Cyngor Cyllido Addysg Uwch Cymru
Higher Education Funding Council for Wales

Cwrt Linden
Clos Ilex, Llanisien
Caerdydd, CF14 5DZ
Ffôn 029 2076 1861

Linden Court
Ilex Close, Llanishen
Cardiff, CF14 5DZ
Tel 029 2076 1861

hefcw

www.hefcw.ac.uk

Cylchlythyr | Circular

Data publication thresholds and aggregation on Unistats: Consultation on thresholds and subject groupings for data on Unistats and the National Student Survey

Date:	16 December 2014
Reference:	W14/46HE
To:	Heads of higher education institutions in Wales Principals of directly-funded further education institutions in Wales Other interested parties
Response by:	13 February 2015, online
Contact:	Name: Dr Cliona O'Neill Telephone: 029 2068 2283 Email: cliona.oneill@hefcw.ac.uk

This circular provides a consultation on a proposal to lower the thresholds for the publication of data on the Unistats web-site and on the outcomes of the National Student Survey. It also seeks views on a revised subject hierarchy for aggregating data where publication thresholds are not met.

If you require this document in an alternative accessible format, please telephone us on (029) 2068 2225 or email info@hefcw.ac.uk.

Noddir gan Lywodraeth Cymru
Sponsored by Welsh Government

Introduction

1. This consultation invites views on a proposal to lower the thresholds for the publication of data on the Unistats web-site and on the outcomes of the National Student Survey.

2. It also seeks views on a revised subject hierarchy for aggregating data where publication thresholds are not met.

Executive summary

Key points

3. The UK higher education funding bodies are committed to ensuring that the Unistats web-site remains valuable to prospective students and those who advise them. In response to feedback from the sector, and in the interest of prospective students and other users of the site, we have been exploring options for changing publication thresholds on Unistats. Currently, National Student Survey results are published using the same approach as Unistats, and we propose that this continues.

4. The current publication thresholds applied to student-related data and their aggregation approach (described in detail in paragraphs 20 to 22) have been intended to ensure the availability of robust, reliable and comparable data. However, following discussions with sector representatives about the limitations these place on the quantity and relevance of publishable data, the funders agreed to explore lowering thresholds and to review the subject hierarchy used for aggregation.

5. After investigating the likely impact on the quantity of publishable data and their robustness, we are proposing to lower the headcount publication threshold to 10, while retaining the response rate threshold at 50 per cent. This consultation seeks views on that proposal.

6. We have also taken this opportunity to consider how we group subjects when aggregating data on Unistats. We have analysed the numbers of students and courses linked to the codes in the Joint Academic Coding System and are proposing new subject groupings that take account of the distribution of provision across subjects. We recognise that our data-driven approach should be modified by feedback from the sector, to ensure that any changes accurately reflect how provision is generally structured. We would therefore welcome comments on the proposed subject hierarchy and on the principles used to develop it. We are also seeking the sector's view on the most appropriate time to implement such changes.

Action required

7. Responses to this consultation should be made online by **noon on Friday 13 February 2015,** using the response form which can be accessed alongside this document on the consultation web page.

8. This is an open consultation. We welcome comments on either or both parts from higher education providers and from other groups, sector representatives, organisations or individuals with an interest in the Key Information Set and Unistats. We recognise that these proposals may be of interest or relevance to different areas within institutions, but we would encourage the submission of a single institutional response where possible. We would therefore be grateful if institutions could collate responses internally where necessary, prior to submission.

Background

9. Unistats (www.unistats.ac.uk**)** is the official site for searching for and comparing data on undergraduate courses from UK higher education institutions (HEIs), and from further education institutions (FEIs) in England and Wales[1].

10. The development of the Key Information Set (KIS), and its presentation on the Unistats web-site, has been informed by a substantial programme of research and evidence collection. It contains the items of information that prospective students have identified as most important in making their decisions, and is intended to provide high-quality information in an accessible, comprehensive and comparable way.

11. Our approach to presenting data on Unistats aims to balance the need to supply data that are statistically robust and meet data protection standards with that of providing information that is relevant and easy to understand, allows straightforward comparison, and will ultimately help prospective students to make informed choices about higher education courses. Our approach includes setting thresholds for publishing data based on a headcount of numbers of students and, in some cases, additionally by response rate. Where these thresholds are not met, we aggregate data over years and across subject areas in an attempt to meet them, and publish these aggregated data where possible. If thresholds are still not met, no data are published.

12. Since the launch of the site, feedback has been provided by some institutions about our approach to setting thresholds and aggregation. It has been argued that high-level aggregation makes the data less comparable and of less interest to those seeking course-level information. It has been reported that it can also bring together data on subjects that may in practice be quite different, and could therefore be misleading. In addition to this, user research has shown that a lack of data can be viewed negatively by potential students and their advisers.

[1] More detailed information on Unistats and KIS can be found at www.hefce.ac.uk/whatwedo/lt/publicinfo/kis/. Further information on research that informed the development of the KIS can be found at www.hefce.ac.uk/whatwedo/lt/publicinfo/kis/kisrd/

13. It is recognised that these issues may disproportionately impact on smaller or specialist institutions, or further education institutions which have smaller cohorts.

14. We held two roundtable events in 2013 to consult sector representatives from all four UK nations on the issues arising from the approach used to provide data on Unistats. The outcome of these roundtable events was that the funders agreed to explore lowering publication thresholds. We originally intended to do this as part of a more fundamental review of Unistats and the KIS, which the UK funding bodies are undertaking as one element of a wider review of the provision of information for students. However, as any changes resulting from the outcomes of the wider review will be implemented from 2017, we are proceeding with this separately, with a view to making changes for the publication of KIS 2015 on Unistats if feasible.

15. We recognise that issues remain around including and presenting data for the full range of provision, which need to be considered as part of the wider review[2]. This consultation will, however, allow us to test whether a short-term change of this nature would be welcome, and whether it is acceptable to set thresholds for publication at a lower level, and we can carry this forward as a principle for any future publication of these data.

Development of proposals

16. The first step of this piece of work was to model the 2014 KIS data using a range of lower thresholds, to assess their impact on the quantity of publishable data and the level of specificity at which they could be published.

17. To ensure any proposal to lower thresholds is sound and statistically defensible, we sought independent statistical advice on the results obtained from modelling work, and asked about the implications of lowering thresholds, such as considerations of robustness of data and statistical uncertainty.

18. We then took the results of the modelling and the statistical advice to a focus group of sector representatives from across the UK, for advice on how to proceed.

Consultation proposals

19. This consultation consists of two parts.
Part 1: A proposal to lower publication thresholds on Unistats.

[2] The wider review is currently underway and HEFCE have published further information about it on behalf of all the funders at www.hefce.ac.uk/whatwedo/lt/publicinfo/review/. We anticipate starting the implementation of any changes deemed necessary in 2017.

Part 2: A proposal to change the way we group subjects on Unistats.

Part 1: Publication thresholds on Unistats

How the data are provided on Unistats currently

20. Our current approach and the rationale for it are outlined in paragraph 11. They are as follows:
    a. For the National Student Survey (NSS), 50 per cent of the eligible students must have responded and these must represent at least 23 students.
    b. For the Destinations of Leavers from Higher Education (DLHE), the publication thresholds differ in that the number of students covered by the indicator must represent a full-person equivalent (FPE) of at least 22[3]. The 50 per cent threshold is not applied, but for salary information at least 50 per cent of the relevant students (those who are employed full-time) must have specified a salary. Thus, for salary data to be published, at least 22.5 students who are employed full-time must have specified a salary, and these must represent at least 50 per cent of the students employed full-time.
    c. For individualised data from the Higher Education Statistics Agency student record, the Individualised Learner Record and the Lifelong Learning Wales Record, there must be at least 22.5 FPE.

21. Where data do not meet the publication threshold at course level, they are aggregated: that is to say, the responses from two years or from a broader subject area are added together, but no further weightings are applied.

