



DLHE Longitudinal 2006/07 Survey – Weighting Process

- 1.1 After an exploration into the similarities and differences between sample A and sample B, the decision was taken that it was appropriate for the samples to be combined and analysed as one. The samples were weighted, firstly to correct for selection bias and then to correct for response bias, to ensure that the weighted survey findings were representative of the early DLHE population.
- 1.2 An additional weight was also developed for use when conducting analysis at individual HEI level.
- 1.3 The remainder of this chapter describes in detail the specific weighting methodology that was used.

Logistic Regression Model for Weighting 'Sample A' and 'Sample B'

- 1.4 The method used to weight the Sample A and Sample B completes followed that used for the 2002/03 and 2004/05 DLHE Longitudinal Surveys. The method was as follows:
 - A preliminary weight to correct for selection criteria was computed by inverting the sampling fraction.
 - Non-response behaviour (i.e. whether they responded to the survey or not) was then modelled using binary logistic regression. The regression model was run on all issued DLHE Longitudinal Survey graduates, weighted by the inverse of the selection weight. A weight was generated using the predicted probabilities saved from the non-response model.
- 1.5 The following variables were used in the model:
 - **Country of Institution** (England, Northern Ireland, Scotland, Wales),
 - **Level of Qualification Obtained** – the 8 way split was used (Higher degree by research, Higher degree by taught course, Postgraduate diploma or certificate, First Degree, Other diploma or certificate, Professional qualification, Other qualification, No formal qualification);
 - **Classification of Degree** (First class honours, Upper second class honours, Lower second class honours, Third class honours / Pass, Unclassified, FE level qualification, Classification not applicable)
 - **Employment categories at Early Survey** (Full-time paid work only (including self-employed), Part-time paid work only, Voluntary/unpaid work only, Work and further study, Further study only, Assumed to be unemployed, Not available for employment, Other)
 - **Subject of Original Course** (Architecture Building and Planning; Biological Sciences; Business and Administrative studies; Creative Arts and Design; Eastern, Asiatic, African, American and Australasian Languages, Literature and related subjects; Education; Engineering; European Languages, Literature and related subjects; Historical and Philosophical studies; Law; Linguistics, Classics and related subjects; Mass Communications and Documentation; Mathematical and Computer Sciences; Medicine and Dentistry; Physical Sciences; Social studies; Subjects Allied to Medicine; Technologies; Veterinary Sciences, Agriculture and related subjects; Multiple subjects)

- **Student Type** (Research PhD and Masters students, Foundation students, Research Council Taught students, HND (England institutions only), TDA funded students, General).
- **Method of Data Collection for the Early Survey** (Standard questionnaire: first mailing, Standard questionnaire: second mailing, Telephone survey: graduate, Telephone survey: third party, Own institutions student record, Other)
- **Contacts available** - mail (yes, no), email (yes, no), and telephone (yes, no),
- **Gender** (female, male)
- **Age** (25 or under; 26 – 30; 31 – 40; 41 -50; 51+)
- **Ethnicity** (Asian; Black; Mixed; Other; White)
- **Number of contact methods**

1.6 Method of Completion was not used as it is a tautological variable.

1.7 The logistic regression model generated the probability of a graduate participating in the survey given their 'type' (based on the predictor variables described above). Not all the variables that were used in the model necessarily made it into the final model.

1.8 After these weights were calculated, the top and bottom 2.5% were trimmed. Finally, a calibration of age bands was required. Apart from age band, all other key variables had very similar distributions to the unweighted population of 332,110.

1.9 The Sample A and B weights were calculated separately, and together provide national level weighting for the all 49,063 interviews completed.

HEI weights

1.10 As well as national weights a set of HEI specific weights were calculated to correct for response bias at HEI level. This was done on the combined completed interviews from Sample 'A' and 'B'. The process for calculating HEI weights was based on that used for the previous DLHE Longitudinal Survey, and differentiated depending on the number of interviews completed for that HEI.

1.11 For HEIs with 400 or more DLHE Longitudinal Survey respondents the survey data are weighted so as to give a close percentage match between the survey and the census in terms of broad subject group, the part-time/full-time split; and the postgraduate/undergraduate split.

1.12 The broad subject groups us DLHE Longitudinal Studyed were: health and welfare; science and agriculture; engineering, manufacture and construction; social science, business, law and combined; humanities and arts; education.

1.13 For HEIs with between 200 and 399 DLHE Longitudinal Survey respondents the survey data were weighted so as to give a close percentage match between the survey and the census in terms of the part-time/full-time split; and the postgraduate/undergraduate split.

1.14 For HEIs with between 100 and 199 DLHE Longitudinal Survey respondents the survey data are weighted so as to give a close percentage match between the survey and the census in terms of the postgraduate/undergraduate split.

- 1.15 For HEIs with fewer than 100 DLHE Longitudinal Survey respondents no HEI level adjustment has been made.
- 1.16 Finally, all HEI weights also include an adjustment so that when they are applied each HEI is scaled in proportion to the number of responses for that HEI in the unweighted data. This means that the weighted base for each HEI is equal to its unweighted sample size.