), University of Bradford

### Abstract

*This article presents an overview of developments leading up to and following the introduction of the Teaching Excellence Framework (TEF), with a particular focus on the most recent iteration of TEF and the notion of learning gain following consultation in 2022. In the face of widespread uncertainty and critique surrounding the framework, the challenges presented to many UK higher education institutions, for whom the adoption of and compliance with still complex and contested TEF requirements is all but obligatory, are considered in detail. With attention on learning gain within the framework including, but not limited to, academic and personal development and work readiness, the actual measurement of learning gain directly or by proxy is, however, far from unproblematic and can, in fact, be undertaken in different ways by different means to suit different purposes, as can the measurement of a broader and high-quality student experience itself. Entry-level reading, and an essential starting point for those interested in the development of learning gain metrics for TEF, those involved in planning and the determination of a strategic and operational response to future TEF exercises and those involved in the establishment of TEF policy in detail, includes the early but comprehensive Rand report published by McGrath et al. (2015), the review of learning gain provided by Evans et al. (2018) and the evaluation and final report of the HEFCE/OfS learning gain pilot projects published by Kandico-Howson (2019).*

*Keywords: Teaching Excellence Framework (TEF), Learning Gain, Learning Gain Metrics, Student Experience, Office for Students*

### Introduction and background

The recently renamed Teaching Excellence Framework (TEF), replacing the Teaching Excellence and Student Outcomes Framework as it was fully and formerly known, remains a controversial, sector-wide exercise designed to assess the quality of the undergraduate student experience in participating higher education institutions across the UK, bringing with it important reputational and other consequences

depending on the devolved funding and regulatory systems operating in each home nation state. Primarily involving institutions in England and driven by the Conservative government of the time's concern at research being privileged over what it saw as 'lamentable teaching' driven in no small measure by the Research Assessment Exercise (RAE), now the Research Excellence Framework (REF), first introduced in 1986. While the notion of a TEF first appeared in the government's Green and White papers of 2015 and 2016: *Success as a knowledge economy: Teaching excellence, social mobility and student choice*, the TEF was later cemented into the *Higher Education and Research Act of 2017* alongside provision for the establishment of the Office for Students (OfS) as the new independent regulator and competition authority replacing the Higher Education Funding Council for England (HEFCE) and the Office for Fair Access (OFFA).

At an unprecedented time of ongoing interventionist higher education evolution, the original intention and stated purpose of the TEF was to:

- Better inform students' choices about what and where to study
- Raise esteem for teaching
- Recognise and reward excellent teaching
- Better meet the needs of employers, business, industry and the professions

Importantly, the proposed TEF was also linked to the funding of teaching in higher education through the tuition fees an institution could charge.

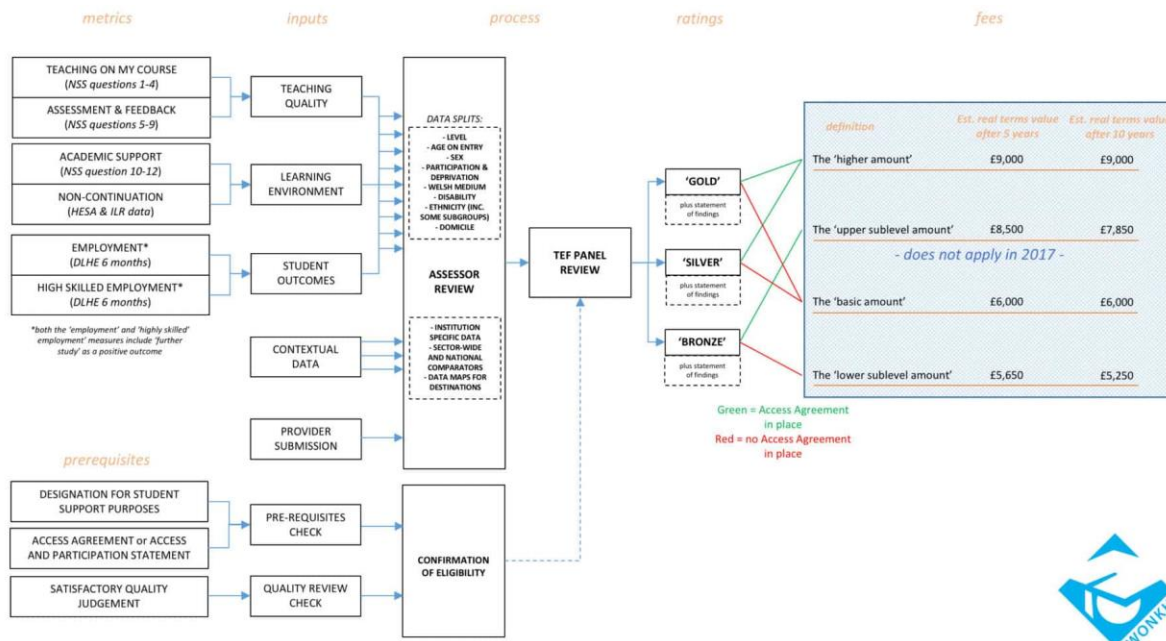
To date, and following an initial consultation and pilot (Year 1 or TEF1), there have been three major rounds of TEF assessment at provider-level, with subject-level pilots incorporated later to run in parallel: 2016-17 (Year 2 or TEF2: Provider), 2017-18 (Year 3 or TEF3: Provider and initial subject pilot) and 2018-19 (Year 4 or TEF4: Provider and additional subject pilot). In essence, and as indicated, each was intended to provide a judgement of relative performance through the establishment of a data benchmarking process drawing on a small number of identified student intake metrics and pre-existing quality assurance measures (e.g. selected National Student Survey results, retention/non-continuation rates from HESA, and leaver destinations, skilled employment and the proportion of students in further study from the Destination of Leavers from Higher Education Survey). These metrics were then supplemented with provider statements (e.g. additional qualitative/quantitative narratives and evidence) to form the basis of the assessment process. This was completed using appointed TEF assessors confirmed by an appointed TEF panel including academics, students and employers.