22. Aggregation occurs in the following order until data that meet the thresholds are achieved:
    - course level, most recent two years
    - current subject level 3, most recent year (108 subjects)
    - current subject level 3, most recent two years
    - current subject level 2, most recent year (42 subjects)
    - current subject level 2, most recent two years
    - current subject level 1, most recent year (21 subjects)
    - current subject level 1, most recent two years.

Modelling of 2014 KIS data

23. The current data elements presented on the Unistats web-site that are subject to publication thresholds are:
    - qualifications on entry
    - tariff scores of students on the course
    - continuation rates from year one into year two
    - class of degree

[3] FPE is a measure of headcount used where students are studying more than one subject as part of an award. For example, a student studying Engineering and French would typically count as half an FPE in Engineering and half an FPE in French. When considered over all subjects, every student will amount to one FPE.

- destinations (for instance, whether students go on to work, study or both)
- job type
- top 10 common jobs
- NSS (question scale, student union and overall satisfaction scores).

24. We have used the DLHE destination and salary data and the NSS data for our modelling, as these elements currently have the lowest levels of publishable data.

25. As only a small proportion of alternative providers[4] with data on Unistats currently participate in relevant surveys, we have not included data for them in our modelling as the results would not be comparable with other institution types.

26. Using the 2014 KIS data, we have summarised the number of KIS courses that are published against the different aggregation levels as a baseline for our modelling. We then reprocessed the same KIS submission data with different aggregation thresholds, to assess whether these had a significant impact on the amount of publishable data. We have modelled the results of reducing both the headcount required and the response rate threshold. The modelling scenarios are:
    - baseline
        - Headcount or FPE ≥ 22.5 and where applicable response rate ≥ 50 per cent (current thresholds)
    - headcount modelling (maintain response rate of 50 per cent)
        - Headcount or FPE ≥ 15
        - Headcount or FPE ≥ 10
        - Headcount or FPE ≥ 5
    - response rate modelling
        - Headcount or FPE ≥ 22.5 and where applicable response rate ≥ 40 per cent
        - Headcount or FPE ≥ 10 and where applicable response rate ≥ 40 per cent.

27. The results of the modelling are provided in **Annexes A** and **B**. **Annex A** provides a summary of the percentages of the number of courses that are publishable (broken down by aggregation type) for the above scenarios. **Annex B** provides the number of courses that are publishable. For simplicity headcount and FPE have been abbreviated to 'HC' in the tables in the annexes.

Measuring uncertainty

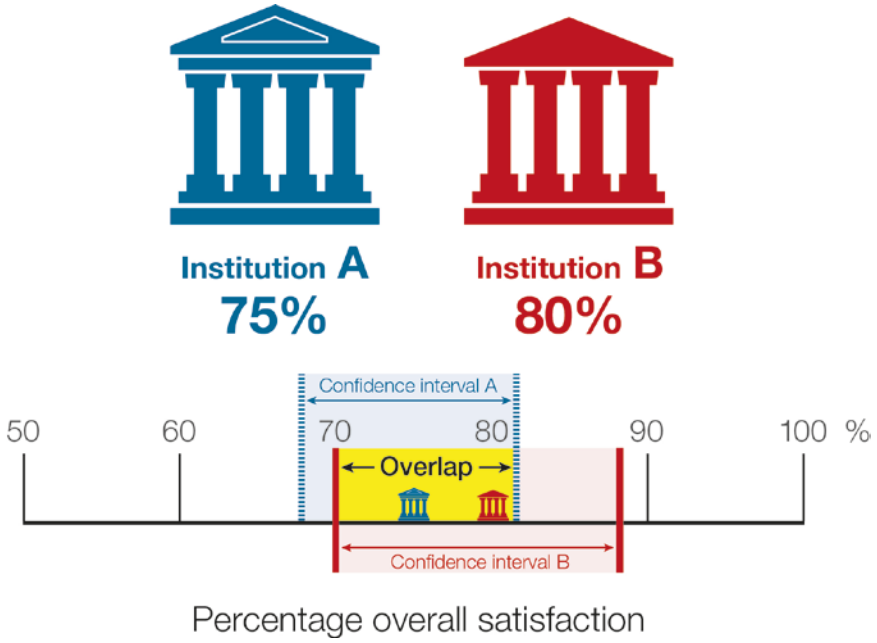
28. An important consideration when assessing whether to lower the publication thresholds is whether this would increase uncertainty in the data. The statistical tool that we have used to measure the level of

[4] An alternative provider means any provider of higher education courses which is not in direct receipt of recurrent funding from one of the funding bodies, or which does not receive direct recurrent public funding, and is not a further education institution.

uncertainty in data is the confidence interval. This is a numeric range with a minimum and maximum bound, within which the true value of an unknown value is likely to fall. This is useful where a hypothetical value has been generated using a sample, providing an estimate of the value for the population. Generally, the smaller the difference between the minimum and the maximum of the confidence interval, the greater the certainty the generated value is representative of the true value (or 'population value').

29. To help us assess whether changing the publication thresholds may increase uncertainty in the data, we considered the NSS data only, taking the percentage agreement scores for question 22 (overall satisfaction) for each course included in the KIS. We considered the spread of the confidence intervals for two modelling scenarios, using headcount thresholds of 10 and 23 with response rates of 50 per cent in both cases. Overlapping confidence intervals indicate that values are not significantly different (meaning that we cannot be sure that they are different).

Figure 1: Confidence intervals overlap



30. Consider the scenario in Figure 1. Institution A has a score of 75 per cent while Institution B has a score of 80 per cent with confidence intervals of 68 per cent to 81 percent and 70 per cent to 88 per cent respectively. These confidence intervals mean that the true value of Institution A's score could be anywhere between 68 and 81 per cent and that of Institution B anywhere between 70 and 88 per cent. The overlap from 70 per cent to 81 per cent between the two institutions' possible true scores means that we cannot be sure their scores are very different. Thus, by calculating confidence intervals and the incidence of overlapping intervals, we are able to assess how much lowering thresholds affects uncertainty in the data.

31. The results of calculating the proportion of confidence interval overlaps across different subject groups are shown in **Annex C**.

Summary of results obtained from data modelling using lower thresholds

32. The modelling results suggest that reducing the response rate (RR) to 40 per cent for headcount or FPE 10 and 23 would have minimal impact. This is not surprising, as the collection process for the NSS is designed to ensure a 50 per cent response rate in each subject area at each institution. This means that the contractors generally achieve a response rate of 50 per cent. Table 1 shows an extract of Table A2 in **Annex A**. The same patterns can be seen in Tables A1 and A3.

Table 1: Extract from Table A2 (NSS data)

		HC=10 RR=40%	HC=10 RR=50%	HC=23 RR=40%	HC=23 RR=50%
All courses					
	KISCOURSE	33%	33%	21%	21%
	Level 3	44%	44%	49%	49%
	Level 2	4%	4%	7%	7%
	Level 1	3%	3%	4%	4%
	Not publishable	15%	15%	19%	19%
Split by institution type					
HEIs	KISCOURSE	35%	34%	23%	23%
	Level 3	47%	47%	54%	54%
	Level 2	4%	4%	7%	7%
	Level 1	3%	3%	3%	3%
	Not publishable	11%	11%	13%	14%
FEIs	KISCOURSE	21%	20%	8%	7%
	Level 3	21%	20%	16%	16%
	Level 2	4%	4%	4%	4%
	Level 1	8%	8%	10%	10%
	Not publishable	46%	49%	62%	63%

33. With current thresholds (FPE=23), about one-fifth of the NSS (21 per cent) and DLHE destinations (19 per cent) data get published at course level (81 per cent and 84 per cent publishable overall respectively). For salary data, this figure is 3 per cent (66 per cent publishable overall).

