

Progress to date on the review of the benchmarking approach

UKPITG 16/06†

Issue

1. In response to the 2013 fundamental review of the UK performance indicators (UKPIs) for higher education (HE), a review of the benchmarking approach for UKPIs was agreed. Timely progress with the review needs to be made in order to effect change in the 2016 publications of UKPIs and experimental statistics, as requested by the UK Performance Indicators Steering Group (UKPISG). UKPITG need to consider progress made since the February 2016 meeting of the UK Performance Indicators Technical Group.

Recommendations

2. It is recommended that UKPITG undertake an assessment of the fit of the existing benchmarking factors with the principles proposed for the selection of factors to include in benchmarking calculations.
3. It is recommended that UKPITG consider the good practice identified through the desk based review of benchmarking in UK public sector, and identify any implications for the UKPIs' benchmarking approach.
4. It is recommended that UKPITG consider the proposed set of principles for defining groupings used within benchmarking factors.
5. It is recommended that UKPITG consider the updates provided, and identify any requirements of them that may arise from the next steps in each of the four work strands.

Discussion

6. The UKPIs provide information on the performance of the HE sector in the UK. The benchmarks included in the UKPIs mean that the measures can be used for comparative and contextualising purposes.
7. In response to the 2013 fundamental review of UKPIs, a review of the benchmarking approach for UKPIs was agreed. Four strands were established, focussing on:
 - (a) Independent assessment of the statistical approach to the benchmarking
 - (b) Review of the principles employed in the selection of factors to include in benchmarking calculations
 - (c) Review of the methodology used to define groupings used within benchmarking factors
 - (d) User consultation with regards to the benchmarking approach used within UKPIs

Update on (a) Independent assessment of the statistical approach to the benchmarking

8. Having confirmed their availability to undertake the assessment of the statistical approach to the benchmarking, the ONS methodology unit provided a further update at

the beginning of May 2016. It was noted at the time that the ONS' work with BIS in relation to reviewing the data sources being used in the Teaching Excellence Framework (TEF) had developed such that they were planning a short piece of further work for that report which would include an initial look at the benchmarking methodology. ONS therefore considered that it would be best to hold off from any work for UKPITG's benchmarking review until the TEF-related report was complete. This was anticipated to be completed by the end of May.

9. The ONS' intention was that an assessment could then be made as to the sufficiency of work to date for UKPITG's purposes, and UKPITG's requirement for any additional stages of work. HESA are awaiting a further update from the ONS.

Update on (b) Review of the principles employed in the selection of factors to include in benchmarking calculations

10. In December 2015 UKPITG members were invited seek advice and guidance from colleagues in their organisations regarding a proposed set of principles for the selection of factors to include in benchmarking calculations (included as Appendix 1). The secretariat have received no feedback on any omissions or refinements deemed necessary, but the planned assessment of the fit of the existing benchmarking factors with the principles has yet to be undertaken.

Recommendation

That UKPITG undertake an assessment of the fit of the existing benchmarking factors with the principles proposed for the selection of factors to include in benchmarking calculations, in preparation for the forthcoming UKPISG meeting.

11. Work on this strand undertaken by HEFCE since the February 2016 meeting of UKPITG has focussed on the desk-based literature review of benchmarking in UK public sector. It has been considered that this would be the most efficient way to find out whether HE sector practice on benchmarking is mirrored by other public sector practice. The full report is available at Appendix 2, and some of the good practice identified is highlighted below.

Taxonomy

12. The review found a breadth of definitions and practices of benchmarking. The most frequently assumed definition was that of simple comparison, whether between providers, or over time. Provider comparison to a sector average was also found to have a variety of definitions. This illustrated the importance of a clear definition of practice.

Evaluative practice

13. There was a lack of evaluation on benchmarking practices within the UK public sector. This made it challenging to summarise good benchmarking practice. A substantial proportion of all documents reviewed focused on publication of information, but not reviewing whether this was acted on and whether that had been effective. This highlighted the need for review of whether data collected was useful and fit for purpose.

Purpose of activity and accountability

14. The purpose of benchmarking should guide the method. Data definitions, responsibilities and accountability should be clear as a part of the benchmarking process. This also involves outlining actions and decision thresholds for various levels of performance. The easiest focus is on what to do in case of 'failure', or what actions and by whom should be taken if a provider falls below a level of acceptable performance.

