



Department
for Education

School Census: Data Quality and Processing

**Information on data quality and
processing for users of DfE's statistics
that are underpinned by data from the
school census**

September 2017

Contents

Introduction	3
The data collections	3
Data supply and data quality	3
Data processing and quality checks	5
Updating the data	6
Other publications	6

Introduction

This report provides detail on the school censuses run by the department to gather information from schools and local authorities on their details and that of their pupils. The report covers the content of the collections, how the data is supplied to us and the various validation checks undertaken to ensure the quality of that data.

It also explains how the data is processed and quality assured within the department and the publications which use the school census data.

Users of the various publications may find the information provided here useful additional context and background to those publications.

The data collections

We collect data on pupils from all schools. Most schools complete the [school census](#) three times per year. Completion of the school census is a statutory requirement under section 537A of the Education Act 1996.

Independent schools are required to complete a [school level census](#) annually. This is a statutory requirement on all registered independent schools.

Eleven general hospital schools complete a [school level census](#)

Pupils having their education funded by a local authority in an alternative setting are included in the annual [alternative provision census](#).

These collections give an excellent coverage of pupils educated in England. The main gap in coverage is pupils who are educated at home by the choice of their parents (without the local authority providing funding). While we recognise there are some limitations and weaknesses in the data, overall we make the judgement that the data are fit for purpose and give a fair representation of the policy area.

Data supply and data quality

Local authority maintained schools and pupil referral units submit census data to their local authority who authorise the data and then submit it to the department. Academies and other schools not maintained by a local authority are responsible for submitting their own data to the department. However, the level of validation within the submission process means we judge that the data quality is good and fit for the reporting purposes for which it is used.

The department provides a specification for the main school census for each academic year (covering three census returns) at a time. The specification and validation rules and a collection guide are sent to the suppliers at least six months prior to the collection

opening. All changes from the previous collection are highlighted and also listed in a version history. These documents are sent direct to suppliers and also published online (see [here](#) for the school census documentation).

For all school-level collections a test site is made available to enable the suppliers to undertake test data loads and to ensure their implementation of the validation rules is correct.

The data collected by the department is based on data we expect a well-run school would already hold to manage their day to day activities. These data are held within the school in a management information system. Schools are free to [select](#) a provider of their management information system but the department sets of minimum criteria for these systems. Those minimum criteria include some validation checks on the data held in the system. Most schools use a system supplied by one of 6 main providers. The department works closely with those providers, sharing information about upcoming changes to the data being collected in the census.

Within the school data items will be checked prior to being recorded. How pupils are counted is often set out in legislation or in guidance issued by the department. For example, the information to be contained in a school register are set out in section 434 of the Education Act 1996 and subsequent regulations. These regulations set out the grounds upon which a pupil's name must be deleted from the register. The checks schools must make to verify a claim for free school meals are specified in guidance.

Analysts in the department carefully consider all the data items submitted in the census. By working through the usefulness of those data to the schools themselves and the uses of the data by the department and others it is possible to identify which data items have the greatest risk of poor quality data. Processes for checking data as it is returned by schools can then be tailored to address that level of risk. In general data which is regularly used by schools in their normal day to day activity is likely to be of very high quality, whereas any data collected primarily to allow schools to complete the school census is at higher risk of poor quality.

For example, all schools need to take a register of pupils twice per day. This requirement is set out in legislation, is necessary for schools in the event they have to evacuate the building and has happened in all schools long before the school census collected absence data. On this basis we can believe the absence data in the school census are likely to be high quality. The data we collect is needed by the school itself and is recorded by the school in real time.

At the other end of the spectrum is class size data and whether a large infant class is lawfully or unlawfully large. Schools have no need to record class sizes. It isn't information that would normally be recorded locally, it is created so schools can complete the census. This is compounded by the fact that schools have no operational reason to record whether a child meets any of the ['exception criteria'](#) which would allow them to

increase a class size above 30 without breaching the lawful limit. The way the admissions process works means some school types (voluntary controlled and community schools run by the local authority) may not know this information. Therefore, this part of the census collects information that a school has no need to record and information they may not hold and this creates a higher risk of poorer quality data. To mitigate this risk data on class sizes and unlawfully large infant classes in particular are scrutinised during the live collection phase. All schools recording unlawfully large classes are asked to check this information and the data are shared with policy experts to provide reassurance the data are accurate. We have worked with the UKSA to ensure we have good processes in place and that we have clearly articulated risk inherent in this part of the data collection. Overall, we consider that, with the continued validations by the department of this data, we can judge the infant class size figures to be as good quality as possible and to be fit for reporting on this subject area.

Data processing and quality checks

Once data is submitted to the department a series of [validation checks](#) are run on the data. These include checks within the data submitted by a school, comparisons to previous data submitted by that school and checks against data supplied by other schools. Any data failing a validation checks is followed up until it is corrected or a reasonable explanation for the unusual data is provided.

To produce the statistics we begin by filtering the data supplied by schools to remove data not to be included in the publications. This includes pupils at schools outside England (service children's education schools submit data on a voluntary basis). We also try to ensure pupils are only counted once, as some pupils are on the roll of more than one school and therefore included in the census return more than once.

The school level data are then created by aggregating the pupil level data submitted by schools. Data from independent schools and general hospital schools are added to the data collected via the main school census. The school level data are published as underlying data. At school level we publish a lot of data to allow users to see detailed information for an individual school or build their own data by aggregating school level data. The published data are not suppressed or rounded (except where stated). Information about the school itself is incorporated from [Edubase](#) which is the department's register of all educational establishments.

The school level data is then aggregated to form the national and local authority tables published in the statistical first release. Other data (from the alternative provision census) are added to complete the national level data.

Once the national data are produced this undergoes further [quality assurance checks](#) with any unexpected changes in the data being investigated. The tables are produced, wherever possible, both from the underlying pupil level data and also from the core

datasets stored in SQL code. They are cross-checked against each other and against results for the previous year.

Sense checks are applied to any changes – for example, a year-on-year increase in the academy population would be expected. In addition, at the start of the process the whole dataset is investigated to identify any additional code which need incorporating into the processing.

Updating the data

Once the school census is complete it is closed and not reopened. This means schools cannot, after collection closure, amend or correct any data submitted. Therefore, the published statistics show a picture of the education system at a point in time, with no retrospective changes. This is important as allows readers to identify changes over time by comparing consecutive publications (for example, the rate at which schools have converted from other school types into academies). It also enforces to schools the importance of submitting correct census data.

Other publications

In addition to the main [schools and pupils](#) publication the school census is the basis for statistical publications on pupils with [special educational needs](#) and [absence](#).

The school census is used in the production of [school performance tables](#), to [calculate school funding](#) and for policy development and monitoring. The teams producing these provide a specification of the data they require and the results are quality assured and, wherever possible, dual run. Thus the quality checks for school census are carried through to the next layer of data.



Department
for Education

© Crown copyright 2017

This publication (not including logos) is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

To view this licence:

visit www.nationalarchives.gov.uk/doc/open-government-licence/version/3

email psi@nationalarchives.gsi.gov.uk

write to Information Policy Team, The National Archives, Kew, London, TW9 4DU

About this publication:

enquiries www.education.gov.uk/contactus

download www.gov.uk/government/publications

Reference: DFE-00260-2017



Follow us on Twitter:
[@educationgovuk](https://twitter.com/educationgovuk)



Like us on Facebook:
facebook.com/educationgovuk