**ROYAL STATISTICAL SOCIETY**

DATA | EVIDENCE | DECISIONS

12 Errol Street
London EC1Y 8LX
020 7638 8998
www.rss.org.uk

Ed Humpherson
Director General for Regulation
UK Statistics Authority
1 Drummond Gate
London
SW1V 2QQ                                                            05 March 2019

Dear Ed,

We are writing to you on behalf of the Royal Statistical Society (RSS) to express our serious concerns about the Teaching Excellence and Student Outcomes Framework (TEF) produced by the Department for Education/Office for Students (DfE/OfS). The TEF is in large part a statistical artefact, and we are concerned that it does not meet the standards of trustworthiness, quality and value that the public might expect. Indeed, the statistical issues are so major that, in our view, the TEF is likely to mislead the public and, in particular, mislead students who use TEF to inform their university choices.

The RSS has written to you before about the TEF and, in doing so, enclosed the key points from our previous consultation responses to the provider-level and subject-level TEF exercises. We are not confident that all of our statistical points have been adequately addressed in relation to the TEF.

We are grateful for your efforts to foster communication between stakeholders, such as the RSS, DfE and OfS. Furthermore, we have welcomed the opportunity to participate in "The Independent Review of the TEF" that is being led by Dame Shirley Pearce. We duly met with Dame Shirley and her team in London in January and explained our concerns. The attached document is our submission to the Independent Review and highlights some of the key problems with the TEF. As you will see, we believe there are several areas where the TEF either does not adhere to, or transgresses, the UK Statistics Authority's Code of Practice, and we have explicitly referred to these in the document we are sending to the Independent Review.

We would particularly draw your attention to two key issues.

1. Transparency and Reproducibility (Section D)

As far as we can discover, there is no complete, transparent description of how the TEF awards are made - especially in relation to the process by which statistical information and flags are provided to the TEF panels. Partial descriptions exist, together with some spreadsheets. However, several of the TEF recipe assumptions that we *can* see are not able to be properly evaluated due to the lack of transparency. If TEF is to continue, we would argue that it must be made fully transparent and easy to check. The whole, specific

and detailed analysis pipeline should be published, making fully clear the methods and software that was used, plus as much data that can be revealed, as well as a proper detailed explanation of how it all works. If there are reasons why some lower level data should not be published, then these should set out and clearly explained.

Our belief is that full transparency would help all concerned. Once transparent, if everything adheres to best practice and can be validated, then this could help gradually establish the trustworthiness of TEF.

2. The multiple hypothesis testing problem. (Section C(vi))

The TEF process produces flags, the collection of which are used to inform the process of discerning the final TEF award. The flags are produced by assessing the size of Z-scores and comparing them to a 'standard' critical value. However, the TEF computes a large number of Z-scores and this is equivalent to conducting a multiple hypothesis test. In such cases, the Z-scores should not be compared to a 'standard' critical value, but typically one that is much larger (and there are various methods to do this, such as Bonferroni correction or using false discovery rate assessment). Using a single test critical value instead of a multiple test value is a serious statistical mistake, which will result in far too many indicators being spuriously flagged. Hence, the RSS believes that all the TEF awards made so far have been based on seriously flawed inputs and that, because of this, all TEF awards made to date are invalid.
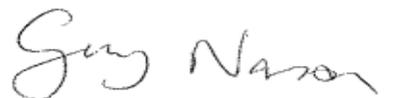
It is hard to discern the exact situation due to the transparency problems mentioned above. However, during our listening session, as part of Dame Shirley's Independent Review, we questioned members from the government department on this point and they confirmed that they were not using methods that appropriately controlled the size of the multiple hypothesis test.

Based on the above, and our written submission to Dame Shirley's review, we would ask the Office for Statistics Regulation to consider the validity of the TEF, and to rule on whether TEF does actually provide the public with information which is trustworthy, of high quality and value.

Yours sincerely,

Deborah Ashby                                    Guy Nason
President, RSS                                    Vice President, RSS