34. Reducing thresholds to 15 has some impact on course-level data. For the NSS data, the percentage moves from 21 per cent to 28 per cent at course level (81 per cent to 83 per cent overall) for the sector as a whole. We see a similar increase at course level for HEIs (23 per cent to 29 per cent) and for FEIs (7 per cent to 14 per cent). There is more impact for FEIs when considering overall publishable data, which moves from 37 per cent to 45 per cent, whereas there is an increase of only two percentage points for HEIs. We see similar patterns for salary and DLHE destinations data.

Table 2: Extract from Table A2 (NSS data)

			HC=15 RR=50%	HC=23 RR=50%
All courses				
		KISCOURSE	28%	21%
		Not publishable	17%	19%
Split by institution type				
HEIs		KISCOURSE	29%	23%
		Not publishable	12%	14%
FEIs		KISCOURSE	14%	7%
		Not publishable	55%	63%
Split by institution type and mode of study				
HEIs	FT	KISCOURSE	33%	26%
		Not publishable	2%	3%
	PT	KISCOURSE	5%	3%
		Not publishable	77%	83%
FEIs	FT	KISCOURSE	19%	10%
		Not publishable	43%	52%
	PT	KISCOURSE	2%	1%
		Not publishable	85%	91%

35. When reducing thresholds to 10 and keeping the current response rate at 50 per cent, there is further improvement in the number of courses that can be published for the NSS data. At course level, this increase is from 21 per cent to 33 per cent, and the decrease in the number of courses that are not publishable is 19 per cent to 15 per cent. FEIs see a further improvement in their overall publishable data, which increase to 51 per cent (from 37 per cent currently).

Table 3: Extract from Table A2 (NSS data)

			HC=10 RR=50%	HC=23 RR=50%
All courses				
		KISCOURSE	33%	21%
		Not publishable	15%	19%
Split by institution type				
HEIs		KISCOURSE	34%	23%
		Not publishable	11%	14%
FEIs		KISCOURSE	20%	7%
		Not publishable	49%	63%
Split by institution type and mode of study				
HEIs	FT	KISCOURSE	39%	26%
		Not publishable	2%	3%
	PT	KISCOURSE	6%	3%
		Not publishable	71%	83%

FEIs	FT	KISCOURSE	26%	10%
		Not publishable	36%	52%
	PT	KISCOURSE	4%	1%
		Not publishable	81%	91%

36. We assessed the potential impact of a threshold reduction on the reliability of the data by calculating confidence intervals, as described in paragraph 28. A summary of results by subject area is available at **Annex C**. These results show that the incidence of overlapping confidence intervals does not increase significantly when the data are modelled with a threshold of 10, when compared with the existing threshold. The results of our modelling therefore suggest that reducing the threshold would not significantly increase uncertainty.

Summary of statistical advice

37. We commissioned expert independent statistical advice on the results obtained, asking what the advantages and disadvantages of lowering thresholds were as well as implications for issues of robustness (relating to bias and uncertainty in the data) and data protection (relating to disclosure). The advice received from Professor William Browne of the University of Bristol can be found in **Annex D**. It highlights the advantages of lowering thresholds but also presents the disadvantages of doing so. In summary, Professor Browne advises that:
    a. Reducing the headcount will increase the likelihood of disclosing information about students. However, this is mitigated by the nature of data on Unistats: it would be difficult to identify individual student responses with a headcount or FPE of 10 or above and the risk of publishing sensitive data on individual students on Unistats is low. The use of interquartile ranges, in salary data for example, provides further mitigation.
    b. Reducing thresholds increases uncertainty but not substantially, as the analysis using confidence intervals (**Annex C**) shows.
    c. The response rate is more important than headcount when considering bias in the data. However, the effect of a biased sample may be more noticeable if the headcount is small. This would depend, however, on whether the student make-up was unusual. Bias occurs if there is a correlation between the propensity to respond to the questionnaire and the honest answers to the questions. Response rates might differ depending on the gender, culture, ethnicity and age of students. There might also be the issue of what Professor Browne calls the 'apathetic' middle – in the NSS for example, students who are very happy or have reason to complain are more likely to respond than the rest.

38. As well as seeking advice on the implications of lowering thresholds, we asked Professor Browne to comment on issues relating to the presentation of data on Unistats, including for suggestions for how to communicate statistical uncertainty.

39. He made several recommendations to improve transparency, which included:

- where data are aggregated, including information on the aggregation method in the tooltip text displayed to the user
- in addition to the sample size, displaying the response rate for each data value
- displaying the sample size for aggregated data
- providing the option for the user to aggregate data to the same level for each course in the comparison
- displaying either a 95 per cent confidence interval or some other form of uncertainty quantification for the binary variables, or otherwise identifying statistically different values.

40. We will seek to implement the first three of these recommendations for the publication of the 2015 KIS, regardless of the outcome of this consultation. We will also investigate ways of communicating uncertainty, and implement an approach to this to the same timescale.

41. Since we are currently reviewing the KIS and Unistats, we do not consider it appropriate to make fundamental changes to the way in which we display data on the site at this time. The options we will explore are likely to include adding confidence interval information to the tooltip text, and explanatory text and graphics or video to guide the user on considering statistical uncertainty when comparing values. We will test before introducing any changes to ensure that they are well understood by users and achieve their aim of increasing understanding.

42. We acknowledge, however, that the findings from our modelling and the statistician's recommendations indicate more fundamental changes to the design approach may be desirable. We will consider this as part of the design of any successor site. Potential solutions could include displaying data values with associated confidence intervals, or banding or colour-coding values rather than displaying actual values. We have not yet identified any exemplars of this type of presentation and would welcome suggestions on alternative presentations that allow users to make comparisons between data with an awareness of the uncertainty around the values presented. Similarly, we will further consider the fourth recommendation (to allow the user to aggregate data to the same level for each course in the comparison) when designing any future solution.

Focus group meeting

43. The UK-wide focus group, convened on 7 November 2014, comprised sector body representatives and representatives from institutions nominated by Universities UK, GUILDHE and the Association of Colleges. The list of participants is provided in **Annex E**.

44. Before discussions[5] at the meeting, the focus group considered the results of the data modelling and the statistical advice, and was asked to comment on the following:

[5] A summary of the discussion at the focus group can be found on the consultation web page.

- whether the evidence presented supported a reduction in thresholds and if so, which options we should consult on
- the sector's likely main concerns if we proposed to change thresholds
- how we might present the information in the consultation so as to ensure it was clear and at the right level of detail
- how we could best present information about uncertainty on Unistats (as recommended by Professor Browne).

Consultation proposal

45. The information in **Annex B** gives an idea of the number of courses published on Unistats for different parts of the sector. As the numbers show, HEIs have more presence on Unistats since they have around 27,000 courses published; FEIs have around 3,000 courses.

46. The results of the data modelling have shown that lowering response rate thresholds would make little difference in the amount of publishable data. Lowering headcount or FPE thresholds, however, would increase the amount of publishable data, both overall and at course level, and the lower the threshold the greater the increase would be. A headcount or FPE threshold lower than 10 is likely to be unacceptable, however, as it could potentially expose us to risk in terms of our data protection obligations.

47. Based on the evidence presented to it, the focus group was broadly in favour of proposing lowering thresholds to a headcount or FPE of 10, while keeping the response rate at 50 per cent. It also recognised the importance of communicating information on uncertainty in the data and that any changes to our current approach should benefit prospective students using this information to inform their higher education choices.

Impact on NSS results

48. The NSS results are published in three locations on behalf of the funding bodies:
    - the NSS dissemination web-site (a secure site available to institutions participating in the NSS)
    - in the KIS on the Unistats web-site
    - on the HEFCE web-site

49. On the NSS dissemination site, institutions can view their own data at both the lower threshold of 10 respondents and the upper threshold (a headcount of 23 and response rate of 50 per cent) by a range of variables including age, gender and subject level. They can also view other institutions' data in less detail at the upper threshold. HEFCE also publishes NSS summary data and data by subject on its own web-site at the upper threshold[6]. Should thresholds be lowered on Unistats, the same approach will be used for the NSS dissemination web-site. Thus an additional advantage of lowering thresholds to a headcount of 10 is that it increases alignment between Unistats and the NSS dissemination site, which should reduce queries from academics.

