

GuildHE Final response to:

OfS Consultation on a new approach to regulating student outcomes

17 March 2022



About GuildHE

[GuildHE](#) is an officially recognised representative body for UK Higher Education. Our members include universities, university colleges, further education colleges and specialist institutions from both the traditional and private sectors. Member institutions include some major providers in professional subject areas including art, design and media, music and the performing arts; agriculture and food; education; business and law; health and sports.

Overarching comments

We wanted to highlight a number of key points by way of summary which relate to this consultation and the data indicators consultation.

We have real concerns that publishing three major – and highly complex - consultations at the same time, with only an eight week consultation period, has placed significant pressure on the higher education sector as a whole and specifically disadvantaged smaller providers. We have received numerous comments from members about feeling totally overwhelmed by the scale and complexity of the consultations, particularly given other consultations and reforms that are also happening. We further expect that institutions have had to prioritise the consultation responses and so the OfS will not receive either as many responses, or that those responses will have been subject to the level of consideration, appropriate for the scale of the changes proposed. We would strongly suggest that the OfS continue to discuss the proposals, implications and next steps with institutions and their representative bodies as it develops its proposals.

Regulatory burden

1. We believe that the regulatory burden resulting from B3 is disproportionate to the scale of the problem. We recognise that the consultation is not seeking to gather additional data, and has sought to mitigate some of the impact on smaller providers by giving providers processed work-books of data. However, we believe that the thousands of indicators – when considering the different levels, split metrics, student characteristics, subject and so on – would result in a significant data analysis burden for both the regulator and for providers and will specifically disadvantage smaller providers in the sector who do not have teams of data practitioners to scrutinise the data workbooks nor have access to additional data/research to provide alternative analysis when responding to the effects of the data approach taken by the OfS. Due to the number of students in many of the split metrics it is also likely to result in data that is not statistically robust in many smaller providers.
2. For a regulator to be effective it is important to focus on the biggest areas of risk and therefore to focus on a smaller number of indicators. It is noteworthy that in Australia TEQSA initially developed 48 risk metrics before deciding that, because of burden and complexity, this did not represent effective regulation and reduced this to 11 key indicators. We would encourage the OfS to be similarly focused on the greatest areas of risk and public interest.
3. The mass of indicators would be a particular challenge for smaller providers with fewer/no specialist data staff working to analyse the data. Providing huge data workbooks may give the impression of transparency and openness, but will have the consequence of diverting

© GuildHE (This information may be freely used and copied for non-commercial purposes, provided that the source is acknowledged.)

considerable staff resource away from established processes of quality management such as annual monitoring and periodic review. This problem is likely to be exacerbated in smaller organisations who often do not have dedicated specialist data teams and may result in additional financial burden for those that will need to buy in expertise. These smaller providers make a significant majority of the providers on the OfS' register and for whom the economies of scale mean they will have significant additional financial costs associated with regulation. Data analysis is a science, and therefore providers who can hire those with professional analytical skills will fare better in this approach to regulation. We therefore believe that a regulatory approach based mainly on data is unfair to smaller providers.

4. The mass of data means that it is impossible to see the wood for the trees and essentially means that OfS can decide which institutions it wants to investigate because every institution is likely to be below an indicator somewhere. We believe that the question of the processes by which the OfS prioritises where it focuses its resources is one of the key areas of this consultation. The need for a clear, transparent and robust prioritisation process, free from political interference will be key to maintaining the confidence in the regulator. We would also argue that whilst publishing the data would be beneficial in terms of public accountability, just publishing too large a dataset would result in obfuscation rather than transparency and will encourage those with personal, political or commercial agendas to manipulate the data to their own will, ignoring the nuances that lay within the sector and undermining the risk based and contextual judgements the OfS intends to make outside of the raw data.
5. The OfS also has duties to have regard to the need to use its own resources efficiently, effectively and economically and to have regard to best regulatory practice. We believe that the scale of data analysis and associated burden is not an efficient and effective use of OfS resources, nor is it proportionate to the regulatory risk.
6. The OfS has put together an approach to regulate provision based primarily on the current traditional model of HE delivery. The HTQ and LLE policies currently being implemented throughout the sector will have a dramatic impact on the data available about institutional provision and as such will ultimately impact on the approach as set out here. This means that the sector will have yet another major overhaul of data collection and onward use in the next 3-5 years which further increases regulatory burden on the sector. It is not the number of data that need to be collected which impacts on regulatory burden, as the OfS suggests in this set of consultations (and previous ones on their overall data strategy), but the uncertainty and continual changes being made to the regulatory approach.
7. The underlying data that drives the B3 approach as set out here is predicated on a design structure that is incredibly complex and knits together different data sets from the HESA and ILR record that do not always have the same definitions or contexts. We would like the OfS to work much harder on the original principles of the Data Futures project of collect once, use often and better align the ILR with the HESA return. Most HEIs now have to return both datasets and this significantly increases burden on the sector (especially as the definitions are not always aligned). As the OfS regulates providers not provision - it should be doing more to reconcile the data and minimise the regulatory overlap between itself and the ESFA/IfATE.

8. Furthermore, because the data design is so complex, the need to change and adapt to new approaches to HE delivery, the OfS not giving enough time in these consultation windows for the algorithm's to be fully scrutinised, and because it is intending to publish these data sets in full we believe that there should be a Data Governance Group made up of expert data practitioners inside HE and external organisations such as the ONS to guide the development of the use of data in regulation going forward and provide ongoing scrutiny and accountability to the sector. This is especially important as there are political decisions made about what data is used and how and there should be a mechanism for ongoing discussion of the approach taken. The developments in the delivery of HE to more flexible learning as posed in the HTQ and LLE policies will also mean the approach to data may need to change in the coming years and having a sector data group in place will be able to support the initial thinking and development of proposals as time goes on.

Prioritisation

9. We believe that if the OfS does decide on this approach to B3 then there needs to be a robust, transparent process for deciding how the OfS will prioritise where it will focus its activities for the year. We would argue that there should be an independent process for deciding on priority areas that includes students – as the main parties affected – as well as academic experts and peers from within institutions. There is also a strong argument that this priority focus should not change every year but rather it would be better to focus on particular areas for a couple of years to assess the impact of action and ensure change is embedded. It will of course be important to consider how the lag in time in the data might be considered in relation to action plans.
10. In terms of the factors on which prioritisation could be based we would argue that distance from threshold, student population, statistical certainty and mode or level study could all be reasonable justifications for prioritisation. However, we do not agree that thematic approach or random selection are appropriate and are more akin to the regulator going on a fishing expedition.

Data as a proxy for quality

11. We wanted to take the opportunity to restate our objections to an over-reliance on data as a proxy for quality, and would highlight the recent HEPI [blog](#) on the issue. It is important to restate that continuation and completion are not the only measures of quality, and at best are poor proxies for it. Specifically retention and completion as measures, at least in an un-benchmarked form, can be problematic as they fail to take account of the life circumstances which may result in certain student groups not completing their studies.
12. We welcome the recognition of the need to consider the statistical significance of the data for providers, and particularly the impact that this will have on smaller providers with smaller – and therefore more volatile - datasets. We would strongly restate the importance of the principle of not forming a negative judgement based on small datasets that have high degrees of statistical uncertainty and it will be important to consider how this data is presented as part

of the transparent approach to publishing datasets and ensuring appropriate caveats are highlighted. We would particularly wish to see a mechanism that enables the public to understand the wider context of the provider and for the OfS to make sure that external audiences are not manipulating the data in ways that go against the principles of sound data analysis.

