

Pilot exercise to provide statistics on the nature of young entrants to HE for local areas (June 2007)

Summary

1. This document and associated downloads form a pilot exercise to provide statistics on the nature (across characteristics such as social background and subject choice) of entrants to higher education at a local level. We are seeking views on whether statistics like this would be a useful extension to the local participation information presented through POLAR (which is due to be updated this year). Feedback is requested by 21 September 2007.

Introduction

2. The publication 'Young participation in higher education' (HEFCE 2005/03) reported that young higher education entrants from high and low participation areas differed in nature across a range of statistics. Users of POLAR (www.hefce.ac.uk/polar) data have commented that knowing more about the nature of entrants from different geographical areas would be of interest, and could help inform widening participation activities. We aim to provide such expanded information about the nature of entrants by local areas following the update of the local participation information that we present through POLAR (expected late 2007). To help in our plans with this, and to give some limited information to users now, we are providing statistics on the nature of entrants to higher education at a local (ward) level based on existing data sets as a pilot exercise.

3. Table 2 of HEFCE 2005/03 (on page 120) shows the proportion of entrants with a particular characteristic for each of the ward participation quintiles used for analysis in that report. For example, that analysis shows that 34 per cent of entrants from the lowest participating quintile of wards lived with their parents during term time, compared to 15 per cent of entrants from the highest participating quintile of wards. This pilot presents selected statistics from the data set underpinning this table for individual wards themselves rather than the quintile groupings to reveal the local distribution of these statistics.

4. The main purposes of this pilot exercise are:

- a. To investigate if this type of information is of interest to users.
- b. If information of this type is of interest, then what data items describing entrants would be useful (including possible items not provided in this pilot).
- c. Identify areas with unexpected values on the statistics that might suggest errors or some other difficulty with the proposed approach (for example, some wards containing public schools have extremely high proportions of male entrants).

5. Comments on this document and its associated data should be sent, by **21 September 2007**, to Mark Corver by e-mail (m.corver@hefce.ac.uk) or post (M. Corver, HEFCE, Northavon House, Coldharbour Lane, BRISTOL, BS16 1QD).

Data used for the pilot

6. The data provided in this pilot is based on that used to create Tables 2 and 3 of HEFCE 2005/03 (that is, YPR(H) English domiciled young entrants) with some changes.

The most important are:

- The time period encompassed is expanded from just the 2000 cohort to the 1994 to 2000 cohorts combined.
- The distance calculations for 'HEI within 30 minutes of home' statistic have been revised with improved information on the location of higher education provision.
- The tuition fee payment statistic has been dropped since the interpretation of this is not constant across the 1994 to 2000 cohorts.
- There is one statistic included that was not used in HEFCE 2005/03 - the proportion of entrants who are male.

7. 1991 Census wards are used to provide the geography to match our previous information on local participation rates. If, as a result of this pilot exercise, we decide to produce any post-pilot local entrant nature data sets, we will use 2001 Census based geography to match that planned to be employed by the updated POLAR. Some data items are suppressed (indicated by a '.'). Data items are suppressed in cases where there are too few entrants (leading to a higher risk of compromising confidentiality) or concerns about how representative the statistic is (the proportion of entrants with known characteristics being too small).

8. The data files for this pilot exercise contain:

- a series of geographical variables from region down to the 1991 Census ward codes that the statistics are reported for (this matches the existing POLAR data files)
- information on the participation band of the ward (from POLAR)
- approximate size of the entrant base used for the calculations
- a set of statistics about the nature of the entrants.

The statistics about the nature of the entrants are all reported as the number of entrants with a specified characteristic (living with parents whilst studying, for example) expressed as a percentage of the number of entrants where it is known whether they have the characteristic or not. For some statistics (such as those relating to subject studied) almost all entrants would fall into the known category. For other statistics, like social class, the number of entrants with known characteristics can be lower. If it falls below a threshold the statistic is suppressed because of the higher risk that it is unrepresentative.

9. The local statistics are intended to be used in conjunction with the discussion in HEFCE 2005/03 (pages 119 to 123). Detailed definitions of the statistics are not given in this pilot, although some notes on selected statistics are provided in Table 1.