[6] See www.hefce.ac.uk/whatwedo/lt/publicinfo/nss/data/

50. On the other hand, should lower thresholds be implemented, institutions might see a large difference in their results from the previous year. The scores could either drop or increase where data were no longer aggregated. For example, courses in Nutrition and Dietetics whose data were currently aggregated might have satisfaction scores of 30 per cent and 70 per cent respectively aggregated to 50 per cent on Unistats. If the courses now met the lower publication thresholds, their satisfaction scores would no longer be aggregated. The Nutrition course would see its score drop to 30 per cent and the Dietetics course a rise in its score to 70 per cent when the new thresholds were implemented.

51. We have considered aggregating data over a larger number of years but are not proposing this, due to the sector's concerns that data from a number of years ago might not reflect current provision.

Consultation question 1

a. Are you in favour of lowering the headcount threshold for publication of data on Unistats to 10, while retaining the response rate at 50 per cent?
b. What are the reasons for your response?
c. Do you have any other comments on this proposal?

Part 2. Subject groupings on Unistats

52. Currently, data on Unistats are aggregated using a subject hierarchy to which Joint Academic Coding System (JACS) codes have been mapped[7]. If aggregation is required to meet publication thresholds (and aggregating course data over two years is not sufficient to achieve this), aggregation occurs across the level 3 subject that corresponds to the JACS code for the course, followed by the level 2 subject, then the level 1 subject, until the threshold is met.

53. The subject groupings within this hierarchy are also used on the Unistats web-site to enable users to browse courses by subject.

54. Feedback from the sector indicates that the current subject hierarchy does not always reflect organisational structures and lacks detail in some areas, and is therefore not the most appropriate way to aggregate data in some subject areas. We are considering revising the subject groupings and hierarchy used for Unistats to better reflect typical organisational structures, and to provide further detail in some areas.

55. Our aim is to identify whether we can improve the way courses are currently grouped when aggregation is required to meet publication thresholds, so that the data published on Unistats are as relevant as possible to the course for which they are being displayed, and therefore likely to provide a good indication of a student's potential experience.

[7] For directly funded FECs in Wales JACS codes are collected in field HE16 of the LLWR

Subject grouping principles

56. Our proposed groupings are based on the following principles:
    - there should only be three levels
    - as far as possible, each level should include aggregations from below
    - at the lowest level, data that are not course-level should be publishable for a reasonable number of institutions
    - subjects should be intuitive or describable
    - as far as possible, the groups should aim to reflect how institutions structure their provision.

Development of the new subject groupings

57. To develop the new subject groupings, we calculated the following measures for each JACS 3.0 code using 2012/13 student data (Higher Education Statistics Agency data and Lifelong Learning Wales Records(LLWR)) in Wales/Individual Learning Record in England, for undergraduate entrants:
    - FPE
    - number of courses associated with the JACS code
    - number of potentially publishable cohorts associated with the JACS cohort[8].

    We then used these measures to identify how subjects could potentially be grouped giving consideration to the principles outlined in paragraph 56.

Consultation proposals

58. Our proposed revised subject hierarchy is described in **Annex F**. This includes the current mapping for each JACS code and the proposed revised mapping, along with the measures we have calculated for the code. For example, JACS code B600 (Aural and oral sciences) currently maps to 'Aural and oral sciences' at level 3, then to 'Other subjects allied to medicine' at level 2 and to 'Subjects allied to medicine' at level 1. We propose that B600 should now map to 'Aural and oral sciences' at level 3, then to 'General anatomy, physiology and pathology programmes' at level 2 and to 'Subjects allied to medicine' at level 1. While the proposal at **Annex F** seeks to reflect the principles, we are not experts in institutional structures or subject taxonomies, and are therefore seeking detailed comments on the mapping (which should, at this stage, be considered no more than indicative).

Consultation question 2

a. Do you agree that the current subject groupings used for Unistats require amendment?
b. Do you agree with the principles for grouping subjects (See Paragraph 56)?
c. Do you have any comments on these principles?
d. Would any areas in the proposed subject hierarchy be problematic for your institution, and can you suggest potential changes to resolve this?

[8] 'Potentially publishable' courses were considered to be those with a cohort size of ten or more.

59. If we were to change the subject groupings the NSS data at the subject level would not be comparable with previous years in the amended groupings. We recognise that the sector may be particularly interested in the outcomes of the NSS in 2015, given that the first cohort of students paying higher fees will complete the survey in this year. We are therefore seeking views about whether we should defer implementing such changes until 2016 (this would not affect the timing of any reduction to publication thresholds). Deferral to 2016 would have the additional advantage of allowing us more time to work with the sector to finalise any changes following our analysis of responses. It would mean, however, that changes to improve the alignment of the subject groupings used in aggregation with organisational structures would not be implemented for a further year.

Consultation question 3

a. Would you support any changes to the subject groupings being implemented in 2015, or would you prefer deferral until 2016?
b. Please explain the reason for your response.

Next Steps

60. We commit to read, record and analyse the views of every response to this consultation in a consistent manner. For reasons of practicality, usually a fair and balanced summary of responses rather than the individual responses themselves will inform any decision made. The merit of the arguments made is as important as representativeness of the arguments made for this consultation. Responses from organisations or representative bodies which have high relevance or interest in the area under consultation, or are likely to be affected most by the proposals, are likely to carry more weight than those with little or none.

61. We will publish an analysis of the consultation responses and an explanation of how the responses were considered in our subsequent decision. We may publish individual responses to the consultation. Additionally, all responses may be disclosed on request, under the terms of the Freedom of Information Act. The act gives a public right of access to any information held by a public authority, in this case HEFCE, who will receive the responses on behalf of all the funders. This includes information provided in response to a consultation. HEFCE has a responsibility to decide whether any responses, including information about the identity of respondents, should be made public or treated as confidential. It can refuse to disclose information only in exceptional circumstances. This means that responses to this consultation are unlikely to be treated as confidential except in very particular circumstances. For further information about the act see www.ico.gov.uk.

62. The analysis of responses and subsequent decision relating to the first part of this consultation (the proposal to lower thresholds) will be considered in March by a sub-group of the Higher Education Public Information Steering

Group (HEPISG), which oversees changes to the Unistats site[9]. If the decision is to lower thresholds to a headcount of 10 with a response rate of 50 per cent, then we intend to implement this, subject to an impact assessment of the necessary system changes, in the 2015 KIS. We do not anticipate a need for institutions to alter their processes and systems for making the KIS return.

63. For the second part of this consultation (the proposal to change subject groupings), we will work with the sector to develop a final structure once responses have been received. Timescales for this will be determined by the views expressed on when changes should be implemented.

Further information/responses to

64. For further information, contact Dr Cliona O'Neill (tel 029 2068 2283; email cliona.oneill@hefcw.ac.uk).

65. Responses to this consultation should be made online by **noon on Friday 13 February 2015,** using the response form on the HEFCE website at www.hefce.ac.uk/whatwedo/lt/publicinfo/kis/aggregation/.

[9] HEPISG is a UK-wide body with higher education stakeholder and sector group representation. It oversees the development of public information about higher education.

Annex A: Course publication percentages

Note: 'HC'='headcount'; 'RR'='response rate'; 'HEI'='higher education institution' 'FEI'='further education institution; 'FT'='full-time'; 'PT'='part-time.