Progression Indicator

13. We also have particular concerns about the use of progression data for whether graduates go on to professional and managerial jobs. The role of the regulator is to regulate on things over which an institution has control. There are structural inequalities within society that impact on the type and level of jobs that people are appointed to. There are also likely to be regional differences in current skills needs and not considering this factor by removing benchmarks might negatively impact on Government plans to “level-up” regional economies. **Not benchmarking this data for progression to professional and managerial employment will unfairly penalise institutions for issues beyond their control.** We therefore propose that the OfS must either remove this Indicator or retain benchmarked data on the progression measure. We would also suggest that the OfS should undertake a full equality impact assessment on the impact of not benchmarking data by particular groups of students, and in particular the impact on mature learners, facing very particular challenges.

14. The Graduate Outcomes dataset was described as an experimental dataset in June 2020 and the survey was not designed for a regulatory purpose. We believe that there are still significant flaws in the dataset. The OfS’ as a provider of Official Statistics has a responsibility to meet the [Code of Practice for Statistics](#) which emphasises the need for “Suitable data sources”. We believe that there are significant questions as to whether Graduate Outcome meets this threshold yet based on the low response rate and the imprecise way in which data is gathered and coded. We would also question whether the census-based nature of the survey – gathering what a student is doing on a particular date in time recognises the portfolio nature of many graduates careers and therefore the survey is not “...the most appropriate data to meet intended uses.” The current Graduate Outcomes was also gathered in the midst of a pandemic, which had particular impacts on some industries, with the creative industries to a large extent being shut for much of the period and so reducing career options for these graduates.

15. We also have concerns about the definitions of professional and managerial jobs as defined by SOC Codes. There are a number of graduate-level jobs – such as in early-years education, working in care homes, veterinary nurses, costume design, fashion merchandising and hair and make-up for TV and Film as well as the challenges of those in portfolio careers such as in the creative industries – that are not currently recognised at this level that should be. There will be many more careers in a similar position and this is also likely to continue to change in the future.

16. We also believe – and have many examples of - the practical challenges of graduates self-reporting their jobs, which is then coded by a non-expert, through Graduate Outcomes means it can result in mis-coding of outcomes.

17. Linking the Progression Indicator to the SOC Codes creates an unnecessarily inflexible system. The SOC Codes are currently only reviewed every 10 years and also feed into other government priorities – such as for visa application purposes - and so are likely to face different political pressures. Some careers will become graduate-level and new and innovative career paths will be created within the 10-year period before the next review. We would therefore propose that if the OfS does insist on using the progression metric that you only use the SOC Codes as a starting point, but have the possibility to create an additional list of graduate-level jobs that the SOC Codes haven't yet caught up with yet.

Case study: Agriculture and land-based subjects

18. There are some subjects where a significant proportion of the students or indicators are below the numerical threshold. For example, looking at the Progression metric for Agriculture, Food and Related Subjects where 51.4% of students are currently at or below the numerical threshold of progression to professional and managerial jobs. The agricultural and land-based industries have well-developed qualification routes from both further education and higher education but the different skill levels for different jobs are often less well reflected in the job titles and so these graduates are disproportionately impacted by mis-coding in Graduate Outcomes returns.
19. It is important that the OfS consider how this data is presented publicly so as not to impact on a strategically important sector of the economy where future food security is likely to be an increasing government priority. It would be worth considering whether there should be a separate numerical threshold for subjects or indicators that are statistically significant distances from the sector-wide numerical threshold or whether it should be more explicitly mentioned in the contextual information. It will be necessary to ensure that reforms are directed at the key areas of major risk to the sector, rather than unintentionally catching up specific subjects by mistake.

Context will be key

20. As highlighted above we do not agree with an over-reliance on purely quantitative measures of quality and therefore welcome the recognition of the role of taking into consideration contextual factors. We do believe that there needs to be more detail provided on the kind of evidence that would be considered as part of this process, for example we believe that the use of contextual information around IMD and ABC characteristics should form part of the benchmark calculation.
21. It will be important that this contextual information is considered alongside the data, rather than being seen as an after-thought and not given due consideration when forming the judgement. There should be transparency in the process of how this will occur, for example it should not just be based on the decision of OfS officials but should be considered through processes – such as the OfS QAC – that include academic experts and peers. We would also want some clarity on how context will be presented alongside the publishing of data workbooks. We would argue that it would need to be presented in some form so as not to result in confusion for people looking at the data.

ESG Compliance

22. Finally, we would restate that the OfS will need to consider the international reputational impact of the quality system and how it makes the case for England's approach in other parts of the world. The UK has a strong international reputation for delivering excellent higher education and this is seen through the number of international students that we recruit every year. This is in part based on the sector's reputation for high quality education. However, the OfS's new quality system – based on data, no cyclical review and no peer review - is looking ever-more out of step with European partners and the European Standards and Guidelines. We may like to think of ourselves as innovating and leading the way, but with a review of whether we are compliant with these guidelines imminent it may further solidify England's position as an outlier and separate us from the EU student market and also impact on our international reputation. If the OfS is going to insist on such a radical approach to assessing quality then it will need to take a more proactive role in publicising of benefits and rigor of this approach to key international markets and their sister regulators.

GuildHE Proposals

In response to the commentary above we are proposing a number of specific amendments to the proposals:

Key Indicators

We propose that the OfS only publicly publishes the data for the three main Indicators (continuation, completion and progression) and that these are only split (publicly) by mode of study (full-time, part-time and apprenticeship). This would provide 9 Key Performance Indicators that would be publicly available. Additionally, the OfS may decide in a particular year to look at a particular split metric (including based on level of study, protected characteristic or subject) but that these would be limited to a maximum of three splits in any year and these could be made publicly available. This set of (a maximum of) 12 Indicators would provide a more manageable set of data to focus public attention on. The OfS could still consider the full proposed dataset and would provide this to institutions for their own internal quality processes and for it to be used internally through the TEF process (but without publication).

Prioritisation

The criteria on which the prioritisation process would be considered should be published at least six months prior to publication and should be agreed by committee/advisory group including academic peers and students. This process would enable institutions to consider their data based on the criteria and develop their own internal action plans and the additional transparency in the process would avoid a perception of politicization of the process.

Progressions Indicator

We have particular concerns about the progression indicator, both on a principled basis – as something outside the control of the institution and not measuring quality - but also on a practical basis. **We would therefore propose that you do not use the progression indicator.** We do however recognise that the OfS has expressed their policy intent of seeking to measure

successful graduate outcomes, we would therefore propose that if you insist on proceeding with the use of the progression indicator that it needs to be strengthened to ensure that it is a robust Indicator and would recommend that the OfS should undertake the following actions:

- i. Change the wording of B3.3 from “must deliver” to “must enable” in relation to positive outcomes for students on its higher education courses.
- ii. Develop a list of Professional and Managerial jobs to sit alongside the formal SOC Codes where the skills requirements of a career have developed, new jobs been created or more nuance is required.
- iii. Further enhance the Graduate Outcomes survey to ensure better training for those calling graduates to ensure that careers are properly coded.
- iv. Graduate Outcomes data to only be used where there is a 50% response rate – similar to the NSS data.
- v. Use the additional Graduate Outcomes questions (My current work fits with my future plans; My current work is meaningful and important to me; I am using what I learned during my studies in my current work)) alongside the SOC 1-3 metric to ensure a more rounded consideration linked to the expectations of the individual. To make the data from these questions more robust they should be made mandatory and elevated higher in the survey to reflect their regulatory.