In even early iterations of TEF, outcomes were grouped into a small number of performance-related 'awards' or ratings:

- Gold: The institution delivers outstanding teaching, learning and outcomes of the highest quality

- Silver: The institution delivers high quality teaching, learning and outcomes exceeding the baseline quality threshold expected
- Bronze: The institution delivers teaching which is satisfactory meeting the baseline quality threshold expected
- Provisional: Insufficient data

Taken together, the complex nature of the emerging TEF was already becoming apparent (Wonkhe 2016):



As the most recent TEF ratings may not provide an accurate or up-to-date reflection of existing teaching quality, particularly following the disruption to higher education caused by the Covid-19 pandemic, universities, colleges and other providers are no longer permitted to advertise outcomes from the 2018-19 exercise.

Following publication of an independent review of TEF (Pearce 2019), including a commissioned evaluation of the statistical element undertaken by the Office for National Statistics (James et al. 2019), the OfS has developed a new TEF framework, with the publication of consultation phase outcomes and decisions in July 2022 (see later). The OfS aims to begin work receiving and assessing submissions for new TEF awards from September 2022 with outcomes published later in 2023 in time for the 2024-25 recruitment cycle.

Complete details of all central matters and reports associated with the TEF are available on the OfS website with an ongoing analysis and critical commentary at every stage of development provided by Wonkhe. For a valuable overview and

broader alternative perspective on what might constitute high-quality higher education teaching and learning, as well as pedagogical research, see Evans et al. (2021).

### **A general critique of TEF**

The fourteen articles selected for mention or review and summarised or paraphrased here reflect the main highlights arising from the introduction of TEF between the initial announcement in 2015 to 2021. These are presented in a broadly chronological order to chart TEF developments over time. While somewhat dated in terms of the dynamic and ever-changing nature of TEF, as well as the higher education sector itself, the central thrust of arguments and points raised across all of these articles remain relevant, particularly where those involve a discussion of metrics.

#### *The Green and White papers*

In an academically informed, rapid-response analysis of the policy proposal for the TEF set out in the original Green paper of 2015, Forstenzer (2016) asks about the meaning of teaching excellence, which, as the sum of countless factors, is almost impossible to define (e.g. addressing teaching quality as process, learning environment as presage and student outcomes and learning gain as product), what happens if excellence is ever universally attained (e.g. excellence becomes normal), which measures stand as proxies for excellence (e.g. easy to measure numerical indicators or contextualized qualitative data), and who should resolve upon the matter of excellence itself (e.g. politicians or educational specialists)? In an excellent deconstruction of the Green paper, Forstenzer concludes by suggesting that the TEF is not really about teaching excellence but about fees and accountability to taxpayers, that the TEF does not serve students or institutions at all but strives to increase fees by stealth and to further extend an already market-driven sector, that the notion of 'employers' is ill-defined, and that the TEF ignores the non-economic, social benefits and public good of undergraduate education focusing instead on the self-interest of the individual and personal growth alone.

Addressing the same Green paper, Berger and Wild (2017) add to the work of Forstenzer by highlighting that while the TEF outcomes prioritize the quantitative component of the process, and the use of 'Big Data', the overall outcome is in danger of succumbing to a number of potential risks in the form of influencing institutional strategic response and decision-making, adversely skewing behaviour and 'homogenizing' provision across the sector. By way of remediation, Berger and Wilde suggest that the qualitative component of the assessment process is essential for providing contextual relevance. In a useful analogy, Berger and Wild indicate that a system relying on 'Big Data' treats all universities as the same 'green apples', but not all 'green apples' are the same shade of green and many perform different functions in different ways with some for eating and some for cooking (it might also be said that some might also be considered inedible). However, no amount of

quantitative or qualitative data would ever be capable of determining a quality threshold.

Gibbs (2017) and McNay (2017) took perhaps a more direct view of initial proposals, commenting that while the TEF had evolved since first announced, its underpinning rationale, that all higher education provision was uniform and therefore directly comparable, was deeply flawed and that the interpretation of evidence about educational quality, employability and value for money were 'irrational' and not supported by evidence, with some of its metrics invalid and not fit-for-purpose. Employability was considered to have little to do with teaching quality, a perception of poor value for money was not necessarily linked to or caused by poor quality and high fees were not a proxy for a better education. Further, it was suggested that the ability of a small panel of experts to make sound quality judgements was not well founded.

#### *Later developments and initial results*

While later critiques of the TEF occasionally took on a deeper and more philosophical stance (Forstenzer 2018; Canning 2019), including the charge of 'symbolic violence' by Tomlinson et al. (2020), others remained committed to addressing its function, design, implementation and intended outcomes, as well as the policy narrative and legislation driving it in more conventional ways, often lamenting the 'black box' methodology and metrics-driven 'deliverology' (Crockford 2020) and the additional layers of bureaucracy and micro-management involved (Barkas et al. 2019).