Table A1: Destinations of Leavers from Higher Education destinations data

			HC=5 RR=50%	HC=10 RR=40%	HC=10 RR=50%	HC=15 RR=50%	HC=23 RR=40%	HC=23 RR=50%
All courses								
		KISCOURSE	39%	31%	31%	25%	19%	19%
		Level 3	44%	46%	46%	48%	48%	48%
		Level 2	5%	6%	6%	7%	10%	10%
		Level 1	4%	6%	6%	6%	8%	8%
		Not publishable	8%	11%	11%	13%	16%	16%
Split by institution type								
HEIs		KISCOURSE	41%	32%	32%	27%	20%	20%
		Level 3	46%	49%	49%	51%	51%	51%
		Level 2	5%	7%	7%	8%	11%	11%
		Level 1	3%	5%	5%	6%	7%	7%
		Not publishable	5%	7%	7%	8%	10%	10%
FEIs		KISCOURSE	25%	17%	17%	11%	6%	6%
		Level 3	29%	27%	27%	24%	18%	18%
		Level 2	3%	3%	3%	4%	5%	5%
		Level 1	7%	9%	9%	9%	10%	10%
		Not publishable	36%	44%	44%	52%	61%	61%
Split by institution type and mode of study								
HEIs	FT	KISCOURSE	45%	36%	36%	30%	23%	23%
		Level 3	48%	52%	52%	55%	56%	56%
		Level 2	4%	7%	7%	8%	11%	11%
		Level 1	2%	3%	3%	5%	6%	6%
		Not publishable	2%	2%	2%	3%	4%	4%
	PT	KISCOURSE	16%	10%	10%	8%	5%	5%
		Level 3	36%	30%	30%	24%	19%	19%
		Level 2	8%	8%	8%	7%	5%	5%
		Level 1	14%	16%	16%	15%	18%	18%
		Not publishable	26%	36%	36%	46%	53%	53%
FEIs	FT	KISCOURSE	30%	21%	21%	15%	8%	8%
		Level 3	31%	30%	30%	27%	22%	21%
		Level 2	3%	4%	4%	5%	6%	6%
		Level 1	7%	10%	10%	10%	11%	11%
		Not publishable	29%	35%	35%	43%	53%	53%
	PT	KISCOURSE	10%	5%	5%	3%	2%	2%
		Level 3	25%	18%	18%	14%	9%	9%
		Level 2	3%	2%	2%	1%	1%	1%
		Level 1	6%	9%	9%	8%	5%	5%
		Not publishable	56%	66%	66%	74%	83%	83%

Table A2: National Student Survey data

			HC=5 RR=50%	HC=10 RR=40%	HC=10 RR=50%	HC=15 RR=50%	HC=23 RR=40%	HC=23 RR=50%
All courses								
		KISCOURSE	41%	33%	33%	28%	21%	21%
		Level 3	39%	44%	44%	47%	49%	49%
		Level 2	4%	4%	4%	5%	7%	7%
		Level 1	3%	3%	3%	3%	4%	4%
		Not publishable	13%	15%	15%	17%	19%	19%
Split by institution type								
HEIs		KISCOURSE	42%	35%	34%	29%	23%	23%
		Level 3	42%	47%	47%	51%	54%	54%
		Level 2	4%	4%	4%	5%	7%	7%
		Level 1	3%	3%	3%	3%	3%	3%
		Not publishable	9%	11%	11%	12%	13%	14%
FEIs		KISCOURSE	27%	21%	20%	14%	8%	7%
		Level 3	22%	21%	20%	18%	16%	16%
		Level 2	3%	4%	4%	5%	4%	4%
		Level 1	6%	8%	8%	8%	10%	10%
		Not publishable	42%	46%	49%	55%	62%	63%
Split by institution type and mode of study								
HEIs	FT	KISCOURSE	47%	39%	39%	33%	26%	26%
		Level 3	46%	53%	53%	57%	61%	61%
		Level 2	4%	5%	5%	6%	8%	8%
		Level 1	1%	2%	2%	2%	3%	3%
		Not publishable	2%	2%	2%	2%	3%	3%
	PT	KISCOURSE	10%	7%	6%	5%	3%	3%
		Level 3	14%	12%	11%	9%	7%	7%
		Level 2	6%	4%	3%	3%	2%	2%
		Level 1	11%	9%	8%	7%	5%	5%
		Not publishable	59%	68%	71%	77%	82%	83%
FEIs	FT	KISCOURSE	34%	27%	26%	19%	10%	10%
		Level 3	26%	25%	25%	22%	21%	20%
		Level 2	4%	5%	4%	6%	6%	5%
		Level 1	7%	9%	9%	10%	12%	12%
		Not publishable	30%	34%	36%	43%	50%	52%
	PT	KISCOURSE	8%	4%	4%	2%	1%	1%
		Level 3	12%	10%	8%	7%	4%	4%
		Level 2	2%	2%	2%	2%	1%	1%
		Level 1	4%	6%	5%	4%	4%	4%
		Not publishable	74%	78%	81%	85%	91%	91%

Table A3: Destination of Leavers from Higher Education salary data

			HC=5 RR=50%	HC=10 RR=40%	HC=10 RR=50%	HC=15 RR=50%	HC=23 RR=40%	HC=23 RR=50%
All courses								
		KISCOURSE	18%	9%	9%	6%	3%	3%
		Level 3	42%	45%	42%	39%	36%	35%
		Level 2	9%	13%	12%	14%	15%	14%
		Level 1	7%	9%	9%	12%	15%	14%
		Not publishable	24%	24%	27%	30%	31%	34%
Split by institution type								
HEIs		KISCOURSE	20%	10%	10%	6%	3%	3%
		Level 3	46%	50%	47%	44%	40%	39%
		Level 2	10%	14%	14%	15%	16%	15%
		Level 1	7%	10%	10%	13%	17%	16%
		Not publishable	17%	16%	19%	22%	23%	26%
FEIs		KISCOURSE	2%	0%	0%	0%	0%	0%
		Level 3	11%	5%	5%	3%	1%	2%
		Level 2	3%	2%	2%	1%	1%	1%
		Level 1	7%	5%	5%	4%	1%	1%
		Not publishable	77%	87%	87%	92%	96%	97%
Split by institution type and mode of study								
HEIs	FT	KISCOURSE	21%	11%	11%	7%	3%	3%
		Level 3	50%	55%	51%	48%	45%	44%
		Level 2	11%	16%	15%	17%	19%	18%
		Level 1	7%	9%	10%	13%	19%	17%
		Not publishable	10%	9%	12%	14%	14%	18%
	PT	KISCOURSE	8%	3%	3%	2%	1%	1%
		Level 3	20%	17%	17%	13%	10%	10%
		Level 2	5%	4%	4%	3%	2%	2%
		Level 1	10%	12%	12%	11%	9%	8%
		Not publishable	58%	63%	64%	71%	78%	78%
FEIs	FT	KISCOURSE	3%	0%	0%	0%	0%	0%
		Level 3	9%	4%	4%	2%	1%	2%
		Level 2	4%	3%	3%	2%	1%	1%
		Level 1	7%	6%	5%	4%	1%	1%
		Not publishable	77%	86%	87%	92%	96%	97%
	PT	KISCOURSE	2%	0%	0%	0%	0%	0%
		Level 3	14%	8%	8%	4%	1%	2%
		Level 2	0%	0%	0%	0%	0%	0%
		Level 1	6%	4%	4%	3%	1%	1%
		Not publishable	78%	88%	88%	92%	97%	97%

Annex B: Course publication numbers

Note: 'HC'='headcount'; 'RR'='response rate'; 'HEI'='higher education institution' 'FEI'='further education institution; 'FT'='full-time'; 'PT'='part-time.