Finally, we wanted to raise the more intangible question of trust. Many of the responses requiring additional information or transparency in OfS processes, or wider outcries and consternation about the proposals, stem from an erosion in trust between the regulator and the sector. In part this arises from the experiences of the last few years with overly legalistic and confrontational communications, lack of genuine consultation and consideration of views of those that the OfS regulates as well as riding roughshod over questions of UK and European regulatory cohesion and international reputation. There will always be a tension in relationships between the regulated and regulator, but this relationship does need to be built on trust. We endorse the view of the NAO that the OfS needs to communicate more effectively with the sector to build trust and recommend that the OfS considers what more it might do to rebuild and strengthen its relationship with the higher education sector.

Response to specific consultation questions

Question 1: Are there aspects of the proposals you found unclear? If so, please specify which, and tell us why.

We recognise that students value successful outcomes and that getting a good job is seen as part of that. However, the role of the regulator is to regulate on things over which an institution has control. We therefore continue to have grave concerns about trying to hold providers to account for whether graduates go on to achieve “professional and managerial level jobs” and we do not believe that it is actually a measure of quality or success of the course. There are structural inequalities within society that impact on the type and level of jobs that people are appointed to. There are also likely to be regional differences in current skills needs. We believe that the provision of this information for students is an important step to ensure greater transparency as part of their decision-making process on which course and institution to apply to, but that it shouldn't be used as a regulatory enforcement tool due to the lack of control that institutions have over the metric and the potential unintended consequences that may result. We believe that this metric is a poor proxy for the quality of the course and should be removed.

We are also unclear whether the OfS has given due consideration to the impact of the proposed changes on the sector's international reputation of the quality system. The UK has a strong international reputation for delivering excellent higher education and this is seen through the number of international students that we recruit every year. This is in part based on the sector's reputation for high quality education. However, the OfS's new quality system – based on data, no cyclical review and no peer review - is looking ever-more out of step with European partners and the European Standards and Guidelines. We may like to think of ourselves as innovative and that we are leading the way, but with a review of whether we are compliant with these guidelines imminent it may further solidify England's position as an outlier and separate us from the EU student market and also impact on our international reputation. If the OfS is going to insist on such a radical approach to assessing quality then it will need to take a more proactive role in publicizing of benefits and rigor of this approach to key international markets and their sister regulators.

The revised wording for condition B3 states five years of cohort data and paragraph 81 says four-year time series. We are not clear on why there is this difference. We would recommend four years for all, and this would align with the TEF cycle.

Question 2: In your view, are there ways in which the objectives of this consultation (as set out in paragraph 7) could be delivered more efficiently or effectively than proposed here?

Key Indicators

We propose that the OfS only publicly publishes the data for the three main Indicators (continuation, completion and progression) and that these are only split (publicly) by mode of study (full-time, part-time and apprenticeship). This would provide 9 Key Performance Indicators that would be publicly available. Additionally, the OfS may decide in a particular year to look at a

particular split metric (including based on level of study, protected characteristic or subject) but that these would be limited to a maximum of three splits in any year and these could be made publicly available. This set of (a maximum of) 12 Indicators would provide a more manageable set of data to focus public attention on. The OfS could still consider the full proposed dataset and would provide this to institutions for their own internal quality processes.

Prioritisation

The criteria on which the prioritization process would be considered should be published at least six months prior to publication and should be agreed by committee/advisory group including academic peers and students. This process would enable institutions to consider their data based on the criteria and develop their own internal action plans and the additional transparency in the process would avoid a perception of politicization of the process.

Progressions Indicator

We have particular concerns about the progression indicator, both on a principled basis – as something outside the control of the institution and not measuring quality - but also on a practical basis. **We would therefore propose that you do not use the progression indicator.** We do however recognise that the OfS has expressed their policy intent of seeking to measure successful graduate outcomes, we would therefore propose that if you insist on proceeding with the use of the progression indicator that it needs to be strengthened to ensure that it is a robust Indicator and would recommend that the OfS should undertake the following actions:

- i. Change the wording of B3.3 from “must deliver” to “must enable” in relation to positive outcomes for students on its higher education courses.
- ii. Develop a list of Professional and Managerial jobs to sit alongside the formal SOC Codes where the skills requirements of a career have developed, new jobs been created or more nuance is required.
- iii. Further enhance the Graduate Outcomes survey to ensure better training for those calling graduates to ensure that careers are properly coded.
- iv. Graduate Outcomes data to only be used where there is a 50% response rate – similar to the NSS data.
- v. Use the additional Graduate Outcomes questions (My current work fits with my future plans; My current work is meaningful and important to me; I am using what I learned during my studies in my current work)) alongside the SOC 1-3 metric to ensure a more rounded consideration linked to the expectations of the individual. To make the data from these questions more robust they should be made mandatory and elevated higher in the survey to reflect their regulatory.

Additionally, we would propose that you identify a standardised student population for all measures and purposes. Utilising different student populations for different purposes (e.g. B3/APP/TEF) and for different measures increases complexity, reduces efficiency and is not felt to increase effectiveness.

Questions relating to Proposal 1: Revising condition B3 and associated guidance in the regulatory framework

Question 3: Do you agree or disagree that the proposed wording of condition B3 will enable the OfS to meet its policy objectives? If you disagree, what changes do you think are necessary to do so?

Agree

Disagree

Don't Know

We disagree with the wording of the proposed B3 Condition for a number of key reasons.

- **B3.1** we do not agree that this Condition should be extended to all provision that the providers delivers (such as non-prescribed HE courses, transnational education and further education qualifications). We remain concerned that this would result in a massive expansion in the OfS and regulatory burden, as well as regulatory overlap with other regulators, and question how this provision would be monitored. We would therefore propose limiting the condition to just the provision that is funded by the OfS/public funding. If the OfS does decide to include Non-Prescribed HE in the dataset this should be a separate dataset (either as Mode or Level of Study) and not just included in “Other Undergraduate”.
- **B3.3** states that “the provider must deliver positive outcomes for students on its higher education courses” – the OfS should either remove reference to progression to professional and managerial job, our preference, or change the words “must deliver” as delivering these jobs are not something that the provider has responsibility over and so cannot “deliver” and propose that you use the phrase “must enable” instead.
- **B3.4a** we believe that considering all the split indicators will result in too many indicators and take focus away from the OfS’s overall responsibility for focusing on the areas of highest risk. Each split indicator may appear reasonable when considered on its own, but as a whole they are completely overwhelming and would potentially result in large amounts of data remaining unanalysed, which is not good regulatory practice.
- **B3.4bi** We believe that in paragraph B3.4 b i. the word “justifies” should be replaced by “explains”. We believe the test of establishing justification of provider data is too strong, and all that could ever be established is that the context adequately explains the data in question.
- **B3.5fiii** we propose the deletion of this sub-section and removal of reference to professional and managerial employment as something not in the control of the provider.
- **B3.5k** we propose the deletion of this sub-section referring to split indicators to prevent an ever-growing mass of data and inability to focus on the key over-arching priorities. The OfS should consider how it addressed gaps in outcomes in a proportionate way.

The principles outlined in the bullet-points above would then feed through and amend the Guidance document as well.

Questions relating to Proposal 2: Constructing indicators to assess student outcomes

Question 4: Do you agree or disagree with the proposals for how we will construct a student outcome measures? Do you have any alternative suggestions?

