



Teaching Excellence Framework Technical Consultation – Response Form

Name/Organisation: Joint submission on behalf of a federation of learned societies of Politics, International Relations/Studies and European Studies. This submission has been endorsed by the British International Studies Association (BISA), Political Studies Association (PSA), and UACES: the Academic Association for Contemporary European Studies.

Address: c/o British International Studies Association, International Politics Building, Aberystwyth University, Penglais Campus, Aberystwyth, Ceredigion, SY23 3FE

Email Address: gpb@aber.ac.uk; helena.djurkovic@psa.ac.uk; elinnemann@uaces.org

Please tick the box that best describes you as a respondent to this consultation:

	Respondent type
<input type="checkbox"/>	Alternative higher education provider (with designated courses)
<input type="checkbox"/>	Alternative higher education provider (no designated courses)
<input type="checkbox"/>	Awarding organisation
<input type="checkbox"/>	Business/Employer
<input type="checkbox"/>	Central government
<input type="checkbox"/>	Charity or social enterprise
<input type="checkbox"/>	Further Education College
<input type="checkbox"/>	Higher Education Institution
<input type="checkbox"/>	Individual (Please describe any particular relevant interest; parent, student, teaching staff etc.)
<input type="checkbox"/>	Legal representative
<input type="checkbox"/>	Local Government
<input type="checkbox"/>	Professional Body
<input checked="" type="checkbox"/>	Representative Body
<input type="checkbox"/>	Research Council
<input type="checkbox"/>	Student
<input type="checkbox"/>	Trade Union or staff association

	Respondent type
<input type="checkbox"/>	Other (please describe)

Question 1 (Chapter 1)

Do you agree with the criteria proposed in Figure 4?

Yes No Not sure

Please outline your reasons and suggest any alternatives or additions.

We broadly support the identification of three aspects of teaching excellence: Teaching Quality; Learning Environment; and Student Outcomes and Learning Gain. That said, we question the wisdom of a variated process which begins with an institutional level assessment before moving towards discipline specific assessments as of 2019-20. We offer the following observations under each of these headings:

1. Teaching Quality.

We welcome the explicit recognition in the criteria that engaging students in stimulating and challenging learning is an indispensable part of Higher Education (HE). We welcome the recognition that institutional cultures and policies which support teachers and the development of education are an important aspect of teaching excellence. We welcome the emphasis within course design on the importance of stretching students and supporting them to reach their full potential. We also welcome the emphasis on assessment and feedback which are core elements of the learning process.

We note that the core metrics for the assessment of these (Figure 2, page 11) will be NSS Q1-4 and Q5-9.

This approach, however, is based on an assumption that students, as consumers of HE, are in the best position to judge teaching quality – the assumption being that the more satisfied they actually are, the higher quality the teaching provision must be. There is a danger that the more powerful the NSS becomes, the more lecturers may feel under pressure to ensure immediate student gratification and the more they will feel the need to entertain and appease students at the expense of intellectual challenge. Students may be further encouraged to see themselves as passive consumers of a particular 'service', with the promise of an eventual financial reward in terms of the graduate earnings premium without sustained commitment and engagement in the learning process.

We also appreciate that, whilst there is sensitivity to the need for reflection on response rates, utilising institutional level response rates may mask significant differences in subject-level responses. Therefore, there is a need for qualitative

analysis to understand why students don't complete the NSS (for example) and what the implications of non-completion are on a subject-by-subject basis.

Overall, while the metrics identified are effective measures of student satisfaction, there is no evidence that these can measure the extent to which the student learning experience is challenging or stretching. Either this heading should be relabelled "Teaching Quality and Student Satisfaction", or more weight should be placed on those metrics that provide a more dispassionate measure of quality, such as those included in the supplementary evidence.

2. Learning Environment.

The three criteria identified in Figure 4 provide a good balance of different aspects of the learning environment, balancing resources to support learning, enrichment of teaching through research and professional practice and support for students to progress in their studies.

