

## **Methodological Changes to the 2006/07 Destinations of School Leavers Survey**

### **Background**

In previous years, Careers Scotland supplied school leaver destinations data to the Scottish Government aggregated to school level. The purpose of the data collection was to allow national monitoring of destinations and for individual schools to report the results in their school handbooks. Analysing data at this level provided basic information about destinations but did not provide a more useful insight into the factors that influence a school leavers destination.

Over time the needs of the data collection have changed and there is increasing demand for explanations of a given change or for information on the destinations of a particular sub-group of school leavers.

To this end, it was agreed by the Education Information and Analytical Services Division of the Scottish Government (EIAS) and Careers Scotland, that in future the destinations data would be supplied at the individual pupil level. This data would then be matched to other data held by the EIAS to provide a more detailed analysis and to better monitor and inform Government policy. The 2006/07 school leavers data collection is the first to be provided at the individual level.

### **The Two Systems**

Every September, Local Authorities supply pupil census data to EIAS and Careers Scotland with data on pupils attending publicly funded schools. The ScotXed Unit use this data for the production of National Statistics and associated analyses whilst Careers Scotland use the data in their administration system, for example to track service delivery and outcomes for young people in school and beyond.

At September following the academic year of leaving, Careers Scotland carry out a survey of the destinations of school leavers and record the destination of the school leaver on their Insight system.

The different purposes of the two systems means that the data is treated in two different ways. The ScotXed system focuses on characteristics and indicators and holds a bare minimum of personally identifiable information, while the Careers Scotland system focuses on contact details.

The intention of the matching exercise is to link the school leavers current destination from the Careers Scotland System with the school pupil level record, which is held on the EIAS ScotXed system, to enable a more detailed and valuable analysis to be produced. The final data set produced is a statistical data set that does not hold contact details as it is intended for statistical analysis rather than to update an administrative system.

### **The matching process**

To carry out the more detailed analysis outlined above we need to match the destinations data held by Careers Scotland to the characteristics data held by ScotXed. In principle, this should be straightforward, since both systems hold Scottish Candidate Number which should be a

unique identifier for each school pupil. However, SCN was allocated when a pupil was entered for exams<sup>1</sup> so any pupil who left at the end of S4 might not have an SCN on either system since they could have received their SCN after the census date. In addition, pupils who left school with no qualifications, i.e. those that are most at risk of not entering a positive destination, are more likely to have left school without an SCN. There are other scenarios where a pupil may not get allocated an SCN, e.g. if they sat A-levels, instead of an SQA qualification.

For these reasons, it was agreed that the matching methodology would use SCN first and then if that is missing, school code, date of birth, postcode and gender, since these four variables are common to both systems and highly likely to be unique.

The matching process was also required to take account of potential differences in the data held about the same pupil on the two systems. Over the year, the two systems begin to diverge since Careers Scotland update their system as and when pupils move, while the statistical database remains a snapshot of census day. In addition, the statistical system only requires the data to be valid and plausible, it is only the Careers Scotland system that will confirm name and address with the actual pupil.

To account for these diverging systems a combination of fuzzy matching and deduction were used to match records from the two systems. The first stage is to match records on the basis of SCN, since these will be unique matches. The matched records are then separated from the unmatched records to reduce erroneous matches based on characteristics. For the 2006/07 leavers, 63% of records were matched at this stage.

The next stage is to match using characteristics. 26% of records matched on all four characteristics: school, gender, postcode and date of birth. These matches should contain enough unique information about the school leaver that we can be sure they are reliable.

The remaining 11% of records were matched using sub-sets of the characteristics, provided that the match was unique. Many of these remaining records had missing or incomplete details. These matches are best described as fuzzy matching or nearest neighbour matching, since we're dealing with incomplete records and some of the matches have different characteristics. Records were matched on each combination of three or two characteristics, in the same order as the table below. Unlike previous steps which were done in bulk, this process worked on a single record at a time. The first unmatched record from the destinations data was chosen and then matched on the basis of school, postcode and birthday. If this did not work, the process would move onto matching on gender, postcode and birthday, etc. until a match was found.

This method was chosen for its computational simplicity, since due to the size of the data sets involved the matching process takes several hours to run. If a record has more than one match, the first record is chosen. Once matched, records are excluded from later steps in the process.

---

<sup>1</sup> This has now changed and all pupils should now be allocated a unique SCN as part of the SCN clean-up exercise associated with the School to School system. The data used here was September 2006 pupil census data, which was collected prior to the clean-up exercise. In future, SCN should be much more complete as all current pupils should have been allocated an SCN and all new pupils will be allocated an SCN when they start school.

All matching is done within certain constraints:

- Pupil Census record must be in SP (Special School), AD (Adult Learner) or S3 or above to be included.
- School Attended must be main school attended.

Matched on	Records	Percentage
SCN	35,857	63%
All Characteristics	15,092	26%
School, Postcode and Date of Birth	25	0%
Postcode, Gender and Date of Birth	95	0%
School, Gender and Date of Birth	3,115	5%
Postcode and Date of Birth	1	0%
Gender and Date of Birth	2,503	4%
School and Date of Birth	211	0%
School, Gender and Postcode	205	0%
Gender and Postcode	46	0%
School and Postcode	79	0%
School and Gender	135	0%
Total Records	57,364	100%

### Reliability of the matches

We do not believe that the process has biased any of the relationships between characteristic and destination. However, it is impossible to put an accurate estimate on the error rate and it is advised that small differences between characteristics should be treated with caution.

89% of the matches are reliable and accurate and, whilst the remaining 11% have a degree of uncertainty to them, all results show expected trends. Where it is possible to compare results with other data sources, the differences are either very small or can be explained via different definitions and terms.

Headline figures are unaffected by this methodology and the actual numbers entering each destination are found using the same method as in previous years. Similarly, school level and Local Authority level results should not be affected by the matching process, since the matching process will not change these variables.

Different sub-groups will have different degrees of reliability. Leavers who went onto Higher Education are most likely to have an SCN and have the most reliable matches. Conversely, leavers who became unemployed are the least likely to have an SCN and are more likely to be matched on the basis of their characteristics leading to a greater likelihood of an incorrect match.

Overall, we are confident that the process has produced a statistical dataset that provides reliable information on the relationships between pupils characteristics and their destinations upon leaving school.

If you have any comments or questions on the methodology used in this survey, please contact: [ewan.smith@scotland.gsi.gov.uk](mailto:ewan.smith@scotland.gsi.gov.uk)