On the matter of informing choice, Gunn (2018a) correctly reminded readers that decisions about where to study, certainly in English higher education, was often determined by institutional reputation rather than factual evidence and therefore not as informed as it might well be. The TEF could therefore be seen as a legitimate response to the need for better information. The TEF could also be seen as a possible solution to the escalating cost of higher education via the control of fees and loans, capable of providing a means to addressing sectoral growth as well as societal change and expectation over time, and raising the status of teaching overall. However, teaching excellence was still viewed as a highly contested, value-laden, context-specific and sometimes nebulous entity remaining impossible to measure in any sensible and meaningful way, despite now being broken down into quality, environment and outcomes and gains (e.g. the acquisition of lifelong learning and other skills that would allow graduates to make a strong contribution to society, the economy and the environment, or to progress to further study or to compete for a graduate level job). The use of National Student Survey metrics was also considered questionable, particularly in terms of equating student satisfaction ratings with teaching quality, which are completely different things requiring separate measurement. National Student Survey metrics also failed to provide direct representation of pedagogy or teaching intensity, subjectively or objectively. Gunn also, perhaps prophetically, indicated that the main ramification of the TEF was as a 'badge of honour'. Interestingly, however, Gunn also reported that while participation

in the Year 2 survey (2016-17) which was available at the time was only 60%, including 134 HEIs and 106 FE colleges and 55 alternative providers, 60 were rated gold (26%), 115 silver (50%) and 53 bronze (23%), the symmetry bearing an uncanny resemblance to a pre-determined rather than naturally occurring outcome. The results also showed no relationship with the age, research income or the socio-economic background of the student cohorts involved, challenging the picture presented by the REF.

In Deem and Baird's (2019) thoughtful contribution and political analysis, the TEF was considered in terms of the concept of 'intelligent accountability' and the notions of self-governance and trust, pointing out that without trust society is drawn into a position of feeling the need to monitor one another and to gauge one another's performance, increasing competition rather than collaboration, and reinforcing a market-driven ideology. In Deem and Baird's opening position, they considered the TEF an index rather than a measurement and made up of a 'basket' of different indicator variables of interest which when combined only define a construct by virtue of its component parts. To call such an index a measure of 'teaching excellence' was simply invalid both by definition and logic. Exemplified using the nature of some of the 'basket' variables incorporated, National Student Survey data was also taken to task in detail. This, the authors suggested, always attracted a notoriously low response rate, with surveys undertaken at a time when responses themselves might be affected by immediate and unconnected events rather than reflect the whole student experience over time. In addition, the student experience of provision at any given institution was always inward looking and provided in the absence of any familiarity with what goes on across the sector. The questions being asked were not compatible with how the responses were being used. Further, it was also stated that HEIs and other providers actually had very little control over the demographic of the students they attracted which may affect how benchmarks derived from HESA data impacted upon them. Deem and Baird also noted that the labour market destinations of an institution's graduates, the employment strategies of employers and labour market discrimination, all of which often exhibited considerable regional variation, sat well outside institutional control. Deem and Baird also drew attention to the fact that with so many institutions receiving only one of three awards in any given year, and with some short-term volatility in movement between ratings, the awards themselves were insufficiently differentiated and had a limited value in terms of shelf-life. In a final statement of relevance, Deem and Baird also reminded readers that while class, ethnicity, gender, race, disability and many other factors besides can and do significantly affect educational outcomes, the same educational outcomes can also be affected by students turning up for class, studying hard, completing assessed work to the best of their ability and making an effort.

### *Final thoughts*

In more recent years, even allowing for a break in TEF activity and the lead-in time between submitting an article to a journal and its eventual publication, academic interest in the TEF appears to have waned, perhaps as a result of the independent

*Pearce Review* published in 2019 (see later). With attention moving on from earlier concerns, most of which had already been clearly and comprehensively articulated, new insights did include how TEF failed to consider and capture the experiences, perceptions and voice of those academic staff actually engaged in teaching and other related matters thereby eroding its credibility (Cui et al. 2021). Critique also extended to the potential for 'gaming' and how the use of certain themes and key words in the accompanying TEF narratives resulted in the subsequent 'upgrading' of initial ratings (Matthews and Kotzee 2021).

### **TEF metrics and learning gain**

Perhaps the single-most important area of concern surrounding the introduction of the TEF was in the establishment and subsequent use of TEF metrics to demonstrate learning gain, an area still shrouded in mystery, and part of the 'black box' methodology described by Crockford (2020) earlier.

#### *TEF metrics*

Presented by Gunn (2018b) as 'another chapter in the application of New Public Management' in higher education, the assessment of teaching quality has always proved problematic, and not just in the UK. Described further by Gunn as far from transparent and not beyond reproach, the actual choice of TEF metrics and the assessment criteria adopted have always drawn criticism, with the initial design of the TEF constrained by the very metrics available at the time. By way of a reminder, and as summarised by Behle and Maher (2018), the TEF has focused on three elements of excellence by way of its own authority and metric selection, thereby cementing the notion of students as consumers in a highly marketized system rather than as learners:

- Teaching quality (e.g. teaching, assessment and feedback data from the National Student Survey)
- The learning environment (e.g. academic support and continuation data from the National Student Survey and HESA respectively)
- Student outcomes and learning gain (e.g. employment and further study data from the Destination of Leavers from Higher Education Survey)

However, and as commented upon by Polkinghorne et al. (2017):

- How is it possible to begin to measure something that is not yet fully understood or agreed upon?
- How can using the results of a process within the marketing of higher education be justified when based upon such high levels of ambiguity and uncertainty?

Viewed as underpinning teaching quality and the learning environment, and with reference to learning gain alone, Polkinghorne et al. go on to point out that:

- The evaluation of learning gain must be reliable and valid if it is to provide a suitable indication of excellence
- New models of evaluating learning gain may be necessary if the sector is to achieve the objectives set out within TEF
- Any proposal to measure learning gain must not become a bureaucratic and administrative burden for both staff and students alike
- Any measure of the learning of students should disregard external factors not linked to teaching quality and over which institutions have no control
- The effect of gaming upon the data collected for the purposes of evaluating student learning should be minimized
- Learning gain metrics must capture the diversity of subjects, accommodate variation in teaching styles and support innovation
- Learning gain metrics should form a natural part of the ongoing course review process and be relevant to institutions, students and employers alike

Indeed, and in the influential RAND report presented by McGrath et al. (2015) some three years earlier, the authors clearly articulate the following position:

*The question of how to measure the student learning gained in higher education has been much debated. The debate is healthy - any attempt to measure a process which is as multidimensional as learning can encourage a focus on what is effective - but it also risks distorting activity. There is an evident need to consider different approaches to the measurement of gains in learning and consider their impact on practice and on outcomes. Debate on the measurement of the learning students gain in higher education has been prominent internationally, but the concept has not been studied extensively in the English higher education context. (xi)*

#### *Learning gain defined*

Echoing the work of McGrath et al. (2015) and Polkinghorne et al. (2018), and in one of the most comprehensive reviews available in the UK higher education context, Evans et al. (2018) summarise a number of available definitions of learning gain, each varying in detail and intent:

- The distance travelled, or the difference in skills, competencies, content knowledge and personal development demonstrated, by students at two points in time
- An attempt to measure the improvement in knowledge, skills, work-readiness and personal development made by students during their time spent in higher education (overcoming the two-point measurement designs required by the previous definition)
- A comparison between performance predicted at the outset of studies and the actual performance achieved (value added)



In combination, Evans et al. also considered an integrated definition of learning gain as ‘a change in knowledge, skills, work-readiness and personal development, to include beliefs and values, and the enhancement of specific practices and outcomes in defined disciplinary and institutional contexts’.

Exploring the rationale for considering the use of learning gain, and with particular relevance to the TEF, Evans et al. make some interesting observations. Widening participation and social mobility agendas have, for example, always provided a particular momentum for measurement and a desire for metrics which extend beyond historic reputation or mission group and which permit the identification of students, courses and institutions that demonstrate the knowledge, skills and other attributes that employers are looking for and the economy needs. This also alleviates the burden and inefficiencies of employers having to do this for themselves. In turn, however, this has also led to a broader conceptual debate questioning the very nature and function of higher education itself as well as what to measure, how to measure it, the level at which the measurement takes place (e.g. the institution, faculty, department, discipline, module or individual) and the purpose for which that measurement is undertaken (e.g. improvements for the different stakeholder groups including government). Interestingly, Evans also notes that the introduction of a TEF may, in fact, have reduced teaching quality (see also Evans et al. 2021).

*The HEFCE/OfS learning gain pilot projects (2015-2018)*

Following on and taking the lead from the RAND report presented by McGrath et al. (2015), HEFCE awarded over £4 million to 13 pilot projects of up to three years in duration involving over 70 higher education institutions with the aim of testing and evaluating measures of learning gain in England (including the University of Lincoln as one of the 13 leads). While the OfS took over managing the projects in 2018, the final evaluation report presented by Kandico-Howson (2019) and summarised here appeared after the introduction of the TEF itself. The extent to which any findings were assimilated into earlier OfS deliberations is not known with any degree of certainty though these are acknowledged and referred to in various ways in the subsequent 2022 consultation and response (see also Behle and Maher 2018 and Arico et al. 2018).

Learning gain across the 13 projects was captured in a variety of ways including different cognitive measures of what students thought and knew, different behavioural measures of what students did and how they engaged, and different affective measures of how students felt. Where developed and employed for purpose, the use of standardised tests did not provide robust or effective learning gain measures at all, but personal and soft skills gains were best captured through surveys and focus groups, and work readiness data was usefully captured through careers registration. It was recommended that any measures employed needed a clear rationale for use and should be done so in partnership with students themselves in order to support learning and self-reflection. At best, these should also be embedded into the curriculum and individual institutional processes in order to be

seen as valued educationally. Measuring learning gain was also seen as a complex and long-term investment, requiring multiple measures to capture diversity over time and direction, and requiring up to 5 years for longitudinal projects to come to fruition in any meaningful way. Contextual factors including institutional type, subject-level differences and student characteristics were also thought to impact on the transferability of learning gain measures and should be considered carefully in the presentation of findings.

*The OfS HELGA report (2019)*

As part of the ongoing learning gain programme initiated by HEFCE (2015-2018), the *Higher Education Learning Gain Analysis* (HELGA) was published by the OfS (2019) in order to assess whether or not it was possible to use only administrative data to measure learning gain as cognitive gain from the higher education experience of students across two points in time. Within the context of the statistical nature of the work undertaken, institutional value-added measures were created as a proxy for learning gain using multi-level modelling and pairwise comparison in order to draw institutional comparisons. In association with potential predictor (e.g. GCSE and A-level results) and output variables (e.g. degree classification and employment status and salary after 6 months), and adjusted for context, the project reported having failed in finding a single measure of learning gain that could be used across the sector based on administrative data alone. This was partly due to the data collected by institutions not being done so with the express purpose of measuring learning gain and the complexity of the concept of learning gain itself.

*The OfS NMMLGP report (2019)*

As a further part of the learning gain programme initiated by HEFCE (2015-2018), the *National Mixed Methods Learning Gain Project* (NMMLGP) undertaken across 10 participating higher education providers was led by a team of researchers from Sheffield Hallam University and reported, with a small review of learning gain literature, to the OfS by Jones-Devitt et al. (2019). As it progressed, the NMMLGP evolved into a study of student perceptions of learning gain itself. The original aim of the NMMLGP was to understand the logistical requirements of administering a sector-wide learning gain survey questionnaire to students across multiple providers in England (e.g. to identify approaches and practices that worked well and those that did not, to identify opportunities and challenges, to identify issues for scaling-up including student support, and to gather student perceptions).