We note as above, that the core metrics are only partially aligned to these criteria. Those proposed in Figure 2, will report on student satisfaction with academic support and on benchmarked rates of retention. They will provide only a very limited basis for assessment on issues such as the extent to which teaching and learning is informed by the research practice of staff and/or research activity undertaken by academic staff and students in collaboration. Moreover, there is a need to ensure that the second criterion allows for as much diversity of practice as possible, so as not to stifle innovative practice. Finally, we are minded that a number of programmes within the social sciences provide what might be termed a common first year in which students undertake a portfolio first year before choosing to specialise at year 2. There is thus an active encouragement of movement between programmes and it is important that 'retention data' is sensitive to such subject-specific variation, particularly for those universities which are disproportionately affected.

3. Student Outcomes and Learning Gain

As for the areas that are outlined above, we broadly welcome the stated criteria, which we believe have the potential to provide a balanced and rounded assessment of student outcomes and learning gains.

However, we here again would draw attention to the limited alignment of these to the proposed core metrics, which focus solely on employment outcomes. Using DHLE figures as a core metric (Figure 2) to measure teaching quality reflects the notion that the more employable graduates actually are, the better the teaching must be. But universities with strong reputations tend to attract high quality students who feed into strong employability statistics. This may occur despite poor quality teaching. Employment statistics can simply reflect the longstanding reputation of the university or its ability to attract good students. Other universities may attract proportionally more women, those who intend to enter public services or lower socio-economic groups who do not have money to fast track their graduate careers. Their employability statistics reflect their own socio-economic intake, not teaching quality.

A disproportionate focus on employment metrics undermines the richness of elements of the criteria which rightly focus on the extent to which higher education contributes to “*personal* and professional lives” (emphasis added). We would support an approach that recognises the extent to which each student’s potential is realised, whatever form that might take.

We would also like to raise a further issue regarding the level of scrutiny at different ends of the assessment range. As stated in the consultation document ‘Where a provider is significantly above or below the benchmark for all or most metrics, the decision is likely to be straightforward unless contextual information and additional evidence present compelling mitigating factors’ (p.32, para. 116). The emphasis here and earlier (p.28, para.102) on the ‘mitigating’ nature of the additional information submitted by the provider, implies that where a provider has benchmark data that has a positive variance, there will be less expectation that the provider submits evidence of meeting the criteria. On this basis, there is a risk that providers with strong core metrics will not adequately be assessed and scrutinised on their performance against all aspects of these criteria.

In summary, while recognising that the core metrics form only part of the overall body of evidence that will be considered by the panels, they are arguably the main way in which panels will identify excellence and also causes for concern. We thus believe that the lack of alignment that we have identified is concerning and more thought should be given as to how better align assessment criteria and the intended metrics.

Question 2 (Chapter 3)

A) How should we include a highly skilled employment metric as part of the TEF?

We note that the way that the question above is framed is somewhat different to the position taken in the text on p.22 (para. 73) which states the consultation will be used in order to inform a final decision on whether or not a highly skilled employment measure should be included as one of the TEF metrics for Year Two – the question thus arguably precludes certain answers being offered.

That said, we welcome the decision to use ‘an adapted version of the UK PI employment indicator which measures the proportion of the provider’s graduates who are working or studying (or both) six months after graduation, as a percentage of all graduates’ (p.21, para. 69), presuming this is a comparison between graduates from an individual institution and all graduates. We welcome that this adapted version of the UK PI indicator covers a wide range of outcomes, including self-employment, unpaid work and voluntary work. The fact it also includes ‘those who are retired, in ill health, looking after the home or family, or taking time out to travel or similar’ (p.22, para. 69) is also positive.

The Technical Consultation goes on to note that such data does not capture whether graduates are using the skills gained through HE, hence the proposal that an additional metric is needed to establish the proportion of graduates in 'highly skilled employment' (page 22, para. 70). This could be problematic, as a) this will be very hard to define and we are not optimistic that there is an 'easily interpreted, widely used and readily available definition' to draw upon (p.22, para. 71), b) skills gained at university are not simply academic or employment focused, but include interpersonal skills such as social and presentation skills, and c) it does not reflect the fact that a student may enter 'lower-skilled' employment post-university for a number of reasons before entering more highly-skilled employment later on (as recognised in the report on page 22, para 72), meaning the six months' time frame is a potential problem. This latter context must be recognised if an additional metric is utilised, with different types of providers also accounted for. For example, part time, distance, and online providers often have students who are already in employment and the consultation does not make reference to how this will be recognised).

