## Annex A
## Data sources and definitions

### Data sources

HESA student record

1. See HESA records for 1997-98 through to 2001-02. Further details can be found at www.hesa.ac.uk under Data Collection.

DfES schools data

2. For further details, see www.dfes.gov.uk/performancetables/.

UCAS data

3. For further details, see www.ucas.ac.uk.

### Data linking

4. The data are linked in three stages. Stage one involved linking yearly individualised student HESA records to create a longitudinal record of student progression. The linking between years was based on a number of variables recorded for each student including a student instance university identifier (more commonly know as a HIN number), date of birth, gender, postcode and other characteristics.

5. After the longitudinal HESA linking has taken place, the individual students are linked to their UCAS data if they exist. The final part is to link the HESA/UCAS combination record to the information recorded on each school (based on the DfES school performance tables).

### Definitions

Cohort

6. The initial selection was made from records from the 1997-98 HESA student records which were identified as:

a. English domiciled.

b. English schooled.

c. Full-time.

d. A-level taken in 1997-98.

e. At English HEI of entry.

f. Aged 18 on the 31 August 1997.

g. On degree level courses (HESA qualification aim field = (19, 20, 21, 22, 23, 24)).

7. Records with the following characteristics were then removed:

a. Students who have no first degree award in the period but have been given a higher degree award (HESA qualification fields, codes 01-14).

b. Expected course length at start of the student's course not equal to three or four years.

c. Students whose degree subject area at the start of the course is medicine, dentistry or veterinary science (HESA subject codes A, D1), or architecture, building or planning (code K).

d. Duplicate records for the same student. The record with the highest first-degree honour award is favoured when multiple records exist for a single student.

e. Students who are recorded as gaining less than eight A-level points.

f. Students who attended schools which could not be linked to schools on the DfES's performance tables.

Achievement

8. Each student in the cohort is given an achievement status for the period 1997-98 to 2001-02 which is derived for each student in the following way:

a. Identify if the student has been awarded a first degree in the period, and in which years. The student may have multiple degree awards for the period. First degree awards are identified using the HESA qualification fields (codes 19 through to 24).

b. Students with no first degree awards are separated into those who are still active in HE in 2001-02 or were given an HE award below degree level in the period, and those who are not active and were not given an HE award. Students are defined to be active in 2001-02 if they are not recorded as leaving (based on the HESA date left field) and they do not have a dominant mode. Students with codes of (25, 26, 27, 28, 29, 30, 31, 32, 41, 42, 43, 61, 62) on the HESA qualification fields are defined as being awarded an HE sub-degree.

c. For those students with a first degree award, the highest classification of first degree awarded in the period is given. First class honours is assumed to be the maximum achievement, followed by an upper second, lower second, third and finally an award

with no honours attached. The classifications are based on the HESA classification field and are allocated as follows:

| | |
|---|---|
| '01' | First class honours |
| '02','04' | Upper second class honours |
| '03' | Lower second class honours |
| '05' | Third class honours |
| '06'-'11' | Award with no honours |
| Otherwise | Assumed not to be a first degree. |

Note that undivided second class honours (code = '04') are assumed to be upper 2nd class awards.

9. There are five measures of achievement for each student:

a. Whether the student gained a first degree or is still active
Positive outcome: students with first class, upper second, lower second, third class honours awards, awards with no honours, or still active in HE in 2001-02.
Negative outcome: Students who are not active in 2001-02 and do not have a first degree award.

b. Whether the student gained an honours degree by 2001-02
Positive outcome: Students with first class, upper second, lower second and third class degree awards.
Negative outcome: All other students in the cohort.

c. Whether the student gained X or better by 2001-02 (where X is a lower second, upper second or first)
Positive outcome: Students with X or better.
Negative outcome: All other students in the cohort.

10. A simplifying assumption for this analysis is that the student was given their award at the institution they started at in 1997-98. This is not strictly correct but transfer rates are low for these data, at around 2 per cent.

School type

11. School type is derived from the type of institution field given in the DfES's school performance tables:

a. Our definition – Local Education Authority (LEA).
DFES institutions included: County school maintained by the LEA; voluntary aided school maintained by the LEA; and voluntary controlled school maintained by the LEA.

b. Our definition: further education colleges (FEC).

DfES institutions included: Further education sector institution.

c. Our definition: grant-maintained (GMS).
DfES institutions included: Grant-maintained school.

d. Our definition: independent (IND).
DfES institutions included: Independent school.

All other DfES school types are omitted from the analysis.

12. The link between a student's HESA record and the school they attended is achieved using a combination of methods including the use of the last institution attended field on the student's HESA record and/or the student's UCAS number (recorded on both HESA and UCAS individualised data).

School performance

13. The school performance measure given in the report is based upon the average A-level points score per examination entry for each school. It does not depend on the number of examinations taken by individual students.

14. For example, imagine we have a school which only has two students entered for A-levels. The first achieves two grade As and a B, and the second gains one grade C at A-level. The school's average A-level score per examination entry would be calculated as follows:

[10 (First student, grade A) + 10 (First student, grade A) + 8 (First student, grade B) + 6 (Second student, grade C) ] / [ 4 (Number of A-level exam entries)]
= 32 / 4
= 8 (the school would average a grade B for each A-level examination entry).