Agree

Disagree

Don't Know

These proposals are totally unmanageable for both providers and the regulator and, when published, for the wider population. The proposals outlined would create thousands of indicators, split indicators and data-points – not the 48 stated by the OfS in the consultation. This would result in a huge regulatory burden – both for the OfS to assess and prioritise the data and for providers to go through the vast swathes of data being proposed.

Indicators, split-indicators and other data-points proposed include:

- Three overarching indicators for continuation, completion and progression
- Which will be split three ways for mode of study - full-time, part-time and apprenticeship
- Which would be further split another 18 ways at level of study - full-time and part-time modes: – other undergraduate – first degree – undergraduate with postgraduate elements – PGCE – postgraduate taught masters – other postgraduate – postgraduate research • apprenticeships: – undergraduate – postgraduate
- Which would each be further split another 8 ways by student population - age on entry to higher education course • disability • ethnicity • sex • domicile • eligibility for free school meals at Key Stage 4 (for young undergraduate students) • English Index of Multiple Deprivation (IMD) • quintile associations between characteristics of students (ABCS) quintile
- Which would be split a further 7 ways by taught, registered and partnership with levels within each category
- Which would then be split by subject level – up to 35 different ways at CAH2
- With further splits by type of course - full-time first degrees with an integrated foundation year separately from other full-time first degrees; Higher Technical Qualifications (HTQs) separately from other full-time first degrees; Level 4 and Level 5 courses separately, within our 'other undergraduate' grouping
- This will also be extended to non-prescribed courses
- Which could be supplemented by TNE data in the future

On a principle level we disagree with a quality and standards assessment approach that is over-reliant on data as a proxy for quality, and would highlight the recent HEPI [blog](#) on the issue. A process that is overly reliant on data ends up needing to create indicators on every possible way of breaking down the data, as shown by the thousands of proposed indicators and split indicators highlighted above. **Each split indicator may appear reasonable when considered on its own, but as a whole they are completely overwhelming and would potentially result in large amounts of data remaining unanalysed, which is not good regulatory practice.**

This mass of data ends up neither focusing on the areas of highest risk nor areas of greatest public concern and swamping both the regulator and providers with data so that it is impossible to see the woods for the trees. The mass of indicators would be a particular challenge for smaller providers with fewer specialist data teams working to analyse the data. The large number of indicators and split indicators means there needs to be a very robust process of prioritisation at the OfS. This problem is likely to be exacerbated in smaller organisations who often do not have dedicated specialist data teams and may result in additional financial burden for those that will need to buy in expertise. These smaller providers make a significant majority of the providers on the OfS' register and for whom the economies of scale mean they will have significant additional financial costs associated with regulation. Data analysis is a science, and therefore providers who can hire those with professional analytical skills will fare better in this approach to regulation. We therefore believe that a regulatory approach based mainly on data is unfair to smaller providers.

The OfS should focus on the areas of highest risk or greatest public concern and identify a limited number of indicators to focus attention where it is most needed. It is noteworthy that in Australia TEQSA initially developed 48 risk metrics before deciding that, because of complexity and burden, that this did not represent effective regulation and reduced this to 11 key indicators. We would encourage the OfS to be similarly focused on the greatest areas of risk and public interest. We would propose that the OfS sets a limited number of key indicators – maybe 10 - that it will assess on a rolling three-yearly basis and publish the areas for prioritisation at least a year ahead of implementation. These indicators should be based on the highest areas of risk or public interest which could include the number of students below the numerical threshold or number of students furthest from the numerical threshold.

The mass of indicators would be a particular challenge for smaller providers with fewer specialist data teams working to analyse the data. Providing huge data workbooks may give the impression of transparency and openness, but will have the consequence of diverting considerable staff resource away from established processes of quality management such as annual monitoring and periodic review. It is questionable how useful these workbooks are, or what OfS expects institutions to do with them. This problem is likely to be exacerbated in smaller organisations without large dedicated specialist data teams.

It is important to restate that continuation and completion are not the only measures of quality, and at best are poor proxies for it. In particular the use of progression data for whether graduates go on to professional and managerial jobs is a real concern. We recognise that students value successful outcomes and that getting a good job is seen as part of that. However, the role of the regulator is to regulate on things over which an institution has control. We therefore continue to have grave concerns about trying to hold providers to account for whether graduates go on to achieve “professional and managerial level jobs” which does not measure of quality or success of the course. There are structural inequalities within society that impact on the type and level of jobs that people are appointed to. There are also likely to be regional differences in current skills needs. We believe that the provision of this information for students is an important step to ensure greater transparency as part of their decision-making process on which course and institution to apply to, but that it shouldn't be used as a regulatory

enforcement tool due to the lack of control that institutions have over the metric and the potential unintended consequences that may result. We believe that this metric is a poor proxy for the quality of the course and should be removed. As highlighted in para 93 of the OfS' own analysis of the earlier consultation document "Only a small number of respondents suggested that progression to higher-level study or professional and managerial employment was an appropriate measure of student success or quality. A larger number commented that it was not." We would argue that the Regulator's Code (5.3) places a responsibility on regulators to not just consult those that they regulate but also that is "meets their needs". We would argue that ignoring consultation responses does little to "improve confidence in compliance for those they regulate" (1.2).

Additionally, there is research which shows that students achieving a 'good degree' are more likely to get a professional or managerial job or go into further study. We are confident that most English providers take their responsibilities for academic standards very seriously, and that 'good degrees' are awarded appropriately. However, the emphasis on progression to managerial and professional employment within 18 months might provide a perverse incentive to encourage higher grades, especially as this also forms part of the regulation on access and participation.

Finally, we believe that the proposals to extend the ongoing conditions to all higher education courses and provision whether they receive OfS/public funding or not raises a number of questions and could prompt a number of unintended consequences. We remain concerned that this would result in a massive expansion in the OfS and regulatory burden – particularly in light of the proposed mass of other data that the OfS will be expecting to review as outlined above. We believe that the additional regulatory cost and burden on the sector would be disproportionate to the risk presented.

Questions relating to Proposal 3: Setting numerical thresholds for student outcome indicators

Question 5: Do you agree or disagree with our proposed approach to setting numerical thresholds set out in Annex E? If you disagree, please provide reasons and any alternative suggestions.

Agree

Disagree

Don't Know

In addition to the principled objections to the use of the progression metric, which does not measure quality and without benchmarking it would penalize institutions for structural societal issues beyond their control, we also have a number of practical concerns.

The Graduate Outcomes dataset was described as an experimental dataset in June 2020 and the survey was not designed for a regulatory purpose. We believe that there are still significant flaws in the dataset. The OfS' as a provider of Official Statistics has a responsibility to meet the [Code of Practice for Statistics](#) which emphasises the need for "Suitable data sources". We believe that there are significant questions as to whether Graduate Outcome meets this threshold yet based on the low response rate and the imprecise way in which data is gathered and coded. We would also question whether the census-based nature of the survey – gathering what a student is doing on a particular

date in time recognises the portfolio nature of many graduates careers and therefore the survey is not “...the most appropriate data to meet intended uses.” The current Graduate Outcomes was also gathered in the midst of a pandemic, which had particular impacts on some industries, with the creative industries to a large extent being shut for much of the period and so reducing career options for these graduates.