Later iterations of TEF also need to take account of the differences between disciplines, with clear understanding that some disciplines have a more direct route to employment than others.

It is important that whatever means is chosen to measure this metric, it does not replicate existing bias in favour of HE institutions with the reputational pull and/or location that enables them to provide their students with, often easy, access to high paying jobs. This would not be an accurate measure of high quality teaching. The temptation to put further weighting onto graduate salaries should be resisted and a more sophisticated metric needs to be developed.

B) If included as a core metric, should we adopt employment in Standard Occupational Classification (SOC) groups 1-3 as a measure of graduates entering highly skilled jobs?

Yes No Not sure

If this must be measured, then the SOC measure is as good as any. However, as the report notes, not all occupations listed within SOC are 'high value graduate jobs' (p.22) meaning the measure needs to be refined. That said, and as suggested in relation to A), this will be problematic as individuals have differing definitions of what a highly skilled job is. It also fails to reflect the fact that students may not necessarily want a 'high value' graduate job and that people study for different reasons, not necessarily focusing on an immediate 'high value' job, or may suffer from mental or physical health conditions that preclude them from working in highly-pressured roles.

It is also important that this metric is not biased towards STEM and business occupations.

C) Do you agree with our proposal to include all graduates in the calculation of the employment/destination metrics?

Yes No Not sure

Please outline your reasons and suggest any alternatives.

Yes, as long as the wide range of outcomes described in A) is utilised.

Question 3 (Chapter 3)

A) Do you agree with the proposed approach for setting benchmarks?

Yes No Not sure

It is important to have some mechanism that addresses the differences in geography, mission and history between institutions. However, it may take a little time for the sector to become entirely convinced that these are fair.

Nevertheless, some questions about the choice of exclusions in Figure 5 (p. 23-24) remain. The White Paper excludes ethnicity, gender and disability from benchmarking for non-continuation on the grounds that universities are expected to mitigate the impact of historical and complex combinations of socio-economic and cultural issues that affect the abilities of different kinds of students to complete their programmes. The same is true for historic patterns of discrimination against disabled people in gaining employment. It may be useful to encourage HEIs to increase their efforts in this regard, but it may also have unintended consequences.

B) Do you agree with the proposed approach for flagging significant differences between indicator and benchmark (where differences exceed 2 standard deviations and 2 percentage points)?

Yes No Not sure

Please outline your reasons if you disagree.

The proposed approach is slightly less demanding than the current UKPI framework, which uses 3 standard deviations. The proposal means 95% of all cases would be

considered as falling within the normal distribution, which means an increase in the potential number of false flags. The inclusion of the z-scores is good, as it should help assessors distinguish between areas of legitimate interest and statistical anomalies.

The macro question relates to the interaction between the assessors' judgement and 'false flags' which are thrown up by the metrics. The inference from the present documentation is that the assessors will provide a check on false flags but there is an inferred assumption that the metrics will trump the judgement of the assessors – clarity here (p.24) is crucial.

Question 4 (Chapter 3)

Do you agree that TEF metrics should be averaged over the most recent three years of available data?

Yes No Not sure

Please outline your reasons and suggest alternatives.

Whilst we understand the focus on a three-year cycle, as on the surface this maps on to most degree programmes, the choice of a three-year average poses some problems. It covers the standard three-year undergraduate degree programme, but not the full lifecycle of students who take a placement or a study abroad year. On the other hand, a longer period would mean that the metrics would remain more stable, as it would take longer for negative cohort elements to work their way out of the metric as well as for policy innovations to impact fully upon it. There is also some evidence that there is a bias in favour of the status quo in this sort of metric, as the result itself influences the assumptions of the people who are completing the surveys that make it up. In the end, the choice made here depends on what kind of measure we want the TEF score to be.