Table B1: DLHE destinations data

			HC=5 RR=50%	HC=10 RR=40%	HC=10 RR=50%	HC=15 RR=50%	HC=23 RR=40%	HC=23 RR=50%
All courses								
		KISCOURSE	11,682	9,147	9,147	7,491	5,639	5,640
		Level 3	13,230	13,896	13,896	14,370	14,222	14,221
		Level 2	1,379	1,900	1,900	2,171	2,964	2,964
		Level 1	1,090	1,695	1,695	1,890	2,280	2,280
		Not publishable	2,508	3,251	3,251	3,967	4,784	4,784
		Total	29,889	29,889	29,889	29,889	29,889	29,889
Split by institution type								
HEIs		KISCOURSE	10,852	8,578	8,578	7,109	5,425	5,425
		Level 3	12,253	12,986	12,986	13,568	13,612	13,612
		Level 2	1,270	1,791	1,791	2,030	2,808	2,808
		Level 1	853	1,382	1,382	1,584	1,952	1,952
		Not publishable	1,281	1,772	1,772	2,218	2,712	2,712
		Total	26,509	26,509	26,509	26,509	26,509	26,509
FEIs		KISCOURSE	830	569	569	382	214	215
		Level 3	977	910	910	802	610	609
		Level 2	109	109	109	141	156	156
		Level 1	237	313	313	306	328	328
		Not publishable	1,227	1,479	1,479	1,749	2,072	2,072
		Total	3,380	3,380	3,380	3,380	3,380	3,380
Split by institution type and mode of study								
HEIs	FT	KISCOURSE	10,272	8,206	8,206	6,840	5,257	5,257
		Level 3	10,968	11,928	11,928	12,695	12,937	12,937
		Level 2	988	1,512	1,512	1,779	2,617	2,617
		Level 1	351	796	796	1,035	1,291	1,291
		Not publishable	350	487	487	580	827	827
		Total	22,929	22,929	22,929	22,929	22,929	22,929
	PT	KISCOURSE	580	372	372	269	168	168
		Level 3	1,285	1,058	1,058	873	675	675
		Level 2	282	279	279	251	191	191
		Level 1	502	586	586	549	661	661
		Not publishable	931	1,285	1,285	1,638	1,885	1,885
		Total	3,580	3,580	3,580	3,580	3,580	3,580
FEIs	FT	KISCOURSE	736	521	521	356	199	200
		Level 3	746	740	740	668	525	524
		Level 2	77	91	91	131	147	147
		Level 1	180	233	233	233	279	279
		Not publishable	701	855	855	1,052	1,290	1,290
		Total	2,440	2,440	2,440	2,440	2,440	2,440
	PT	KISCOURSE	94	48	48	26	15	15
		Level 3	231	170	170	134	85	85
		Level 2	32	18	18	10	9	9
		Level 1	57	80	80	73	49	49
		Not publishable	526	624	624	697	782	782
		Total	940	940	940	940	940	940

Table B2: NSS data

			HC=5 RR=50%	HC=10 RR=40%	HC=10 RR=50%	HC=15 RR=50%	HC=23 RR=40%	HC=23 RR=50%
All courses								
		KISCOURSE	12,113	9,936	9,790	8,233	6,324	6,281
		Level 3	11,787	13,253	13,226	14,015	14,782	14,752
		Level 2	1,186	1,325	1,315	1,550	2,062	2,052
		Level 1	908	973	934	969	1,096	1,083
		Not publishable	3,895	4,402	4,624	5,122	5,625	5,721
		Total	29,889	29,889	29,889	29,889	29,889	29,889
Split by institution type								
HEIs		KISCOURSE	11,215	9,238	9,125	7,743	6,066	6,029
		Level 3	11,032	12,542	12,538	13,406	14,228	14,217
		Level 2	1,077	1,191	1,190	1,395	1,912	1,910
		Level 1	707	696	672	699	760	760
		Not publishable	2,478	2,842	2,984	3,266	3,543	3,593
		Total	26,509	26,509	26,509	26,509	26,509	26,509
FEIs		KISCOURSE	898	698	665	490	258	252
		Level 3	755	711	688	609	554	535
		Level 2	109	134	125	155	150	142
		Level 1	201	277	262	270	336	323
		Not publishable	1,417	1,560	1,640	1,856	2,082	2,128
		Total	3,380	3,380	3,380	3,380	3,380	3,380
Split by institution type and mode of study								
HEIs	FT	KISCOURSE	10,864	8,991	8,899	7,569	5,941	5,912
		Level 3	10,521	12,104	12,139	13,080	13,963	13,964
		Level 2	864	1,064	1,069	1,297	1,831	1,832
		Level 1	326	364	385	460	582	593
		Not publishable	354	406	437	523	612	628
		Total	22,929	22,929	22,929	22,929	22,929	22,929
	PT	KISCOURSE	351	247	226	174	125	117
		Level 3	511	438	399	326	265	253
		Level 2	213	127	121	98	81	78
		Level 1	381	332	287	239	178	167
		Not publishable	2,124	2,436	2,547	2,743	2,931	2,965
		Total	3,580	3,580	3,580	3,580	3,580	3,580
FEIs	FT	KISCOURSE	824	658	627	468	248	242
		Level 3	639	618	610	547	519	500
		Level 2	94	115	106	136	142	134
		Level 1	163	224	214	236	301	288
		Not publishable	720	825	883	1,053	1,230	1,276
		Total	2,440	2,440	2,440	2,440	2,440	2,440
	PT	KISCOURSE	74	40	38	22	10	10
		Level 3	116	93	78	62	35	35
		Level 2	15	19	19	19	8	8
		Level 1	38	53	48	34	35	35
		Not publishable	697	735	757	803	852	852
		Total	940	940	940	940	940	940

Table B3: DLHE salary data

			HC=5 RR=50%	HC=10 RR=40%	HC=10 RR=50%	HC=15 RR=50%	HC=23 RR=40%	HC=23 RR=50%
All courses								
		KISCOURSE	5,258	2,738	2,734	1,645	841	841
		Level 3	12,627	13,394	12,530	11,676	10,718	10,421
		Level 2	2,822	3,881	3,695	4,105	4,354	4,120
		Level 1	2,100	2,732	2,831	3,571	4,626	4,276
		Not publishable	7,082	7,144	8,099	8,892	9,350	10,231
		Total	29,889	29,889	29,889	29,889	29,889	29,889
Split by institution type								
HEIs		KISCOURSE	5,174	2,727	2,723	1,644	840	840
		Level 3	12,265	13,216	12,358	11,580	10,669	10,367
		Level 2	2,722	3,809	3,622	4,057	4,326	4,099
		Level 1	1,877	2,549	2,662	3,438	4,583	4,243
		Not publishable	4,471	4,208	5,144	5,790	6,091	6,960
		Total	26,509	26,509	26,509	26,509	26,509	26,509
FEIs		KISCOURSE	84	11	11	1	1	1
		Level 3	362	178	172	96	49	54
		Level 2	100	72	73	48	28	21
		Level 1	223	183	169	133	43	33
		Not publishable	2,611	2,936	2,955	3,102	3,259	3,271
		Total	3,380	3,380	3,380	3,380	3,380	3,380
Split by institution type and study mode								
HEIs	FT	KISCOURSE	4,894	2,618	2,614	1,581	802	802
		Level 3	11,559	12,600	11,760	11,097	10,305	10,003
		Level 2	2,547	3,654	3,476	3,954	4,257	4,030
		Level 1	1,530	2,107	2,222	3,057	4,250	3,943
		Not publishable	2,399	1,950	2,857	3,240	3,315	4,151
		Total	22,929	22,929	22,929	22,929	22,929	22,929
	PT	KISCOURSE	280	109	109	63	38	38
		Level 3	706	616	598	483	364	364
		Level 2	175	155	146	103	69	69
		Level 1	347	442	440	381	333	300
		Not publishable	2,072	2,258	2,287	2,550	2,776	2,809
		Total	3,580	3,580	3,580	3,580	3,580	3,580
FEIs	FT	KISCOURSE	63	9	9	–	–	–
		Level 3	231	106	100	60	35	39
		Level 2	96	70	71	44	24	18
		Level 1	170	148	134	103	30	20
		Not publishable	1,880	2,107	2,126	2,233	2,351	2,363
		Total	2,440	2,440	2,440	2,440	2,440	2,440
	PT	KISCOURSE	21	2	2	1	1	1
		Level 3	131	72	72	36	14	15
		Level 2	4	2	2	4	4	3
		Level 1	53	35	35	30	13	13
		Not publishable	731	829	829	869	908	908
		Total	940	940	940	940	940	940

Annex C: Measures of variance and uncertainty using NSS question 22 (overall satisfaction) data

Note: 'NSS'='National Student Survey'; 'HC'='headcount'; 'RR'='response rate'.