Reflective and flexible practice

15. The purpose of the method should also reflect stakeholder needs. If the information published is too complex for the intended audience then the process defeats the point. A periodic review of the purpose, method, data and audience was suggested as good practice on the basis that context, technology and populations change.

Recommendation

It is recommended that UKPITG consider the good practice identified through the desk based review of benchmarking in UK public sector, and identify any implications for the UKPIs' benchmarking approach.

Update on (c) Review of the methodology used to define groupings used within benchmarking factors

16. HEFCW took the lead in this work area. A review of existing documentation available on the HESA and HEFCE website, and held in paper form at HEFCE, was conducted. It should be taken into consideration that paperwork covering the entire history of the UKPIs was not available, particularly paperwork relating to the model which was subsequently adopted, an adaptation of work developed by Dr. N. G. McCrum of the University of Oxford, and papers between 2009 and 2010.

17. The full report is available at Appendix 3, and the emerging set of principles proposed for defining groupings used within benchmarking factors are provided below. Throughout the discussion it is assumed that the reader is familiar with the method for calculating the HESA UKPI benchmarks, and the terms factor, group and category within that method. Reference to deciding the overall method, and how to select the factors is avoided as this forms part of work areas (a) and (b).

18. The desk review of existing documentation from HEFCE, HESA, UKPITG and UKPISG suggests the following set of principles.

The groupings used within benchmarking factors should:

- 1) have as few categories as possible;
- 2) be homogeneous with respect to the indicator to which they refer;
- 3) make practical sense;
- 4) be relatively evenly spread;
- 5) be used consistently across all UKPIs where possible, allowing other UKPIs to act as context statistics;
- 6) be determined empirically;
- 7) be reviewed periodically to ensure they continue to be homogeneous, make practical sense and are relatively evenly spread.

Recommendation

It is recommended that UKPITG consider the proposed set of principles for defining groupings used within benchmarking factors.

Update on (d) User consultation with regards to the benchmarking approach used within UKPIs

19. The secretariat had not received an update on this strand of activity at the time of writing. Members will be invited to provide an oral update during the meeting.

Recommendation

It is recommended that UKPITG consider the updates provided, and identify any requirements of them that may arise from the next steps in each of the four work strands.

Further information

20. For further information contact Alison Brunt (Phone: 0117 931 7166; e-mail: a.brunt@hefce.ac.uk) or Mark Gittoes (Phone: 0117 931 7052; email: m.gittoes@hefce.ac.uk).

Appendix 1 – Proposed guiding principles for UKPI benchmarking factors

1. The selection of benchmarking factors should adhere to the same principles as the indicators themselves.
2. In addition, benchmarking factors should:
 - a. Be correlated with what is being measured.
 - b. Be material, in that they should vary significantly from one institution to another. Materiality should be considered in the context of an ambition to minimise the risk of an institution having a significant impact on their own benchmark.
 - c. Be outside of the institutions' control, or otherwise undesirable for them to control for.
 - d. Not be uniformly distributed across institutions, rather the factor should differentially affect institutions' benchmarks.
 - e. Be a direct measure rather than a proxy.

Appendix 3 – Review of documentation relating to the definition of groupings within benchmarking factors

1. HEFCW undertook a review of existing documentation available on the HESA and HEFCE website, and held in paper form at HEFCE. It should be taken into consideration that paperwork covering the entire history of the UKPIs was not available, particularly paperwork relating to the model which was subsequently adopted, an adaptation of work developed by Dr. N. G. McCrum of the University of Oxford, and papers between 2009 and 2010.
2. Throughout the following discussion it is assumed that the reader is familiar with the method for calculating the HESA UKPI benchmarks, and the terms factor, group and category within that method. Reference to deciding the overall method, and how to select the factors is avoided as this forms part of work areas (a) and (b).

There should be as few categories as possible

3. The factor groups in the arithmetic method used to calculate the UKPI benchmarks are used to explain some of the variation between institutions, if there are not enough factor groups, or there are too many factor groups this could affect the robustness or usefulness of the method.
4. If there are not enough groups then less of the variation is explained, however having too many groups could result in a model with sparse data, many blank cells and many cells with small values, adding volatility to the model. A single group would simply provide the sector average, while the maximum number of groups, i.e. one cell per student, would provide an institution's own value of the UKPI.