Question 5 (Chapter 3)

Do you agree the metrics should be split by the characteristics proposed above?

Yes No Not sure

Please outline your reasons and suggest alternatives.

Our one caveat centres on disaggregation in terms of 'disability'. In a number of universities large number of students have dyslexia or dyspraxia, which explicitly impact on their learning whereas other disabilities may have less clear impacts. We thus would suggest that there is consideration of disaggregating disabilities to emphasise those cohorts of students who require explicit support for their learning as opposed to those who may require a more generalised form of support.

Question 6 (Chapter 3)

Do you agree with the contextual information that will be used to support TEF assessments proposed above?

Yes No Not sure

Please outline your reasons and suggest any alternatives or additions.

In terms of subject of study, we understand the desire to use the 18 subject areas but we are concerned that these actually mask considerable disparities. Even within these subject areas there can be wide variation in entry requirements and mappings to professional association needs etc. We also question the value of the subject information given that it will be superseded by more granular subject analysis by 2019. We question why this process does not begin with more granular contextual analysis from the outset (i.e. subdividing the subjects). We would also propose capturing information in relation to parental education, i.e. whether or not the student is the first person in their family to attend higher education.

In terms of the other contextual information, the student population characteristics data appears uncontroversial, especially as they already appeared earlier in the section. There is however some concern in relation to some of the information which is to be provided by the 'data maps'. There is a question as to how this information is to be defined. For example, how do we classify where students grew up? Is it their home address at point of entry? We are also unsure what level of geographical organisation is being measured (local authority, town, nation, or postcode). The same might be said for where students find employment. There is a danger that those students who choose to live in the place where they studied could inadvertently penalise institutions which are based in areas of high unemployment? Similarly, if the student happens to remain in an area of high employment the institution may similarly be 'rewarded'. There needs to be clarification as to the purpose of the data maps if universities are to have an understanding of how to explain and account for such maps.

Question 7 (Chapter 3)

A) Do you agree with the proposed approach for the provider submission?

Yes No Not sure

On the surface we welcome the desire to approach this as a 'light touch exercise', and an approach which combines metrics with accompanying textual commentary seems sound. Yet there is a real danger that a process which is designed to evaluate teaching quality – often a 'lived experience' for students and staff – is to be exclusively assessed through metrics and accompanying paperwork. In the short term, whilst we understand the need to pilot the TEF (hence the institutional level approach), we wonder about how effective this will be at offering meaningful information to potential students, which is, after all, the object of the exercise. It is crucial that the provider submissions are able to capture the diversity of provision, articulate best practice, inform students meaningfully and celebrate excellence. We question if an approach which is absent of site visits can achieve this.

We would welcome more thought and creativity regarding how diversity of practice and innovation in teaching can best be captured. The metrics currently being used are not best able to provide a rich account of unique aspects of teaching at different provider institutions.

B) Do you agree with the proposed 15 page limit?

Yes No Not sure

Please explain your reasons and outline any alternative suggestions.

On the surface a fifteen page limit will ensure that the TEF will be a light touch process that does not distract universities from the actual activities that they should be engaged in: teaching, learning and research. Yet there are two important caveats to this: a) we wonder if length is a sound criteria for judging the burden on an institution. Writing a 'short document' is not necessarily any less time consuming than writing a longer text. Furthermore, the details of what is required are crucial, with the nature of the prescribed template being most salient. b) A further issue is whether or not institutions will be allowed to provide supporting evidence and documentation (the inference is not). Yet we wonder if the fifteen page limit will allow institutions to adequately capture the diversity of their provision in fifteen pages, particularly if this does not allow an appendix.

Question 8 (Chapter 3)

Without the list becoming exhaustive or prescriptive, we are keen to ensure that the examples of additional evidence included in Figure 6 reflect a diversity of approaches to delivery. Do you agree with the examples?

Yes No Not sure

Please outline your reasons and suggest any additions or alternatives?