Despite details of the circulated questionnaire being seemingly unavailable, the NMMLGP evaluation team at Sheffield Hallam concluded that a 'one-size-fits-all' measure of learning gain, based upon the principles adopted, held minimal value for students and that the search for a single indicator should be discontinued. It was also noted that the sector needed to consider whose interests were being served by the measurement of learning gain and to what end (e.g. institutional positioning within a market-driven system or a process of progression reflecting the student journey),

and that the measurement of learning gain, if attempted, needed to be related to the student context and embedded at a local level within subject or discipline, and that student perceptions of learning gain required further exploration beyond impressionistic accounts and idiosyncratic meaning. Providers were also advised to consider a repertoire of approaches as part of a learning gain toolkit which might be accessed by students as part of a process underpinned by student choice rather than normative comparison as well as to consult the findings of the HEFCE/OfS led learning gain programme for approaches most suitable to their own local context.

### **The Pearce Review (2019)**

Tasked by the then Secretary of State for Education to provide an independent and objective review of the TEF 'scheme' as a whole, the Pearce advisory group of nine 'stakeholder experts' invited and led by Dame Shirley Pearce herself got to work. Describing TEF as 'still at an early stage of development', the 124-page review presented by Pearce (2019), clearly drawing on the research literature and other evidence available, together with the 129-page commissioned evaluation of the statistical elements of TEF conducted by the Office for National Statistics and reported by James et al. (2019), make very interesting reading. The UK government's response, which ran to only three printed pages of largely complementary if somewhat delayed and tokenistic text, was published over a year later (DfE 2021), prompting the most recent phase of OfS consultation nevertheless (see later).

The Pearce advisory group specifically investigated:

- The process by which ratings were determined under the scheme for assessing quality in higher education and the sources of statistical information used in that process
- Whether that process, and those sources of statistical information used, were fit for the purpose of determining ratings under the scheme
- The names of the ratings under the scheme and whether those names were appropriate
- The impact of the scheme on the ability of higher education providers to which the scheme applies to carry out their functions
- An assessment of whether the scheme is in the public interest
- Any other matters that the person appointed to lead the review considered relevant

The review methodology, including the various workstreams identified and from within which the evidence was gathered, included the following:

- A call for views: Public consultation from within higher education and beyond
- Listening sessions: A series of over 60 meetings with a wide range of stakeholder groups

- Statistical evaluation: The use of metrics in assessment, the role of split metrics, the number of metrics and their weighting, the use of and method for identifying flags, the treatment of very high and very low values, data sources used to generate the metrics, 'initial hypotheses', and the presentation and transparency of the statistical information and process
- Analysis of the public interest: Desktop review of existing research covering costs and benefits, quantitative cost estimates for provider-level and subject-level TEF, a targeted seminar with relevant sector organizations and individuals to discuss existing evidence and views
- International perceptions: From independent research conducted by the British Council
- Employers: Focus group meetings with employers, employer representative groups and professional bodies
- Applicants and careers advisors: Desktop review of existing research about the role of TEF in providing information to existing and prospective higher education students

From the volume of evidence gathered and considered, a series of 10 recommendations were presented under six overarching themes. These are summarized here along with additional pertinent information from within the body of the report as follows:

#### *Purpose*

- To achieve the greatest impact for students and employers, the TEF should have the 'primary purpose' of identifying excellence and encouraging the enhancement of educational experiences and outcomes. The term educational excellence was preferred over teaching excellence as studying in higher education involved more than being in receipt of excellent teaching. The TEF included no direct measure of teaching and there was no general agreement over which metrics to use. Employers of graduates and other stakeholders seemed largely unaware of the TEF at all.

#### *Principles*

- To improve confidence and respect in the framework, three principles should guide subsequent development and delivery: Transparency (e.g. clarity and understanding), relevancy (e.g. purpose and provision) and robustness (e.g. statistical and operational). This had not always been clear in the past and required a targeted communication programme. Results from the TEF should address only those matters to which it pertains and no other political issue or concern. Metrics over which an institution has no control result in poor measures of an institution's ability to deliver excellence. Higher education providers have diverse missions that result in very different provision, despite the use of common metrics across all. The analysis of data and ratings

judgements should be valid and reliable and derived by experts rather than government or other quasi-governmental bodies.

#### *The assessment process*

- Statistical improvements were needed to address concerns in the management and communication of statistical uncertainty at all levels of the process, the use of small numbers (e.g. from small providers and/or small datasets and split metrics) and non-reportable metrics. The assessment of outcomes sometimes appeared too focused on employment, without consideration of regional variation, and at the expense of wider social and cultural benefits which are not included. The most polarized views involved the use of benchmarks over absolute performance.
- A subject-level exercise should be developed within the provider-level assessment to inform ratings at provider-level rather than subject-level. The process and statistical risks become exacerbated at subject-level where the impact of problems due to subject categorization and small numbers becomes greater, though this has value for driving internal enhancement. Work was needed to develop an effective way of achieving this at scale. Students did report that a subject-level analysis would be more useful to them for decision-making.

#### *Developing the framework*

- The structure of the framework should be adjusted to consider *Educational Experience (teaching and learning environment and student satisfaction)* and *Educational Outcomes (educational gains and graduate outcomes)* to improve transparency about what is being measured, its relevance to all institutions and the balance between nationally comparable and institutionally determined evidence. The Pearce advisory group favoured the term educational gain to learning gain, recognizing the importance of increasing the focus of the TEF 'scheme' on learning and not just teaching and allowing institutions to demonstrate excellence specific to their mission.
- National Student Survey metrics should only be used to assess student satisfaction and not teaching quality and the learning environment which they were never designed to do, as indicated.
- Each institution should be expected to demonstrate how, within their own particular mission, they articulate and measure the educational outcomes and the learning they aim to provide for their students. This could be achieved by incorporating the measurement of educational rather than learning gains to address what students get from their educational experience and how that can be evidenced within the *Educational Outcomes* dimension in the revised framework, as indicated.
- In assessing graduate outcomes, broader outcome metrics should be adopted rather than be restricted to employment and earnings, and the metrics used to assess employment and earnings should control for regional variation. The

location of employment should be used as a benchmarking factor or in creating the metric.