'...there should be as few categories as possible.' PISG 03/08

'At the extremes a single group would not take account of any variation between institutions, while a system that put each student into a separate group would account for all the variation.' HEFCE 2006/34

'For example, if the number of benchmarking groups was increased to the extreme, there would be the same number of benchmarking categorisations as individual students, with each categorisation containing one student. A feature of the benchmarking approach is that where differences exist between the indicator and the benchmark, this may be due to the institution's performance, or due to some other factor which is not included in the benchmark. If each benchmarking categorisation contained only one student, the benchmark would then be the same as the indicator in all cases: all variation would have been explained and the notion that differences may be due to the institution's performance would be lost.' PISG12/01

'...the increase in small numbers in benchmarking categorisations demonstrates an increased potential for the information provided by the benchmarks to be reduced unnecessarily. A second effect is that the proportion of an institution's benchmark that will be determined by itself is likely to increase.' PISG12/01

'Larger numbers of benchmarking groups leads to groups being smaller in size and increased potential for volatility within the PI benchmarks as a result.' PISG12/01

Groups should be homogeneous with respect to the indicator to which they refer

Groups should make practical sense

Groups should be relatively evenly spread

'...groupings have been chosen so that as far as possible the students within each group are relatively homogeneous.' HEFCE99/66

'...each category should be homogeneous with respect to the indicator(s) to which they refer.' PISG03/08

'... they were combined provided that that also made practical sense (so A-level points categories, for example, were only combined if they were contiguous.)' PISG 03/08

'...it was also necessary to define relevant groupings for each factor. In general this was done pragmatically by considering the spread of values of the indicator for different factor values, and grouping together those with similar values. The aim was to produce as small a set of groups as possible, where each group was homogeneous with respect to the characteristic being measured.

'In some cases, the original factor groupings combined areas which at the time appeared to show similar characteristics. Some such areas have now been split because they later became more diverse, or because there was a perception that the two areas should be kept separate. For example, initially the subjects of mathematics and computing were combined into one category, but by the time the new JACS codes were introduced it seemed more sensible to keep them separate.' HEFCE 2006/34

5. Paper PISG 09/07 particularly demonstrates where choosing groupings should make practical sense and be more evenly spread. The paper examined the possibility of extending the coverage of the retention UKPI to include part-time first-degree entrants. Analysis of factor groups showed that using the existing groups for full-time entrants was not appropriate for part-time entrants.

6. The majority (94%) of part-time students fell into the mature age group as defined for the full-time entrants UKPI (young and mature are under 21 or 21 and over for full-time). Analysis of age of part-time students showed that age was mostly evenly spread, and in terms of the indicator 'Absent from HE' there was no consistency across any potential age groups. The age group proposed (30 and under, over 30) was decided upon as it mirrored the HEIPR, and the data were more evenly spread.

7. Similarly with the entry qualification groupings, almost half of part-time entrants were found to hold 'HE qualifications' on entry so the spread across the data was not very even. It was proposed that the HE qualifications group was split, and for some of the existing full-time groups with small numbers of part-time entrants it was proposed that they be aggregated.

8. It was also proposed that some full-time subject groups with small numbers of part-time entrants be aggregated to form an 'Other subjects' group.

9. It could be argued that in this case although the data were evenly spread with respect to age, there was no homogeneity with respect to the indicator 'Absent from HE' and therefore including age as a factor does not add any explanation of the variation amongst institutions. The analysis of spread for subject and qualifications on entry in this paper, did not appear to consider the value of the indicator, and therefore it is not possible to say if the groupings chosen exhibited homogeneity.

10. Paper UKPITG 14/04 discussed bringing the qualification on entry groups for indicator E1 into line with the other UKPIs. Table E1 uses more benchmark factors than the other UKPIs (ethnicity and gender) and therefore has fewer qualifications on entry groups in order to not significantly increase the number of categories. The paper discusses merging groups defined by specific A level grade combinations and groups defined with similar tariff scores but achieved from a wider range of qualifications. Merging this data would have resulted in a much bigger top entry qualification group than used for the other PIs and compared to the other groups to be used with E1, and would also have resulted in treating students with the same tariff score but different A level grade combinations differently. The proposal which was agreed upon created groups for A level grade combinations separately from groups with similar tariff scores achieved from a wider range of qualifications. The number of groupings decided upon evened up the spread of the data across the groups, whilst retaining the same number of categories.