We agree that the examples listed provide an initial indication of some of the types of evidence that may prove useful to assessors and panels in making judgements. We recognise that the list is not intended to be exhaustive, but recognise that the choice of items that are included within this will be a powerful steer to institutions submitting for assessment and to TEF panel members and assessors.

In terms of additions, we would welcome the possibility of considering ‘weighted’ contact hours. This would potentially mitigate against the potential distorting effect of rewarding high contact hours at the expense of those institutions, often in the social sciences, which offer high quality, small group teaching for sound pedagogical reasons. It is unclear whether universities would continue to use contact hours as calculated by KIS. Using this methodology, only select modules actually count, not all. In addition, the enormously broad range of learning events which are not linked to any modules, such as invited talks and research seminars, are not captured. Finally, whilst we acknowledge that student dissatisfaction in terms of perceived ‘value for money’ is often linked to contact hours, we note that it is unclear how module contact time relates to module attainment.

However, it is our view that the lists may: (i) have an unintended bias towards some disciplines and; (ii) not fully align to the criteria for teaching excellence set out in Figure 4.

1. Unintended bias

We observe that a number of the possible examples of evidence identified are more appropriate to disciplines in STEM or Health and less relevant to the social sciences. These include PSRB recognition and number of contact hours. Employer engagement is also found more often in more vocational courses than in degrees that facilitate a wider range of employment, as is typical in the social sciences. We also note that given the higher estates and capital costs of STEM and health subjects, that these will more readily provide quantitative evidence for investment in the teaching and learning infrastructure. As stated above, there is a risk that the list of examples will unintentionally steer provider, assessor and panel members to focus on data that is more relevant to some discipline than others. In the longer-term, this could also affect planning and funding decisions by individual institutions.

This leads into a broader issue of whether and how intra-institutional variation is noted, since the structural disciplinary differences we note above will produce potentially misleading aggregate figures for students who normally only encounter one part of an institution’s provision. Solutions to this might include some indication of the standard deviation on core metrics, or separate figures by broad categories, in addition to an overall weighted institutional figure.

There is a case for balancing these by including examples such as:

Teaching Quality

- Extent to which academic staff are supported to engage in pedagogic research and develop innovative approaches to teaching that are appropriate to their disciplinary contexts.

Learning Environment

- Extent, nature and opportunities for students to engage with external organisations, including local and global civic organisations.

Student Outcomes and Learning Gain

- Extent to which students are provided with opportunities to engage with civil society and develop as active global citizens.

2. Alignment to Criteria in Figure 4

We are concerned that a number of the possible examples of evidence are not aligned to the criteria listed in Figure 4. We are most concerned with the following two examples.

- Teaching Quality. Contact hours as an indicator of teaching quality as defined by the criteria. It is not clear how the number of teaching hours provides an indication of the “stimulation and challenge” of learning; the extent to which course design is “effective in stretching students”; or the quality of assessment and feedback.
- Student Outcomes and Learning Gain. It is unclear how the implementation of GPA by a provider is likely to demonstrate evidence for the criteria relating to progression to further employment or study; preparation for ‘personal and professional life’ or differential attainment by background. While the use of assessment data can be of use in developing effective practices in each of these areas, we are aware of no evidence that suggests that a GPA based methodology provides an additional benefit.

Question 9 (Chapter 4)

A) Do you think the TEF should issue commendations?

Yes No Not sure

B) If so, do you agree with the areas identified above?

Yes No Not sure

Please indicate if you have any additional or alternative suggestions for areas that might be covered by commendations.

Whilst we fully understand that it is valuable to differentiate providers from one another in terms of quality, we are concerned that there are assumptions built into

the analysis (p.33, para. 119) that the distribution of ratings would follow a bell curve. We believe it is possible that all institutions will strive for excellent ratings and may achieve them. Furthermore, the inference of the bell curve thinking is that assessors will be strongly beholden to the values within the metrics rather than those contained within the accompanying textual commentary.