We are very concerned about the Progression measure, for the following specific reasons:

1. It is based upon Graduate Outcomes (GO) data which is flawed in its interpretation and currently has a response rate below that deemed publishable for NSS data.
2. 15 months is too short to decide if the course helped gain SOC 1-3 employment, not all careers work like that, not all people work in a linear fashion, and jobs is not the only positive outcome for graduates.
3. The definition of a graduate level outcome depends upon the interpretation by HESA of SOC for each GO respondent. It relies on how a graduate describes their own job, and then how this is coded. There’s a whole lot of interpretation going on and it relies on a non-expert describing their role to someone who doesn’t understand it, who then categorises it.
4. SOC is also out of date in its understanding of graduate level roles within the creative industries and many other industries and professions.
5. The GO misses some students who have undertaken a one year postgrad course – with one provider highlighting that they are in Cohort D so the census date is after a course has finished but easily before someone might have secured employment.

There are particular challenges with the definition of professional and managerial jobs as defined by SOC Codes. There are a number of graduate-level jobs – such as in early-years education, working in care homes, costume design, fashion merchandising and hair and make-up for TV and Film as well as the challenges of those in portfolio careers such as in the creative industries – that are not currently recognised at this level that should be. There will be many more careers in a similar position and this is also likely to continue to change in the future.

Linking the Progression Indicator to the SOC Codes creates an unnecessarily inflexible system. The SOC Codes are currently only reviewed every 10 years and also feed into other government priorities – such as for visa application purposes - and so are likely to face different political pressures. Some careers will become graduate-level and new and innovative career paths will be created within the 10-year period before the next review. We would therefore propose that if the OfS does insist on using the progression metric that you only use the SOC Codes as a starting point, but have the possibility to create an additional list of graduate-level jobs for careers where the SOC Codes haven’t yet caught up with yet.

We believe that 15 months is too short to decide if the course helped gain SOC 1-3 employment, not all careers work like that, not all people work in a linear fashion, and jobs is not the only positive outcome for graduates. There are many reasons why this is too short a timeframe to make robust judgements about how the university experience has prepared students for work, and taking a measure so early after completion does not offer a true reflection of the impact of the university experience on the functional value of the qualification received. Indeed for some industries gaining the qualification is just a pre-requisite for starting

a career and additional years of experience are needed to reach SOC codes that consider that profession 'highly skilled'.

Furthermore, this does not reflect the transient nature of work in modern Britain, with many young people undertaking multiple jobs, continually developing their talents and portfolios and/or working in industries such as creative or agriculture where self-employment is the norm. We are therefore worried that using progression as a proxy for quality is misleading to the public not only because universities have very little influence over students graduate career paths, but the data is collected way too early in the graduates working life to make a meaningful contribution to the assessment of the degree's impact. There are also of course other intrinsic outcomes for students gaining a HE qualification (such as happiness, job stability, good health, and confidence - especially important for students of colour, mature students and first-generation students) and we think the OfS can and should do more to capture and use this data in its assessment of quality. Securing a specific job 15 months after graduation is not the only positive outcome for graduates.

Additionally it is important to note that judgement based upon GO is unsafe as:

- The second GO survey took place at the height of the pandemic when employment in the creative industries was artificially suppressed (it is a sector dominated by self-employment which is now recovering gradually)
- There appears to be no contextualisation for sub-regional labour markets and as yet we do not understand how benchmarks (above the 60% threshold) have been determined.

Case study:

One institution provided some examples of the concerns about the over-reliance on how a graduate describes their own job, and then how this is coded. The system relies on a whole lot of interpretation and it relies on a non-expert describing their role to someone who doesn't understand it, who then categorises it. One institution provided specific examples of GO respondents very obviously doing a graduate level job and yet they have been coded as unskilled. For example:

- a filmmaker, active in his profession, being classified as sports and leisure assistants SOC code 62110
- a graphic designer leading a complex graphic design and typography project being coded as a pre-press technician SOC code 54210
- graduates working on film sets within their profession were also coded as leisure assistants
- all of the costume graduates we could identify have been coded outside the top three SOC codes

The scale of the impact of these various issues with the HESA data is significant. In year 1 progression data there are errors relating to around 11% of GO respondents so we calculate that we have been negatively impacted by between 14 and 19% points, meaning our indicator value should be 60 to 65%. In year 2 progression data there are errors relating to around 12% of GO respondents so we believe that we have been negatively impacted by c.12% points.

In summary we would therefore strongly encourage the OfS to reconsider its proposal to use the progression metric, ideally at all since it doesn't measure quality and at the very least until the processes for data collection are more robust and reporting against SOC codes more accurately reflects the graduate labour market.

Finally, whilst we recognise the movement on benchmark data to consider the impact that student characteristics may have had on historical data we would still have concerns of the unintended consequences of this proposal to move away entirely from benchmarking of threshold data. We would also suggest that the OfS should undertake a full equality impact assessment on the impact of not benchmarking data by particular groups of students, and in particular the impact on mature learners, who face very particular challenges.

Question 6: Do you agree or disagree with the proposed numerical thresholds set out in summary in Table 1 and shown in full in 'Setting numerical thresholds for condition B3'?

Agree

Disagree

Don't Know

Our position is that the numerical thresholds in Table 1 is simply too complex a set of indicators to use to effectively regulate quality and standards in the sector. While we do not disagree with any particular indicators, we believe that a much smaller subset of those shown in Table 1 would be the right basis on which to regulate quality.

We feel unable to comment on each of the proposed metrics put forward in Table 1 without a detailed understanding of how they have been compiled and whether they appropriately reflect benchmarked performance across the sector.

Questions relating to Proposal 4: Publishing information about the performance of providers in relation to the OfS's numerical thresholds

Question 7: Do you agree or disagree with our proposal to publish information about individual providers' student outcomes and performance in relation to our numerical thresholds, as well as sector-wide data, on our website?

Agree

Disagree

Don't Know

We would also argue that whilst publishing the data could be beneficial in terms of public accountability, we believe that the thousands of indicators – when considering the different levels, split metrics, student characteristics, subject mix and so on – would result in just publishing too large a dataset that would result in obfuscation rather than transparency. The data would then rely on third parties – such as league table providers with their own commercial interests – interpreting and presenting the data.

If the data is published, but there are valid contextual reasons for being below a numerical threshold it will be important that this context is presented in a clear and transparent way so as not to create

unnecessary confusion of why a provider is below a threshold in a priority area but no action is being taken.

Questions relating to Proposal 5: Making judgments about compliance with condition B3, including consideration of context

Question 8: Do you agree or disagree with the proposed approach to assessment set out in Annex F? Is there anything we could do to improve the clarity of this information for providers?

Agree

Disagree

Don't Know

We would strongly emphasise the importance of not forming negative judgements where the data is either too small or not statistically robust enough to form a judgement. In para 38a of the consultation document it states that, "Where it is not clear whether a particular outcome should be viewed as positive (either because interpretation of the outcome is debatable, or because existing data does not provide sufficient granularity of information), we propose to interpret it as either positive or neutral for the purposes of constructing student outcome measures, rather than treating it negatively." For many smaller providers much of their data is not likely to be statistically significant or will be volatile on a year-on-year basis and this should not be seen as a negative unless there is a rigorous discussion with the provider based on other evidence. **It should also be noted that this assumption of more detailed conversations with providers with smaller datasets bakes-in additional regulatory burden for smaller providers and the OfS should reflect on whether this is in-line with the Regulators Code and how this will be implemented in practice.**

The proposal to apply the initial condition to any provider that is able to produce at least one indicator for at least one of the previous five years could end up relying on very historic data. For example, the progression data from five years ago would be based on students that graduated almost 7 years ago. It might be better to have data for one indicator in the previous three years or two indicators within the last five years. It would be even more important to consider the context and comparability of this historic data, especially if five years old, with the current profile of students and courses at the provider.