Table 1: Local entrant nature statistics provided in pilot

| Statistic heading | Notes |
|--|---|
| POLAR group indicator | Numeric flag for the POLAR participation quintile that the ward is assigned to (1=lowest, 5=highest). These are the POLAR quintiles - not the 2000 cohort YPR(H) quintiles used in Table 2 in HEFCE 2005/03 |
| Young participation POLAR ward bands | POLAR participation quintile range |
| Combined entrant count | Indication of the size of the total entrant base from the seven combined cohorts (the actual base for each statistic will usually be less since only entrants with known properties are used) |
| Social Class IIIM, IV, V (UCAS entrants) | Performance Indicator definitions (www.hefce.ac.uk/learning/perfind/default.asp) |
| 16-18 Institution is an Independent School | Performance Indicator definitions (www.hefce.ac.uk/learning/perfind/default.asp) |
| Entry at 19 through a deferred UCAS application | Entry at 19 through a deferred UCAS acceptance gained in the previous cycle (for example, accepted in the 2000 entry UCAS cycle for actual entry aged 19 in 2001). Consistent with a planned 'gap year' |
| Entry at 19 not through a deferred UCAS application | All other means of entry at age 19 |
| Entry through the UCAS clearing process | |
| Best entry qualification not A-level | |
| A-levels 1-17 points or non A-level entry qualifications | This group of qualifications are broadly equivalent to those that were recognised as being at increased risk of non-completion for the purposes of the improving retention funding method (HEFCE 2005/03, page 121) |
| Entry to medicine, languages, humanities or agriculture | Subjects associated with higher participation backgrounds |
| Entry to education, mathematics, computing and nursing | Subjects associated with lower participation backgrounds |
| Studying at HND/HNC level | |
| HEI within 30 minutes of home (revised) | Proportion of entrants who attend a HEI estimated to be within 30 minutes driving time of where they lived prior to entry. Takes account of campus location where possible |
| Lives with parents during term time | Based on the HESA student record term time accommodation field |
| Studying at UK HEI outside England | |
| Studying at HEI formerly UFC funded | |
| Male entrant | Proportion of entrants that are male. Around 51 per cent of the young population is male. |

Using the pilot data

10. As noted in HEFCE 2005/03, HE entrants from low participation areas are, by definition, likely to be atypical of the area they come from. Analysis in HEFCE 2005/03 (Annex F) suggests that only around 5 per cent of children in low participation wards live in more advantaged, higher participation rate, micro-areas within the ward. However,

children from these atypical micro-areas would be expected to form a higher proportion of entrants from these low participation wards (because of their higher participation rates). So whilst attending an independent school might be rare among children living in a poor ward, it could be much less rare among the HE entrants from that poor ward. Also, the number of HE entrants from higher participating wards is so large that they usually form the largest group of entrants with any characteristic, even if that characteristic is more likely to be present among entrants from a low participation background (HEFCE 2005/03, page 122-123).

11. The total number of young entrants from a ward is often small, especially so for low participation wards. This has two consequences. The first is that situations where there is a disclosure risk are more likely. Measures to reduce this risk, either data suppression or data modification, will reduce the accuracy of the information available for these areas.

12. The second consequence is that, even when the entrant base is large enough to avoid the suppression thresholds, the number of entrants is often so small that the reported proportions would be expected to be subject to a substantial component of random variation. So, for example, if we measured the same statistics for different cohorts we would be likely to get different values, even if the underlying propensity for entrants from a particular ward to have the measured characteristic had not changed. Similarly, different reported values for wards could arise from these random fluctuations as well as from any differing underlying propensity for entrants to have a particular characteristic. The approximate size of the combined cohort entrant base used for the statistics is provided for each ward to help users assess the likely impact for their analysis of these random fluctuations.

Disclosure protection

13. We have taken measures, with reference to Office for National Statistics guidelines, to minimise the risk of data disclosure (the revealing of information about potentially identifiable individuals) from these local statistics. These measures include the combining of seven cohorts, data modification to mask sensitive cells and suppression of data for sensitive cells. In particular, users should be aware that zero values in the presented tables do not confirm that no entrants from that area have that property. Inevitably these necessary measures will act to reduce the accuracy of the reported data, especially where the entrant base is small.