- The submission process should have a standard structure for submissions used by all. The student body should be enabled to contribute their own independent and structured written input that reflects their view of the institution's performance in all aspects of any revised framework.

#### *The rating system*

- The rating system should be improved to provide greater and more nuanced information under the four dimensions considered for the revised framework structure indicated above. A single rating, attractive as that may be, risks over-simplifying any outcome. Rating names should make clear that the awards demonstrate excellence that meets or exceeds the UK quality baseline, replacing gold, silver and bronze with outstanding, highly commended and commended and, finally, meets UK quality requirements.

#### *The name of the scheme*

- The name of the 'scheme' should be changed to the *Educational Excellence Framework* to better reflect what is being measured and assessed.

The *Pearce Review* further addressed the matter of educational gain briefly over an additional two pages of the report (pp.56-57). Commenting that while some institutions refer to the learning gain of their students (e.g. knowledge, skills, experience, readiness for work or further study, personal development and resilience), it was noted that different institutions and different disciplines conceptualize and measure learning gain in different ways and that no single nationally comparable definition or measure currently exists. A better and closer working relationship between the DfE, OfS and interested academic researchers was encouraged to develop this area further.

### **Subsequent consultation, outcomes and decisions (2022)**

Following the hiatus in TEF activity since 2019, and the UK government's brief response to the *Pearce Review* in 2021, the OfS embarked upon a sector-wide consultation to inform future developments and develop a new approach. This ran from January to March 2022 (OfS 2022a), with a subsequent analysis and report on findings appearing in July of the same year (Alma 2022). An OfS response with decisions followed on almost immediately (OfS 2022b). This was published alongside an additional report considering student perceptions of the proposed TEF naming and rating options (YouthSight 2022).

Of the 239 responses to the 15 proposals presented in the consultation document, the majority were received from employees on behalf of a higher education provider, though by no means exclusively. Incorporating framework, scope, evidence,

assessment, outcomes and implementation, an overview of the multiple-choice (strongly agree, tend to agree, tend to disagree, strongly disagree, don't know/prefer not to say) and free-text findings is presented as follows. Interestingly, and throughout the consultation, proposals were linked to a 'general duty' to promote quality, choice and opportunity to students, 'regulatory practice' and existing OfS resources in an 'efficient and economic way', thereby limiting the extent to which the Pearce recommendations and a complete revision might be fully adopted (indeed, some recommendations from within the *Pearce Review* were completely overlooked albeit with justification). It is also worth pointing out that behind the apparent simplicity of each of the 15 proposals and descriptive narratives clarifying coverage and meaning lay a complex level of confounding detail reinforcing a clear unevenness in item weighting and significance. In many instances, the basic principles of questionnaire design were also violated (e.g. asking multiple, conflicting or leading questions in a single statement).

In the overview provided, with the final OfS decision presented alongside, respondents clearly 'disagreed' with the statements of three items with the responses to a further three items perhaps best described as mixed.

#### *Framework*

- Provider-level, periodic ratings: An overall rating should be awarded to a provider reflecting the quality of its undergraduate courses, and these ratings should last for four years. 83% 'agree'.

OfS Decision: Proceed with no change.

- Aspects and features of assessment: Two aspects should be assessed and rated, the student experience and student outcomes. 77% 'agree'.

OfS Decision: Proceed with no change. Further clarification on the scope of learning gain articulated in a provider's submission. This should go beyond the B3 measures of continuation, completion and progression used for TEF indicators, be relevant to a particular mix of students and courses at a provider, and ideally take account of student differences in starting points and the distance travelled (see below).

- Rating scheme: There should be three rating categories – gold, silver and bronze – signifying the degrees of excellence above baseline quality requirements. 68% 'disagree'.

OfS Decision: Proceed with no change.

- Absence of excellence: Where there is an absence of excellence, no rating should be awarded. 74% 'disagree'.

OfS Decision: Proceed with no change.

### *Scope*

- Provider eligibility: To be eligible to participate in TEF, and to retain a rating once awarded, a provider must satisfy baseline quality and standards requirements. 61% 'agree'.

OfS Decision: Proceed with one change. To participate in TEF a provider must have at least two indicators with a minimum denominator of 500.

- Courses in scope: All of a provider's undergraduate courses, and the students on those courses, should be within the scope of a TEF assessment. 42% 'agree', 51% 'disagree' (mixed).

OfS Decision: Proceed with one change. Apprenticeships will only be in scope if a provider chooses to include evidence about them in its submission.

### *Evidence*

- Provider submissions: Participating providers should submit evidence of excellence in relation to the experience and outcomes of their students. 85% 'agree'.

OfS Decision: Proceed with one change. Page limit for provider submissions increased from 20 to 25 pages.

- Student submissions: Students should be encouraged to submit their views on the quality of their experience and outcomes. 76% 'agree'.

OfS Decision: Proceed with one change. Outcomes focus on current rather than past students. Optional to include students who are registered elsewhere.

- Indicators: The OfS should produce numerical indicators based on the National Student Survey responses, and student outcomes indicators defined consistently with the indicators to regulate student outcomes through condition B3 (continuation, completion, progression). 56% 'agree', 39% 'disagree'(mixed).

OfS Decision: Proceed with several changes. The positive interpretation of a provider's benchmark for continuation falling not materially below 95% will extend to experience and outcomes measures. Simplification of where students are either taught or sub-contracted out. No final decision on proposed choice of benchmarking factors.



### *Assessment*

- Expert view: Ratings should be decided by a TEF panel applying expert judgement. 88% 'agree'.

OfS Decision: Proceed with no change.

- Assessment of evidence: The panel should interpret and weigh up the evidence within a set of principles and guidelines. 74% 'agree'.

OfS Decision: Proceed with no change.