The consultation proposes that the OfS itself will assess the provider's compliance and will not commission the designated quality body to undertake assessment activity. The DQB is able to draw on academic expertise when making its judgements and in particular when considering the contextual evidence for continuation and completion rates that might require academic judgement as well as just regulatory judgement. In complex cases the OfS should draw on the expertise of the DQB or should demonstrate how they will make these assessments in the context of wider academic judgement.

A concern with examining below threshold performance across indicators and split indicators is the very high likelihood of sampling errors – i.e., differences from threshold being judged significant (or a matter of quality) when they almost certainly result from statistical sampling instead.

We have two concerns with the approach outlined in paragraphs 237-240.

First, applying confidence limits e.g., 90%, to each indicator and split indicator will lead to a large family-wise error rate – the problem of multiple comparisons. To illustrate, comparing one indicator with threshold may give 90% confidence that the difference from threshold is genuine (and not explained by sampling error). However, if two comparisons are made, the probability that both are unaffected by sampling error falls to $90\%^2$ or 81%. We have not fully calculated the number of comparisons the OfS's proposals involve, but it is very large indeed. What this means is that the confidence that the overall pattern of multiple comparisons is unaffected by sampling error will be vanishingly small, and each provider's set of comparisons is almost certain to contain one or more differences that are claimed, in error, to be statistically significant.

We believe the OfS is aware of this problem – the problem of multiple comparisons – but we are concerned that it does not have serious or statistically informed proposals for how to deal with it. Paragraph 29 of the document “Supporting information about constructing student outcome and experience indicators for use in OfS regulation: Description of statistical methods” explains why statistical adjustments should generally be made in the case of multiple comparisons. But paragraph 30 then goes on to say that “We have proposed not to make any such adjustments for multiple comparisons within our construction of student outcome and experience indicators to inform our regulation of student outcomes and the TEF”.

We simply do not understand why the OfS is not proposing a statistically informed treatment of such an important and well-understood problem as this. The OfS will be aware that the past President of the Royal Statistical Society said of this problem, that "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". Their letter went on to explain that failure to deal with this problem in the appropriate way had rendered invalid all TEF awards previously given. We believe that failure to deal with this problem in relation to the regulation of quality would be an equally serious mistake and urge the OfS to adopt an appropriate scientific method for dealing with multiple comparisons.

Paragraph 31 goes on to explain how OfS would like users of these complex datasets to behave, given it is not proposing to make adjustments for multiple comparisons:

“While we have proposed not to adjust for multiple comparisons, we do ask users who wish to make multiple comparisons to consider adjusting to a higher level of confidence when making their judgements because of the higher risk of false discovery when using lower levels of statistical confidence. Users should be more conservative in their interpretation of statistical uncertainty the more comparisons they are making.”

To be absolutely clear, it is not there may be some users who wish to make multiple comparisons – this sentence suggests that the OfS has not fully grasped the nature of the statistical problem. It is that the proposals from the regulator intrinsically commit to multiple comparisons being made for each provider.

Nor does it seem appropriate for users of the statistics to consider adjusting to a higher level of confidence – inexperienced statisticians will almost certainly fail to understand how they could appropriately adjust to a higher level of confidence; and experienced statisticians are almost certain to understand that an explicit statistical adjustment must be made.

We believe it is unreasonable for the OfS to introduce a problem of statistical interpretation but to eschew established statistical methods to control for it (e.g. Bonferroni adjustments). We believe this is a significant weakness in the OfS proposals and urge that it is rectified. Either explicit statistical adjustments for multiple comparisons should be introduced, or the process for assessing compliance should involve only a very few statistical comparisons for each provider.

The second challenge with the proposed approach is that it undermines any claim made by the OfS that it has evidence of non-compliance. All the OfS data will show is a certain probability that the below threshold performance signifies an underlying performance differential. But it is equally open to the argument, from the provider, that there is no underlying performance differential at all, and that the statistical significance must derive from sampling error. Note, that in those cases where the statistical significance does indeed derive from sampling error, neither the OfS nor the provider could possibly demonstrate this. Yet the OfS's proposed procedures place the burden of proof on the provider to show that they remain compliant. But, in this case, there can be no empirical evidence they can cite, since the statistical significance arises from sampling and does not point to an underlying performance differential. In cases such as this, the provider would end up in a search for a spurious contextual explanation to rebut a spurious argument for non-compliance. We believe that the burden of proof cannot be placed on the provider and must therefore remain on the OfS. In this regard, we find paragraphs 237-240 very worrying.

Paragraph 237: we believe the OfS should only use a very high confidence level of at least 95%. Values as low as 80% or even 90% mean that errors will typically be made in 10%-20% of cases with perhaps no possibility of resolution either way.

Paragraphs 239-240: we believe that the OfS should not have the freedom to choose to adopt lower statistical confidence values. The point of statistical confidence is to give confidence that the difference from threshold is likely the result of a genuine performance differential. In giving itself the freedom to choose lower confidence values, the OfS would have the freedom to pursue differences from threshold that could not confidently be said to reflect any performance differential at all. We simply do not understand how such freedom could be consistent with good regulation.

Question 9: Do you agree or disagree with our proposed general approach to prioritisation? If you disagree, do you have any alternative suggestions for how we should approach prioritisation?

Agree

Disagree

Don't Know

We believe that the question of how the OfS prioritises where it focuses its resources is one of the key areas of this consultation. As we highlighted above the vast swathes of indicators, split indicators and other data points that will result from the proposals are likely to result in most providers having some data below a threshold somewhere. There therefore needs to be a robust, transparent process for deciding how the OfS will prioritise where it will focus its activities for the year.

We believe that this should not just be decided by officials within the OfS but that there should be an independent process for deciding on priority areas that includes students – as the main parties affected – as well as academic experts and peers from within institutions. There is also a strong argument that this priority focus should not change every year but rather it would be better to focus

on particular areas for a couple of years to assess the impact of action and ensure change is embedded.

In terms of the factors on which prioritisation could be based we would argue that distance from threshold, student population, statistical certainty and mode or level study could all be reasonable justifications for prioritisation. However, we do not agree that thematic approach or random selection are appropriate and are more akin to the regulator going on a fishing expedition.

Question 10: Do you think that the OfS should adopt Option A or Option B (see paragraphs 207) when defining the scope of each assessment for ongoing condition B3?

If the OfS go with Option B we would argue that a full review of all indicators below the numerical threshold process is only triggered once it has been established that the indicator that has been prioritised is actually below threshold once contextual factors and statistical certainty have been considered.

Option A would be more appropriate if the OfS decides to use thematic enquiries for its prioritization. If you decide to prioritise for example part-time students on other undergraduate courses – and you select providers on the basis of being below thresholds for this data then it shouldn't be used to then explore other indicators.

Question 11: Do you agree or disagree with our proposals for considering the context of an individual provider when assessing compliance with condition B3?

Agree

Disagree

Don't Know

We do not agree with an over-reliance on purely quantitative measures of quality and therefore welcome the recognition of the role of taking into consideration contextual factors. We do believe that there needs to be more detail provided on the kind of evidence that would be considered as part of this process, for example we believe that the use of contextual information around IMD and ABC characteristics should form part of the benchmark calculation. It will be important that this contextual information is considered alongside the data, rather than being seen as an after-thought and not given due consideration when forming the judgement.