The same groups should be used across all UKPIs where possible allowing other UKPIs to be context statistics

11. Both supplementary data and context statistics are provided to facilitate the understanding of the UKPIs. Context statistics currently provided are:

- a. the average number of HE providers in the adjusted sector benchmark comparison
- b. the average proportion which the HE provider's own students contribute to the benchmark

12. However, the UKPIs themselves can be context statistics to each other.

'For both access and student progression indicators, it is proposed that adjusted sector indicators are derived for comparison. If possible, the intention is to use the same grouping of subjects and entry qualifications to calculate adjusted sector averages for the access and student progression measures. This is partly to make it easier to appreciate what is being done, and partly to facilitate the use of access indicators as a context for the progression indicators, and vice-versa.' HEFCE 99/11

13. This criteria should be applied after testing for the first four criteria, as it wouldn't make sense to use groupings for another UKPI should they not prove to be homogeneous, make practical sense, or be relatively evenly spread, as the amount of variation to be explained in the model would not then be being explained appropriately.

Groups need to be determined empirically

'...the categories must be determined empirically' PISG03/08

14. There does not appear to be any other way to choose the groups. It has already been seen that there should be as few categories as possible, they should be

homogeneous with respect to the indicator to which they refer, they should make practical sense and that they should be relatively evenly spread.

Groups should be reviewed periodically to ensure they continue to be homogeneous, make practical sense and are relatively evenly spread

15. Factor groups have been reviewed each time a change is made to the underlying data in order to determine that they still meet the above criteria. Generally this has mostly been due to changes to qualifications on entry data and subject of study data. When changes have been introduced, groups have generally been reviewed and amended to make practical sense and to ensure a relatively even spread of the data. Not so often is the homogeneity of the data examined, i.e. the values of the indicator calculated and examined to see if they are similar or different for each group, or sub groups within groups.

16. Examples where groups have been reviewed for homogeneity and changes have been recommended can be found in PISG 05/09 which reviewed the subject of study groups, and PISG 09/02 and PITG 13/05 which examined qualifications on entry groups.

17. In PISG 05/09, the access and the non-continuation indicators for young entrants to full-time degrees were calculated for JACS Subject Groups, and JACS Principal Subject groups within each JACS Subject Group. Most subjects groups were fairly homogeneous, but the paper went on to identify three subject groups that seemed more disparate and contained large enough numbers to allow a split to be viable.

18. Due to the introduction of additional qualifications on entry data in the 2007/08 HESA student record, paper PISG 09/02 considered four possible approaches to grouping the qualifications on entry for UKPIs based on entrants. For each country of the UK, the size of the groups and the indicator (state school, SEC 4-7, LPN) for each group was analysed. For the state school indicator only, the mean institutional indicator and benchmark for each country of the UK were also examined. Practical sense and number of groups was considered in developing the four approaches, whilst homogeneity and spread was considered in the analysis.

19. Paper PITG 13/05, looked at updating the qualification on entry groupings for the PIs not based on entrants. Analysis of the effect on the benchmarks for T7 was undertaken by running the UKPI with the old and the new qualification on entry groups. For each group, new and old, the number of students and the indicator were calculated.

20. When a change to groups is required this is known in advance, but data are not always available when the groups are agreed. When the data are available then homogeneity should be examined alongside making practical sense and a good spread of the data. When data are not available in advance then the groupings should be reviewed for homogeneity when it becomes available.

21. Groups should be reviewed regularly for homogeneity, practical sense and spread, as changes due to external factors and the way things are perceived or coded can change over time.

Other considerations

22. Factor groups may be changed due to changes to the data being collected. Historically this has mostly related to qualifications on entry and subject of study. Changes included moving from using A level point score, to UCAS tariff and subsequently to using QVALENT2, QVALENT3 and QVALETYPE and QVALEGRADE for qualifications on entry, and from HESACODE to JACS for subject of study.

23. The HEDIIP subject coding project means that it is likely that there will be a future change from JACS to HECOS, which may impact on subject of study groupings.

24. Changes in policy may impact on deciding factor groups directly in terms of what makes practical sense, or indirectly in terms of what data is collected, but the choice of factor groups still has to be determined by meeting the criteria of as few groups as possible, homogeneity, practical sense and spread.

25. Changes due to institutional factors are part of the institution's performance and are therefore part of what is being measured by the UKPI. The factors and their groups are used to remove variation assumed to be outside of the institution's control.

'Factors should not be in the HE providers' control, and so not be part of their performance (www.hesa.ac.uk/pis/benchmarks)

26. It has not proved possible to find information on similar benchmarking methods via an internet search, and therefore not possible to make any observations on how factors are grouped elsewhere.