Table C1: Proportion of confidence interval overlaps across different subject groups

NSS level 1 subject code	NSS level 1 subject name	HC = 10, RR=50%	HC = 23, RR=50%
1	Medicine and dentistry	73%	72%
2	Subjects allied to medicine	92%	88%
3	Biological sciences	95%	94%
4	Veterinary science	97%	97%
5	Agriculture and related subjects	96%	93%
6	Physical sciences	97%	96%
7	Mathematical sciences	96%	94%
8	Computer science	91%	89%
9	Engineering and technology	93%	89%
A	Architecture, building and planning	97%	94%
B	Social studies	93%	90%
C	Law	97%	95%
D	Business and administrative studies	94%	92%
E	Mass communications and documentation	88%	85%
F	Languages	97%	95%
G	Historical and philosophical studies	98%	97%
H	Creative arts and design	92%	87%
I	Education	95%	93%
J	Combined	94%	93%
K	Initial teacher training	88%	83%
L	Geographical Studies	96%	95%
Multiple subjects		95%	93%

Annex D: A report for HEFCE on publication thresholds in the Unistats web-site

Professor William Browne, Professor of Statistics, Graduate School of Education, University of Bristol
28 October 2014

Introduction

The Unistats web-site (www.unistats.ac.uk) is a web-site produced by HEFCE as a publicly available interface into (primarily) student-reported data on all undergraduate courses at all UK higher education institutions. The data are made available in the form of aggregate summary measures, here for example percentages of students replying positively to questionnaire questions or 'typical' salary ranges reported by students post studying. Such data are made available only for courses (or groups of courses) where the response rate is above a specific threshold (currently 50 per cent) and the resulting headcount is also above a threshold (currently 23). In the case that these thresholds are not satisfied several approaches are adopted:

(i) The data for several years of responses for a course are merged until the conditions are satisfied.
(ii) The data are merged with 'similar' subjects (as defined in the hierarchical structuring of courses defined by HEFCE) at the same institution until the conditions are satisfied.
(iii) A null return is given for the specific data item.

The motivations behind the restricted access to direct data on courses with low headcounts and/or low response rates are many but can be split into three main headings which we will consider in more detail later:

(i) The data protection worry of unintentionally indirectly disclosing individual data on students.
(ii) The worry that the accuracy of data on returns with low headcounts and low response rates might be compromised in particular as the site does not offer any confidence intervals around statistics presented and low headcounts will result in wide confidence intervals.
(iii) The worry that a low response rate may result in bias in the statistics presented.

These motivations need to be balanced with the perceived problem of null returns reflecting negatively on the institutions concerned plus problems of potentially unfair comparisons across institutions through differing levels of aggregation skewing the results. The main focus of this report is to look at the potential advantages and disadvantages of reducing the headcount threshold from 23 to 10 for reporting of figures. It will firstly consider the advantages which are in fact well documented in the appendices of the specification document and then will consider the three motivations listed above and how changing the headcount threshold would impact on each of them. It will finish with some thoughts on ideas how the web-site could best consider incorporating uncertainty and other possible presentational improvements.

Data availability rates and associated benefits

As described in the specification document and in particular in the annex tables[10] clearly the lower the headcount threshold the larger the number of undergraduate courses for which data can be shown directly (at the KISCOURSE level) or aggregated at higher levels. There are three sources of data that are used in the web-site: destination on leaving higher education (DLHE) destination and salary datasets along with National student survey (NSS) data.

For illustration, if one considers the NSS data then reducing the threshold from 23 to 10 whilst maintaining a response rate of over 50 per cent will increase the percentage of courses that can be viewed at the KISCOURSE level from 21 per cent to 33 per cent and reduce the percentage of courses not publishable from 19 per cent to 15 per cent. Another interesting comparison is that when one considers the two finest levels of aggregation (KISCOURSE and Joint Academic Coding System (JACS) level 3) then the percentage that can be viewed at one or other level is 70 per cent for a 23 threshold and 77 per cent for a 10 threshold. This pattern repeats for higher levels of aggregation and shows that reducing the threshold basically allows finer grain data to be exposed but that this has a bigger effect for courses in moving them from aggregated data to KISCOURSE-level data than for the data that were originally not publishable.

One other way of presenting these data not considered in the annex is to tabulate the data in terms of percentage of the total student entry that each row describes, i.e. the fact that 21 per cent of KISCOURSE by institution combinations can be viewed directly suggests a poor access rate, however if let's say 80 per cent of students were contained in those combinations and therefore one might expect 80 per cent of the possible applicants to choose one of those combinations then perhaps things aren't so bad as the 21 per cent figure suggests. Taking this one step further perhaps the change of threshold may move the percentage of combinations to be viewed from 21 per cent to 33 per cent but this may only increase the proportion of students in those combinations from 80 per cent to 82 per cent. Without tabulating the data in this way one doesn't of course know if this is the case but one would expect the percentage of student entry to be far larger than the percentage of KISCOURSE by institution combinations.

A related point which might be helped by the increase from 21 per cent to 33 per cent is the comparability across institutions of similar data. Let's consider for illustration two institutions A and B and a KISCOURSE with across the sector poorer scores. Now institution A has a headcount of 25 whilst institution B has a headcount of 20 and therefore its data are aggregated up to the JACS level 3 level with another KISCOURSE which by chance has across the sector higher scores and a larger headcount resulting in a total headcount of 100. The reader would potentially therefore get two wrong impressions from the data:
(i) Institution B has a bigger cohort of students.

[10] The tables made available to Professor Browne were those in **Annexes A** and **B**.

(ii) Institution B has much higher scores (due in fact to 80 students not on that KISCOURSE). In this case lowering the threshold would alleviate the need to aggregate and remove these issues although it is an idealised example and in practice one would probably anticipate similar scores within JACS levels.

So to summarise possible advantages of reducing the threshold are:
(i) More course by institution combinations having data available at lower level of aggregations.
(ii) More course by institution combinations being comparable at the same level of aggregation.

Disclosure issues

Statistical disclosure and data protection are important topics and in particular here the issue would be that the sensitive data on individual students should not be disclosed. In fact such disclosure is a much bigger issue when we have anonymised individual data where specific characteristics of an individual might identify them and allow the reader access to other fields for that student. However in Unistats all data are aggregated.

With aggregate data disclosure is generally less of an issue. There are possibilities of what I will describe as 'full disclosure', i.e. if we know a student has responded (either because all students on the course responded or they have told us they responded) then we know some of their data. This would occur if there is a lack of variability in response for example if all students were 'not satisfied with the quality of the course' or all students were 'in a professional or managerial job'. There is in fact no sample size (headcount number) that guarantees that such a disclosure does not occur. That said, the smaller the required headcount the more likely to have all students responding and responding in the same manner.

A slightly less concerning disclosure is what I will describe as 'range disclosure' where for example if the range for a continuous response like salary was quoted, then we would be able to identify that a student earned more than (or equal to) the minimum quoted and less than (or equal to) the maximum. Here Unistats gets around this successfully by quoting only the inter-quartile range for salaries.

So to summarise reducing headcount will increase the likelihood of disclosure for combinations where there is a 100 per cent response rate.