### *Outcomes*

- Published information: TEF outcomes and the evidence used in assessment should be published in an accessible and timely way. 59% 'agree', 33% 'disagree' (mixed).

OfS Decision: No final decision at this time but minded to proceed with no change.

- Communication of ratings by providers: A provider should be able to display and promote its own TEF rating in accordance with a set of guidelines. 86% 'agree'.

OfS Decision: Proceed with no change.

### *Implementation*

- Name of the scheme: The scheme should be named the Teaching Excellence Framework. 69% 'agree'.

OfS Decision: Proceed with no change.

- Timing of the next exercise: The next exercise should be carried out during 2022-23 and outcomes published in spring 2023. Future exercises should be conducted every four years. 85% 'disagree'.

OfS Decision: Implementation timetable changed. In detail, OfS will appoint a panel by September 2022 with guidance for submissions in the same month. The submissions window will be open from September 2022 to mid-January 2023. Provisional decisions will be made known July to August 2023 with data published by September 2023 for providers that do not make representation following provisional decisions.

For the most part, many of the concerns raised in the free-text responses at consultation reflected those already identified in the research literature, and which the OfS did draw upon in its decision-making, but with a particular call for further clarity as indicated:

- The definition and measurement of learning gain
- The interconnection between the TEF ratings and the baseline quality and standard requirements (mainly the B3 requirements)
- The decision-making process, particularly in relation to how the panels will be comprised and how they will operate, the representations to be made by providers and the communication of the ratings
- The fee level that could be charged by a provider choosing not to participate in the TEF or receiving a 'requires improvement' judgement
- The student submission, particularly regarding who would be responsible for producing the student submission and whether there would be compensation for students' time

#### *Learning gain revisited*

On learning gain specifically, the OfS consultation did make provision which recognised that higher education benefits students in ways that extended beyond the outcomes measured as part of baseline regulation. As such, providers wishing to be recognised for excellence through the TEF would be able to explain what they intended their students to gain from their education, and to provide evidence of how well they were succeeding. The differences between intended gains and how these might be measured were acknowledged. Given the lack of a national measure, such gains would most likely be assessed on the basis of self-determined quantitative and qualitative evidence submitted including, where appropriate, the student submission itself.

The OfS response and decisions document (OfS 2022b) devotes four pages to the issue of learning (pp.28-31). At consultation, some respondents had suggested delaying the inclusion of learning beyond the next TEF to give providers more time either to develop definitions and measures at a provider-level or to arrive at common understandings at a sector level. More information on the scope of such gains was also requested. Other respondents were concerned about the administrative and other burdens involved as well as inconsistencies in content and reporting. OfS indicated clearly that the benefits of inclusion outweighed the alternatives and that their view of 'features of excellence' could incorporate learning gain more broadly. OfS considered it reasonable for providers to articulate for themselves what learning gain means in their own context, and that the concept in TEF was neither new nor unfamiliar. The OfS did go on to state, however, that 'a provider would not be prevented from being awarded a higher TEF rating solely on an absence of developed learning gain measures'. For the OfS, it should also be noted that learning gain must extend beyond B3 requirements, be relevant to a mix of students

and courses at the provider, and ideally take account of different starting points and distance travelled.

For the OfS, learning gain could include but not be limited to:

- Academic development: Subject knowledge and academic skills including critical thinking, analytical reasoning, problem solving, academic writing and research and referencing skills
- Personal development: Student resilience, motivation and confidence, communication and presentation skills, and time management, networking and interpersonal skills
- Work readiness: Employability skills such as teamworking, commercial awareness, and leadership and influence

As part of its evaluation of TEF, OfS encourages providers to 'establish their own projects and resources, to share best practice, and to develop their [own] definitions of learning gain in collaboration with their students'.

#### *YouthSight (2022)*

The YouthSite (2022) assessment of student perceptions commissioned by OfS, a valuable exercise but one which could in this instance also be viewed as a transparent attempt by OfS to justify its actions, involved a 30-minute online qualitative task survey with 20 respondents followed up with a 10-minute online quantitative survey with a sample of 1,112 respondents, 826 of whom were at the application stage for university and 286 were first year undergraduates. Overall findings, which should be treated with caution in terms of the small sample sizes involved, representativeness and the limited information provided to respondents with which to make a fully informed decision, indicated that most believed that the TEF would have a positive impact on the higher education sector and agreed that it would help inform decision-making on where to study.

## **Conclusions**

In summary, this article presents an overview of developments leading up to and following the introduction of a still complex Teaching Excellence Framework (TEF) from initial proposals and adoption in government Green and White papers in 2015 and 2016 and the *Higher Education and Research Act of 2017* to the most recent OfS iteration in 2022, with a particular focus on learning gain. Originally and still intended to better inform students' choices about what and where to study, raise esteem for teaching, recognise and reward excellent teaching and better meet the needs of employers, business, industry and the professions, the TEF remains highly contested and continues to attract criticism from across the sector and from other interested stakeholder groups.

While definitions and measures of learning gain within the TEF framework might include, but not be limited to, academic and personal development and work readiness, the measurement of learning gain directly or by proxy is, however, far from unproblematic and can, in fact, be undertaken in different ways by different means to suit different purposes, as can the measurement of a broader and high-quality student experience itself. With some awareness of learning gain across the UK higher education sector as a whole, this remains somewhat patchy and uncertain, with robust systems of measurement remaining elusive and in their early stages, leading to difficulties of comparison both within (e.g. at a disciplinary level) and between (e.g. at a cultural level) institutions, possibly resulting in the OfS encouraging providers to 'establish their own projects and resources, to share best practice, and to develop their [own] definitions of learning gain in collaboration with their students'.

