
Chair of the UK Statistics Authority, Sir Michael Scholar KCB

David Willetts MP
House of Commons
LONDON
SW1A 0AA

20 May 2009

Dear David

HIGHER EDUCATION INITIAL PARTICIPATION RATES

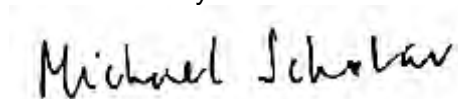
Thank you for your letter dated 2 April 2009. Please accept my apologies for the length of time it has taken to provide a substantive reply to you. I understand that my officials were in touch with your office about this.

Your letter raised a number of important questions and I wanted to ensure these were properly worked through by the National Statistician and the Department for Innovation, Universities and Skills (DIUS), so as to establish what progress we might make in this area. The attached note provides a summary of advice to me from the National Statistician drawing on the department's own advice to her.

I hope that you find the information helpful.

Best wishes – I am sorry that I will miss you when you visit St John's next month.

Yours sincerely



Sir Michael Scholar KCB

ANNEX: HIGHER EDUCATION INITIAL PARTICIPATION RATE

The Higher Education Initial Participation Rate (HEIPR) is the National Statistic that, as you note in your letter, the Government uses to measure progress of the Comprehensive Spending Review 2007 target to "increase participation in Higher Education towards 50 per cent of those aged 18 to 30 with growth of at least a percentage point every two years to the academic year 2010-11". For this purpose, participation is defined as a minimum of six months' study. The HEIPR approximately equates to the probability that a 17-year old will participate in higher education (HE) by the age of 30 given the age-specific participation rates.

The HEIPR essentially counts English-domiciled 17 to 30 year-old HE students. Students are counted if they participate for at least six months on a course expected to last six months or more. Students should be excluded if they have participated in HE previously for at least six months. A key element in the estimation process is, therefore, ensuring that there is no over-counting of students re-entering HE who have had a break in their education.

When the HEIPR time series was introduced in April 2004 (covering 1999/00 to 2002/03), it was based on the self-reported information on initial participants in HE taken from both the Higher Education Statistics Agency (HESA) Student Record and the Learning and Skill Council (LSC) Individualised Learner Record. The methodology underwent major changes in 2007 to reflect the recommendations of both an earlier National Statistics Quality Review and a review undertaken by the National Audit Office in 2006 regarding the data systems underpinning target measurement in a number of government departments.

These changes included *inter alia* improvements to the matching techniques between the HESA and LSC records to identify students with six months prior HE experience. They also widened the geographical coverage to include Scotland and Wales. The introduction of the changes took place in 2007 and was pre-announced. They were explained in advance and revised estimates were provided back to 1990/2000.

Notwithstanding these changes, a further change took place, also in 2007, when HESA introduced a revised variable on their Student Record to identify students who had prior HE experience lasting six months or more. This was intended to be the final improvement to the HEIPR methodology and had been identified in the NAO's earlier report. Unfortunately, serious problems were discovered with this variable and at a late stage in preparation for release of the annual figures, DIUS were required to adopt a new methodology. Rather than delaying release of the latest HEIPR figures, DIUS worked to the existing timetable using the new methodology. This did not allow the department sufficient time to consult users. DIUS are satisfied that the new methodology is robust but it has led to a discontinuity in the time series.

Your recent parliamentary question asked the Secretary of State to recalculate the HEIPR for each year between 1999-2000 and 2005-06 using the new methodology used for 2007-08 that were published on 31 March 2009. The answer provided (HC Deb 28.04.2009, cc1268-9W) noted that additional estimates using the new methodology, for years back to 2003-04, alongside information on how to interpret them and compare them to the existing published trend, have been added as an appendix to the Statistical First Release dated 15 May 2009 and also published separately at <http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000839/index.shtml>

DIUS advises that it has not been possible to recalculate the HEIPR for years between 1999-2000 and 2002-03 using the new methodology as the margin for error would be too significant.

Your second parliamentary question asked what discussions the Department had with the UK Statistics Authority regarding this matter. The answer confirmed that DIUS officials liaised with statisticians at the Higher Education Funding Council for England (HEFCE) prior to the release of data using the new methodology, and more generally regarding the robustness of the old and new methodologies. The National Statistician, as the Authority's principal adviser on professional statistical matters, was kept fully informed at all times regarding the planned changes to the methodology discussed above.

We are keen to provide as clear an answer as we can to each of your specific questions, particularly where they have not been addressed above.

1. Is it acceptable for the HEIPR dataset to be subject to large methodological changes of the sort introduced this March without any prior consultation or even warning?

The Statistical First Release (SFR) is timed to allow a minimum of time between production and release of the HEIPR, to minimise the risk of a breach of the rules regarding pre-release access to official statistics. When DIUS discovered that the HEIPR was affected by a data quality issue, there was little time available to announce the need for a change. The options considered were as follows:

1. In the week before publication, to make arrangements to announce a change to the methodology, and release the data based on a more robust methodology.
2. Cancel the SFR altogether, again during the week before publication, announce the planned changes, and publish the data at a later date;
3. Not release a 2007/08 figure.

It was considered that all three options might attract criticism of some sort. Given that a robust alternative methodology was possible in the time available, the first option was chosen. The decision was made to publish to the announced timescale using the new methodology with clear descriptions, and include figures for 2006/07 based on both the old and new methodologies, to be transparent about the impact of the change.

DIUS analysts obtained advice from HEFCE on the robustness of the old and new methodologies, given the 2007/08 data problems and that HEFCE's agreement with the approach taken was clear.

The Statistics Authority's view is that methodological changes of this sort should be pre-announced as early as possible, together with accompanying explanation of the likely impact and analytical consequences that may flow from such changes.

2. Have there been sufficient grounds for the frequent and major methodological changes to this dataset that have occurred since it first appeared in 2004?

The background to the methodological changes are set out above.

3. Should the HEIPR now be recalculated for historical years using the new methodology so that year-on-year comparisons can still be made?

Provisional HEIPR data based on the new methodology for years back to 2003/04 were published on 15 May 2009 as an appendix to the Statistical First Release published on 31 March. DIUS advises that it has not been possible to recalculate the HEIPR for years between 1999-2000 and 2002-03 using the new methodology as the margin for error would be too significant.

4. Should more detailed information be placed in the public domain about the most recent set of methodological changes?

Additional information on how to interpret provisional HEIPR data based on the new methodology for years back to 2003/04, and how to compare them to the existing published trend, was published on 15 May 2009 as an appendix to the Statistical First Release published on 31 March.

5. How robust is the final data for 2006/07, as calculated under the old methodology, given that it is identical to the provisional data for that year and when, in every previous year, there has been a fall between the provisional and the final data?

The provisional HEIPR is revised to produce a final figure using updates to the Individualised Learner Record (Learning and Skills Council) and population estimates (calculated by the Office for National Statistics and converted to an academic year basis by DCSF, updated on a regular basis).

It is not always the case that finalised data are lower than provisional data in respect of the HEIPR. However, DIUS advises that recently this has been the case principally because population estimates have been revised significantly upwards on several occasions. This did not occur in 2006/07, and therefore the provisional and final HEIPR figures are similar to the accuracy quoted in the SFR.