There should be transparency in the process of how this will occur, for example it should not just be based on the decision of OfS officials but should be considered through processes – such as the OfS QAC – that include academic experts and peers. We would also want some clarity on how context will be presented alongside the publishing of data workbooks. We would argue that it would need to be presented in some form so as not to result in confusion for people looking at the data.

Case study: Agriculture and land-based subjects

There are some subjects where a significant proportion of the students or indicators are below the numerical threshold. For example, looking at the Progression metric for Agriculture, Food and

Related Subjects where 51.4% of students are currently at or below the numerical threshold of progression to professional and managerial jobs. The agricultural and land-based industries have well-developed qualification routes from both further education and higher education but the different skill levels for different jobs are often less well reflected in the job titles and so these graduates are disproportionately impacted by mis-coding in Graduate Outcomes returns.

It is important that the OfS consider how this data is presented publicly so as not to impact on a strategically important sector of the economy where future food security is likely to be an increasing government priority. It would be worth considering whether there should be a separate numerical threshold for subjects or indicators that are statistically significant distances from the sector-wide numerical threshold or whether it should be more explicitly mentioned in the contextual information. It will be necessary to ensure that reforms are directed at the key areas of major risk to the sector, rather than unintentionally catching up specific subjects by mistake.

Questions relating to Proposal 6: How the OfS will address statistical uncertainty in the assessment of condition B3

Question 12: Do you agree or disagree with the proposed approach to using statistical measures when considering a provider's performance in relation to numerical thresholds?

Agree

Disagree

Don't Know

We welcome the recognition of the need to consider the statistical significance of the data for providers, and particularly the impact that this will have on smaller providers with smaller – and therefore more volatile - datasets. We would strongly restate the importance of the principle of not forming a negative judgement based on small datasets that have high degrees of statistical uncertainty and it will be important to consider how this data is presented as part of the transparent approach to publishing dataset and appropriate caveats highlighted. We would particularly wish to see a mechanism that enables the public to understand the wider context of the provider and for the OfS to make sure that external audiences are not manipulating the data in ways that go against the principles of sound data analysis.

However, we also believe the OfS's proposed approach to statistical certainty is unclear and not supported by scholarship or research around the use of statistics. In particular, we are unaware of any scientific analysis that supports the four confidence values outlined in paragraph 237. Moreover, use of lower confidence values will lead to a higher error rate. Paragraphs 238-240 are unclear in many points. For example, paragraph 238 proposes that the OfS would use 90% and 95% confidence values but only "in general". We do not understand why the OfS would wish to hedge its use of confidence values, but in doing so it is introducing a lack of clarity as to the actual confidence values it would use. Similarly, paragraph 239 states that "We propose applying regulatory judgment in each case, rather than using the categories as a determinative way of deciding whether and how we act." It is unclear exactly what the OfS proposes, but we would be very concerned if the OfS is proposing drawing on statistical certainty to assess whether performance falls below threshold, and then proposes that it does not have to follow those same statistical precepts. Paragraph 240 indicates that a 90% confidence values would "normally" be used, but that the OfS may choose to use lower confidence values, e.g., 88%. We do not believe the OfS is using the statistical notion of a

confidence value correctly, and believe it would not be appropriate for the OfS to set out the confidence values it will use, and then give itself the freedom to use different ones in different circumstances. We simply do not understand the statistical argument for using lower confidence values for protected characteristics, where multiple indicators are below threshold, nor where performance is further from threshold. In the latter case, this appears to be an entirely spurious and invalid consideration, as distance from threshold and confidence value are not independent constructs.

A concern with examining below threshold performance across indicators and split indicators is the very high likelihood of sampling errors – i.e., differences from threshold being judged significant (or a matter of quality) when they almost certainly result from statistical sampling instead.

We have two concerns with the approach outlined in paragraphs 237-240.

First, applying confidence limits e.g., 90%, to each indicator and split indicator will lead to a large family-wise error rate – the problem of multiple comparisons. To illustrate, comparing one indicator with threshold may give 90% confidence that the difference from threshold is genuine (and not explained by sampling error). However, if two comparisons are made, the probability that both are unaffected by sampling error falls to $90\%^2$ or 81%. We have not fully calculated the number of comparisons the OfS's proposals involve, but it is very large indeed. What this means is that the confidence that the overall pattern of multiple comparisons is unaffected by sampling error will be vanishingly small, and each provider's set of comparisons is almost certain to contain one or more differences that are claimed, in error, to be statistically significant.

We believe the OfS is aware of this problem – the problem of multiple comparisons – but we are concerned that it does not have serious or statistically informed proposals for how to deal with it. Paragraph 29 of the document “Supporting information about constructing student outcome and experience indicators for use in OfS regulation: Description of statistical methods” explains why statistical adjustments should generally be made in the case of multiple comparisons. But paragraph 30 then goes on to say that “We have proposed not to make any such adjustments for multiple comparisons within our construction of student outcome and experience indicators to inform our regulation of student outcomes and the TEF”.

We simply do not understand why the OfS is not proposing a statistically informed treatment of such an important and well-understood problem as this. The OfS will be aware that the past President of the Royal Statistical Society said of this problem, that “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”. Their letter went on to explain that failure to deal with this problem in the appropriate way had rendered invalid all TEF awards previously given. We believe that failure to deal with this problem in relation to the regulation of quality would be an equally serious mistake and urge the OfS to adopt an appropriate scientific method for dealing with multiple comparisons.

Paragraph 31 goes on to explain how OfS would like users of these complex datasets to behave, given it is not proposing to make adjustments for multiple comparisons:

“While we have proposed not to adjust for multiple comparisons, we do ask users who wish to make multiple comparisons to consider adjusting to a higher level of confidence when making their judgements because of the higher risk of false discovery when using lower levels of statistical confidence. Users should be more conservative in their interpretation of statistical uncertainty the more comparisons they are making.”

To be absolutely clear, it is not there may be some users who wish to make multiple comparisons – this sentence suggests that the OfS has not fully grasped the nature of the statistical problem. It is

that the proposals from the regulator intrinsically commit to multiple comparisons being made for each provider.

Nor does it seem appropriate for users of the statistics to consider adjusting to a higher level of confidence – inexperienced statisticians will almost certainly fail to understand how they could appropriately adjust to a higher level of confidence; and experienced statisticians are almost certain to understand that an explicit statistical adjustment must be made.

We believe it is unreasonable for the OfS to introduce a problem of statistical interpretation but to eschew established statistical methods to control for it (e.g. Bonferroni adjustments). We believe this is a significant weakness in the OfS proposals and urge that it is rectified. Either explicit statistical adjustments for multiple comparisons should be introduced, or the process for assessing compliance should involve only a very few statistical comparisons for each provider.

The second challenge with the proposed approach is that it undermines any claim made by the OfS that it has evidence of non-compliance. All the OfS data will show is a certain probability that the below threshold performance signifies an underlying performance differential. But it is equally open to the argument, from the provider, that there is no underlying performance differential at all, and that the statistical significance must derive from sampling error. Note, that in those cases where the statistical significance does indeed derive from sampling error, neither the OfS nor the provider could possibly demonstrate this. Yet the OfS's proposed procedures place the burden of proof on the provider to show that they remain compliant. But, in this case, there can be no empirical evidence they can cite, since the statistical significance arises from sampling and does not point to an underlying performance differential. In cases such as this, the provider would end up in a search for a spurious contextual explanation to rebut a spurious argument for non-compliance. We believe that the burden of proof cannot be placed on the provider and must therefore remain on the OfS. In this regard, we find paragraphs 237-240 very worrying.