From within the self-selected but hopefully representative literature consulted here, entry-level reading, and an essential starting point for those interested in the development of learning gain metrics for TEF, those involved in planning and the determination of a strategic and operational response to future TEF exercises and those involved in the establishment of TEF policy in detail, includes the early but comprehensive Rand report published by McGrath et al. (2015), the review of learning gain provided by Evans et al. (2018) and the evaluation and final report of the HEFCE/OfS learning gain pilot projects published by Kandico-Howson (2019).

## References

Alma (2022) *The Teaching Excellence Framework (TEF): Analysis of consultation responses*. London: Alma economics.

Arico, F., Gillespie, H., Lancaster, S., Ward, N. and Ylonen, A. (2018) Lessons in learning gain: Insights from a pilot project, *Higher Education Pedagogies*, 3(1), 249-265.

Barkas, L.A., Scott, J.M., Poppitt, N.J. and Smith, P.J. (2019) Tinker, tailor, policy-maker: Can the UK government's teaching excellence framework deliver its objectives? *Journal of Further and Higher Education*, 43(6), 801-813.

Behle, M. and Maher, S. (2018) Measuring 'teaching excellence' and 'learning gain' in the UK. *Beiträge zur Hochschulforschung*, 40, 48-66.

Berger, D. and Wild, C. (2017) The Teaching Excellence Framework: Would you tell me, please, which way I ought to go from here? *Higher Education Review*, August, 1016.

Canning, J. (2019) The UK Teaching Excellence Framework (TEF) as an illustration of Baudrillard's hyperreality. *Discourse: Studies in the Cultural Politics of Education*, 40(3), 319-330.

Crockford, J. (2020) 'Wishing won't make it so': Deliverology, TEF and the wicked problem of inclusive teaching excellence. In: Thomas, K.C. and French, A. (eds.) *Challenging the Teaching Excellence Framework (Great Debates in Higher Education)*, Emerald Publishing Limited, Bingley, 95-127.

Cui, V., French, A. and O'Leary, M. (2021) A missed opportunity? How the UK's teaching excellence framework fails to capture the voice of university staff. *Studies in Higher Education*, 46(9), 1756-1770.

Department for Education (2021) *Government response to Dame Shirley Pearce's independent review of the Teaching Excellence and Student Outcomes Framework (TEF)*. London: HMSO.

Evans, C., Kandiko-Howson, C. and Forsythe, A. (2018) Making sense of learning gain in higher education. *Higher Education Pedagogies*, 3(1), 1-45.

Evans, C., Kandiko-Howson, C., Forsythe, A. and Edwards, C. (2021) What constitutes high quality higher education pedagogical research? *Assessment & Evaluation in Higher Education*, 46(4), 525-546.

Forstenzer, J. (2016) *The Teaching Excellence Framework: What's the purpose?* Sheffield: The Crick Center for Engaged Philosophy and the University of Sheffield. 1-31.

Forstenzer, J. (2018) The Teaching Excellence Framework: Epistemic insensibility and the question of purpose. *Journal of Philosophy of Education*, 52(3), 548-574.

Gibbs, G. (2017) Evidence does not support the rationale of the TEF. *Compass: Journal of Learning and Teaching*, 10(2), 1-10.

Gunn, A. (2018a) Metrics and methodologies for measuring teaching quality in higher education: Developing the Teaching Excellence Framework (TEF). *Educational Review*. 70(2), 129-148.

Gunn, A. (2018b) The UK Teaching Excellence Framework (TEF): The development of a new transparency tool. In: A. Curaj et al. (eds.), *European Higher Education Area: The Impact of Past and Future Policies*, Springer. 505-526.

James, G., Ayoubkhani, D., Ralph, J. and Roarson, G.D. (2019) *Evaluation of the statistical elements of the Teaching Excellence and Student Outcomes Framework: A report by the Methodology Advisory Service of the Office for National Statistics*. ONS/MAS.

Jones-Devitt, S., Pickering, N., Austen, L., Donnelly, A. Adesola, J. and Watson, A. (2019b) *Evaluation of the National Mixed Methods Learning Gain Project (NMMLGP) and Student Perceptions of Learning Gain*. London: OfS.

Kandico-Howson (2019) *Final evaluation of the Office for Students learning gain pilot projects*. London: KCL/OfS.

Matthews, A. and Kotzee, B. (2021) The rhetoric of the UK higher education Teaching Excellence Framework: A corpus-assisted discourse analysis of TEF2 provider statements. *Educational Review*, 73(5), 523-543.

McGrath, C.H., Guerin, B., Harte, E., Frearson, M. and Manville, C. (2015) *Learning gain in higher education*, RAND/HEFCE.

McNay, I. (2017) TEF: Why and how? Ideological and operational imperatives driving policy. *Compass: Journal of Learning and Teaching*, 10(2), 1-5.

Office for Students (2019a) *Higher Education Learning Gain Analysis (HELGA)*. London: OfS.

Office for Students (2022a) *Consultation on the Teaching Excellence Framework (TEF)*. London: OfS.

Office for Students (2022b) *Consultation on the Teaching Excellence Framework (TEF): Analysis of consultation responses and decisions*. London: OfS.

Pearce, S. (2019) *Independent review of the Teaching Excellence and Student Outcomes Framework (TEF): Report to the Secretary of State for Education*. London: Crown.

Polkinghorne, M., Roushan, G. and Taylor, J. (2017) Considering the marketing of higher education: The role of student learning gain as a potential indicator of teaching quality. *Journal of Marketing for Higher Education*, 27(2), 213-232.

Tomlinson, M., Enders, J. and Naidoo, R. (2020) The Teaching Excellence Framework: Symbolic violence and the measured market in higher education. *Critical Studies in Education*, 61(5), 627-642.

Wonkhe (2016) *The even more incredible machine: Our visual guide to TEF*. Online: [www.wonkhe.com](http://www.wonkhe.com).

YouthSight (2022) *Assessing student perceptions of proposed TEF naming and rating options*. London: YouthSight.