In terms of the list in para 124 (p.34), we can see merit in many of these categories but would add some caveats. The first – research-led teaching – is a highly complex issue and one which would need to be clearly specified to institutions. There is little consensus within the pedagogy as to what research-led teaching actually is and indeed of the values which students see in it. We completely support the notion that teaching and research can be seen as working symbiotically but what assessors are looking for will need to be clearly specified. We are less convinced of the merits of ‘business engagement’ unless it is central to the curriculum. Furthermore, we are concerned that a focus on metrics may actually be ill-suited to reach judgements of excellence in any of these measures – how, for example, would innovation be captured through metrics? The inference therefore is that the textual commentary would be the crucial determinant of the granting of a commended status. More clarity on this needs to be provided.

Finally, we are concerned by the inference that only 5-10% of providers ‘would be likely’ to be commended. We see this as something that should, in principle, be open to more institutions than this range suggests. Why not encourage sharing of best practice and outstanding achievements more widely?

Question 10 (Chapter 4)

Do you agree with the assessment process proposed?

Yes No Not sure

Please outline your reasons and any alternative suggestions. The proposed process is set within a relatively tight timescale, reflected in the key dates included in Annex B. Responses should be framed within this context.

In line with concerns expressed above about the metrics of the TEF aligning more suitably with STEM and health degrees, we think it is crucial to ensure that social science degrees are assessed by TEF assessors with extensive experience within social science teaching. Ultimately when TEF moves to discipline level assessment, possibly from 2019 onwards, assessors should be drawn from specific disciplines. This will ensure that a more accurate assessment, reflective of the realities of individual disciplines, can be made drawing on the experience of the assessor and a provider’s metrics and submission.

We also question the lack of an appeals process after the Year 2 TEF outcome on two grounds:

i) Whilst all providers will in the first year be awarded ‘Meets Expectations’, it is simply unfair to not provide an appeals process, as would be normal in other

assessment processes. This is particularly important given that the panel as a whole is inevitably going to be made up of assessors from a wide array of disciplines and perspectives which may make agreement more taxing.

ii) The inclusion of an appeals process could be helpful for the TEF Panel and BIS more generally in reviewing the effectiveness or otherwise of TEF.

Question 11 (Chapter 4)

Do you agree that in the case of providers with less than three years of core metrics, the duration of the award should reflect the number of years of core metrics available?

Yes No Not sure

Please outline your reasons.

In principle we can see that it makes sense for the length of the awards to be shorter than three years' duration in those cases where the core metrics are available for less than three years. However, we would be concerned that new providers with limited data had the opportunity to provide information that demonstrated excellence purely on the basis of one or two years of data. This is important because it guards against a situation in which, for example, an HE provider with three years of data that included one year of weaker performance would be (in essence) penalised in comparison to a new provider with one year of data that was acceptable, even though the former had many more years of high performing data.

Question 12 (Chapter 5)

Do you agree with the descriptions of the different TEF ratings proposed in Figure 9?

Yes No Not sure

Please outline your reasons and any alternative suggestions.

We think that the description of TEF ratings should be the same for all 3 categories: Meets Expectations; Excellent; Outstanding. There seems to be a discontinuity between the description for 'Meets Expectations' and the other categories. In particular, the criteria for 'Meets Expectations' do not map clearly onto the criteria within Figure 1 (which form the foundation of Q1)

It would be better to standardise all categories and state that the TEF ratings describe teaching quality; the learning environment; student outcomes and learning gain as drawn from assessment criteria in Figure 4.

We also feel that each of the criteria should be more explicit and provide some exposition. In short, accepting that teaching quality, the learning environment and

student outcomes and learning gain are the correct measures, the descriptors should explicitly articulate what 'meets expectations', 'excellent' and 'outstanding' actually looks like in these three fields – a list is not enough.

Thank you for taking the time to let us have your views.

We do not intend to acknowledge receipt of individual responses unless you tick the box below.

Please acknowledge this reply

At BIS we carry out our research on many different topics and consultations. As your views are valuable to us, would you be happy for us to contact you again from time to time either for research or to send through consultation documents?

Yes

No

BIS/16/262/RF