Paragraph 237: we believe the OfS should only use a very high confidence level of at least 95%. Values as low as 80% or even 90% mean that errors will typically be made in 10%-20% of cases with perhaps no possibility of resolution either way.

Paragraphs 239-240: we believe that the OfS should not have the freedom to choose to adopt lower statistical confidence values. The point of statistical confidence is to give confidence that the difference from threshold is likely the result of a genuine performance differential. In giving itself the freedom to choose lower confidence values, the OfS would have the freedom to pursue differences from threshold that could not confidently be said to reflect any performance differential at all. We simply do not understand how such freedom could be consistent with good regulation.

Question 13: Do you have any suggestions for additional steps the OfS could take to provide greater clarity about the impact that the proposed approach to statistical confidence may have for individual providers?

We believe the approach would be much improved by focusing on a much smaller set of indicators, with fewer multiple comparisons. To do this, we believe the OfS should set and maintain consistent use of a much higher confidence value (e.g., 95%) and use statistical adjustments for multiple comparisons (e.g., Bonferroni) to maintain the family-wise confidence value at the same high level. This would give much greater confidence to the sector and to users of regulatory information that where the OfS had identified possible non-compliance, it was highly likely to be indicative of a genuine compliance problem. The current proposal will, almost as a matter of certainty, generate far too many spurious cases of statistical significance.

We recommend that whatever final proposals the OfS adopts it commissions the Royal Statistical Society or similarly respected and recognised body to endorse the use of statistics proposed.

Questions relating to Proposal 7: Taking regulatory intervention when a breach is identified

Question 14: Do you agree or disagree with our proposals to impose an ‘improvement notice’ where we find a breach of condition B3?

Agree

Disagree

Don't Know

The term improvement notice is more understandable to a wider audience than “specific condition of registration” and so would aid transparency on the action taken by the OfS. There will need to be clear guidance on how a provider can remove their improvement notice and what steps they need to take – something that was missing in the earlier enhanced monitoring processes that the OfS used.

The improvement notice should only be issued in the areas of prioritization in any given year – rather than for every provider below the threshold for an indicator - and this should only be given after discussion with the provider about their context and statistical reliability of the data.

Question 15: Do you agree or disagree with our proposals to take account of a provider’s compliance history in relation to condition B3 for the purpose of determining eligibility for other benefits of OfS registration?

Agree

Disagree

Don't Know

We believe it is appropriate to consider a provider’s compliance history. It is reasonable to draw conclusions about the credibility of a provider’s intentions based upon the manner in which it has conducted itself on previous occasions. It follows that institutions which have made efforts to conduct themselves in an open, honest and self-critical manner should be afforded a greater amount of trust in their efforts to address areas of concern.

Questions relating to Proposal 8: Timing of implementation

Question 16: Do you agree or disagree with the proposals for the implementation of the proposed approach to regulating student outcomes? If you disagree, do you have suggestions for an alternative timeline?

Agree

Disagree

Don't Know

The timing looks unnecessarily rushed, especially since this timescale is a one-off cycle based on establishing the process whereas in future years the cycle will be aligned to HESA data. Most providers will only have received the data, in this format, for the first time in January 2022 and for those smaller providers with smaller data teams working through it all is a huge task.

We would also argue that publishing the data in September and then making decisions in October gives little time for providers to check the data given the vast swathes of data being published.

We believe that there should be at least three months between the publishing of the data and the OfS announcing which institutions will be subject to scrutiny. This will allow time for institutions to check the validity of the data and also analyse the data and consider initial action plans for how they will tackle any areas of weakness – something that the OfS has said it will consider as part of its contextual considerations. We would also suggest moving to the same deadlines and timescales that will be used in future years to get institutions in sync with the correct data publications and allow more time to prepare.

We do not think that starting the implementation in June 2022 is realistic, considering the tasks that we think would be required to do this:

- OfS to fully consider all the responses to this consultation and other relevant consultations,
- OfS to adjust their proposals as required and publish these with feedback.
- OfS to then train internally and launch training for providers.
- Providers to be able to prepare and resource (which may include making additional appointments) to ensure they can engage and respond,
- OfS to advance the time series a further year and publish this
- Providers to engage with this first publication of a time series of confirmed methodology, which is likely to require more capacity than in future years as the OfS training is applied for the first year.

We believe the combined burden of possibly preparing a TEF submission at approximately the same time as being assessed against B3 compliance is too great. Institutions will want to invest considerable time in both exercises, ensuring they fully understand their performance indicators, and any splits, as well as context. We believe the regulatory burden on providers would simply be too great and suggest that at least one of these exercises (probably the TEF) is moved back one year.

We also believe that the OfS should only identify additional providers to consider “in-year” on the basis of new data, rather than just having completed the proposed reviews and wanting to move on to the next tier.

Question relating to Considering regulatory burden on registered providers

Question 17: Is there anything else we could consider that would reduce regulatory burden for providers while regulating minimum requirements for student outcomes?

We have real concerns that publishing three major – and highly complex - consultations at the same time, with only an eight week consultation period, has placed significant pressure on the higher education sector as a whole and specifically disadvantaged smaller providers. We have received numerous comments from members about feeling totally overwhelmed by the scale and complexity of the consultations, particularly given other consultations and reforms that are

also happening. We further expect that institutions have had to prioritise the consultation responses and so the OfS will not receive either as many responses, or that those responses will have been subject to the level of consideration, appropriate for the scale of the changes proposed. We would strongly suggest that the OfS continue to discuss the proposals, implications and next steps with institutions and their representative bodies as it develops its proposals.

Whilst we welcome that these consultations were not published over the Christmas period we would reiterate our concerns that the three consultations were published at the same time, with only an eight week consultation period. This places undue pressure on smaller providers and will have an impact on the number of responses that you will receive. In smaller providers it is likely to be the same person responding to all three consultations and that they are unlikely to have a large data team able to analyse all the implications of the consultations. To ensure that appropriate consideration is able to be given to consultations we would suggest in future that you either only publish one consultation at a time or give a longer period for responding, say a minimum of 12 weeks.

Finally, we would reiterate the point that we have throughout our response that for a regulator to be effective it is important to focus on the biggest areas of risk and therefore to focus on a smaller number of key indicators. We believe that the regulatory burden resulting from B3 is disproportionate to the scale of the problem. We recognise that the consultation is not seeking to gather additional data, and has sought to mitigate some of the impact on smaller providers by giving providers processed work-books of data. However, we believe that the thousands of indicators – when considering the different levels and split metrics – would result in a significant data analysis burden for both the regulator and for providers and will specifically disadvantage smaller providers in the sector who do not have teams of data practitioners to scrutinise the data workbooks nor have access to additional data/research to provide alternative analysis when responding to the effects of the data approach taken by the OfS. Due to the number of students in many of the split metrics it is also likely to result in data that is not statistically robust in many smaller providers. We would encourage the OfS to focus on the small number of key indicators that represent the greatest areas of risk and public interest.

GuildHE, Woburn House, 20 Tavistock Square, London, WC1H 9HB

Tel: 020 3393 6132, email: info@guildhe.ac.uk

Charity Number: 1012218