Issues of accuracy/confidence intervals in the quoted statistics

Generally on the Unistats web-site statistics are quoted without any uncertainty estimates around the statistics. The one commendable exception is the salary information where a 'typical salary range' is given and in fact this is the case for displaying most other continuous variables on the site for example 'typical annual cost of private accommodation'. I think that here Unistats has done a good job although I am hoping the 'average' figures associated with the typical

range given are median salaries rather than means and perhaps this should be more explicit. I also found it confusing that when aggregate figures were given (and sometimes more than one average figure!) it was hard to know what subjects were aggregated over and this could be made more explicit.

Of course the inter-quartile range estimates will be sensitive to the sample collected and as with all statistics the larger the sample the more confidence we have in our estimates. We will illustrate this in more detail in the case of NSS variables next.

Many of the variables (in particular the NSS responses) that are displayed on the Unistats web-site are binary in nature, i.e. predominately yes/no questions that are then quoted as the percentage of students that answer the question yes. These numbers are quoted without any range/confidence interval given around the average figure. Interestingly I have been faced with a very similar question in a completely different application area, namely welfare assessment of chicken farms (see Main et al. (2013)). Here the farm assurance schemes wished to know the impact of only sampling 50 chickens per farm on estimates of the underlying farm level prevalence of various binary welfare indicators.

We can apply similar logic to the issue of accuracy with estimates obtained from the NSS for small samples and here we will consider the two possible threshold sizes, 10 and 23. When dealing with a variable described as a percentage, the biggest uncertainty occurs when there is a 50 per cent yes (and 50 per cent no) response rate. In this case our estimate is 50 per cent but one common method (as illustrated in Main et al. 2013) is to assume an underlying binomial response distribution for the student responses. Then one can use a Normal approximation and would expect a 95 per cent confidence interval of this estimate to be 50 per cent +/- 100 per cent*(1.96 * sqrt(0.5*0. 5/n)). For a headcount of 10 this gives an interval of (19.0 per cent, 81.0 per cent) whilst for a headcount of 23 this gives an interval of (29.6 per cent, 70.4 per cent). As can be seen even for a headcount of 23 these confidence intervals are wide (with width 40.8 per cent).

The above confidence intervals are based on the assumption of an ‘infinite population’ and in fact if we are prepared to assume a ‘finite population’ then this can greatly reduce the width of the confidence intervals through a finite population correction. Here to explain what is going on consider for example a course where only 10 people enrolled, all filled in their NSS questionnaires and eight of them thought it was well taught. Then our estimate would be 80 per cent with no uncertainty whatsoever as we have data on all the students on the course. This would (disregarding the possibility of wrong inputting of answers and potential variability within students, i.e. they might change their answers over time etc.) be acceptable if we only wished to describe what students thought that year and not extrapolate our results. Generally however a potential student looking at the site is interested in whether he/she will think the course is well taught or whether the course is taught better or worse than another course and here this introduces a ‘super population’ of potential students who might take the courses in question of which our sample is those students who actually

took the course and filled in the NSS form. This suggests therefore that it is better in such cases not use a finite population correction.

To summarise, the uncertainty in the statistics given in Unistats will greatly increase with a reduction in the threshold. For example we can be 95 per cent confident that for a course with a 50 per cent yes rate and a sample of 10 students that the real yes rate is between 19 per cent and 81 per cent whilst for a course with a 50 per cent yes rate and a sample of 23 students then we would be 95 per cent confident that the real yes rate lies between 29.6 per cent and 70.4 per cent.

Issues of bias due to non-response

Statistics constructed from samples of students on courses have the potential to exhibit bias if we do not have a 100 per cent response rate. Such biases will occur if there is a correlation between the propensity to respond to the questionnaire and the true answers to the questionnaire questions. For example if people are less likely to respond to the salary questions on the DLHE salary questionnaire if they earn less money, then the estimate for the average salary derived from the respondents only will be too high.

There are many other reasons why people may not respond, for example there may be different response rates for different genders, cultures, ethnicities and ages of students. There may also be an issue of an ‘apathetic’ middle for, for example, the NSS survey where those students who are very happy or have reason to complain are more likely to respond than the rest.

In terms of the headcount, this shouldn’t directly affect the bias unless certain courses with lower headcount also have unusual student make up. However the effect of a biased sample may be more telling in a small sample situation.

To summarise, the response rate is more important when considering bias in estimates than headcount however the effect of a biased sample may be more noticeable if the headcount is small.

Other suggestions for improvements to the data presentation

Firstly I would say that overall the Unistats site has a very nice interface and is generally very clear and easy to use. I think it could benefit a little more from transparency in a few ways:

(i) When the data displayed are actually an aggregate it would be better to give precise detail of the aggregate in question – the hover over text is nice but if it could be specific rather than general that would be great.

(ii) It is great that the sample size is given when no aggregation has occurred but it might be good to include also the response rate, or if simpler something like ‘based on 30 out of a possible 40 (75 per cent) students’.

(iii) When aggregating it would also be good to have sample size in a similar format for the aggregated level data that are displayed.

(iv) When comparing institutions it might be nice to have the option to aggregate the data to the same aggregate level for each institution in the comparison.
(v) It was unclear whether the 50 per cent response rate criterion was also used on the aggregated data and this might be clarified.
(vi) In line with the typical ranges given for continuous measures it might be good to give either a 95 per cent confidence interval or some other form of uncertainty quantification for the binary variables. Perhaps even pairwise comparisons could be performed to identify statistically significantly different records.
(vii) It was not clear that currently any aggregation across cohorts was occurring on the web-site so some clarification here would be good and if this is being used then some acknowledgement in the hover over text would be useful for data thus aggregated as to what aggregation has occurred.

Summary

My brief here has been to describe some of the issues that a change of headcount threshold might raise in the data produced on the Unistats web-site. As has been described above reducing the threshold will increase the number of course by institution combinations that can be displayed at all and can be displayed without aggregation. This has to be balanced with the added uncertainty inherent in the data produced and the increased risk of data disclosure and possible effects of bias. Some of these issues have been explained through example in the text above in such a way that a policy maker might weigh the positives and negatives and decide between the two suggested headcount thresholds.

References

Main, D.C.J., Mullan, S., Atkinson, C., Bond, A., Cooper, M., Fraser, A., and Browne, W.J. (2012) ‘Welfare outcome assessments in laying hen farm assurance schemes’. Animal Welfare. 21:389-396.

Annex E: Focus group participants

Participant	Organisation
Stephen Batchelor	Midkent College
Katherine Bevan	Queen Mary University of London
Matthew Bollington	Department for Business, Innovation and Skills
Tom Corfield	Student Room
Nick Davy	Association of Colleges
Dee Easter	GuildHE
Keith Herrmann	Careers Stakeholder Alliance
Ed Hughes (Chair)	HEFC E
David Morris	National Union of Students
Melanie Siggs	IFS University College
Andrew Walker	Rose Bruford College
Jonathan Waller	Higher Education Statistics Agency
Jennie Walmsley	University of the West of England
Alistair Wilson	South Devon College

Apologies
Greg Wade Universities UK

Funding council attendees
Gordon Anderson (SFC Senior Policy and Analysis officer)
Gemma Bowers (HEFCE Analyst)
Liz Heal (HEFCW Statistical analyst)
Marie-Helene Nienaltowski (HEFCE Policy adviser)
Catherine Nixon (HEFCE Policy adviser)
Richard Puttock (HEFCE Head of data and management information)

Annex G: List of abbreviations

DLHE	Destinations of Leavers from Higher Education
FEI	Further education institutions
FPE	Full-person equivalent
FT	Full-time
HC	Headcount (also used in this document to refer to FPE)
HEFCE	Higher Education Funding Council for England
HEFCW	Higher Education Funding Council for Wales
HEI	Higher education institutions
HEPISG	Higher Education Public Information Steering Group
JACS	Joint Academic Coding System
KIS	Key Information Set
NSS	National Student Survey
PT	Part-time
RR	Response rate
SFC	Scottish Funding Council