Quality assurance of administrative data used in education statistics

Quality assurance report produced by NAES to investigate the administrative data sources used in the production of short-term economic indicators as set out by the UK Statistics Authority.

Contact:
Richard Perry
stoi.development@ons.gsi.gov.uk

Release date:
13 June 2017

Next release:
To be announced

Table of contents

1 . Introduction

1.1 Background

The National Accounts and Economic Statistics (NAES) Group within Office for National Statistics (ONS) download data from the Independent Schools Council (ISC) website to obtain the number of private schools and the split of day and boarding school pupils. These data form one source in the calculation of short-term economic output indicators, namely gross domestic product, output approach, (GDP(O)) and Index of Services (IOS) for the UK.

This report outlines the process data take from initial collection through to the output of the release. It identifies potential risks in data quality and accuracy as well as details of how those risks are mitigated.

This report forms the latest in a series of quality assurance of administrative data (QAAD) reports produced by NAES to investigate the administrative data sources we use in the production of short-term economic output indicators, as set out by the UK Statistics Authority's Administrative Data and Official Statistics guidance. As such, this report focuses on our administrative data use for the education industry, specifically Standard Industrial Classification (SIC) 85.2 and 85.3, whereas NAES use pupil numbers in private schools to produce aggregate figures and do not aim to cover other aspects of Division 85, as these are collected using non-administrative data sources. Separate industries where we use administrative data will be considered in other QAAD reports in the series.

Further information relating to quality and methodology for the short-term economic output indicators can be found in the Gross domestic product, preliminary estimate and the Index of Services Quality and Methodology Information document.

1.2 Standard industrial classification (SIC) overview

The activities of education cover all activities under UK Standard Industrial Classification 2007 (SIC 2007) Division 85. This includes activities of organisations representing all levels and types of education in schools, colleges and other training institutions.

Based on the UK Standard Industrial Classification 2007 (SIC 2007), the industry is classified into a number of sub-groups but we are only concerned with two:

- 85.2 – primary education

- 82.3 – secondary education

According to the Inter-Departmental Business Register (IDBR) [1] there were 9,655 enterprises classified under SICs 85.2 and 85.3 in March 2016. This is a decrease of approximately 150 enterprises (minus 1.5%) from the previous year (March 2015).

The majority of enterprises within these two SICs were allocated to "85.3 – secondary education"; which equates to 5,015 enterprises (52%).

Within the two SICs, there were 1,835 enterprises with fewer than 5 employees (19% of the division), compared with 5,895 enterprises with 20 or more employees (61% of the division).

Of the 9,655 enterprises in the two SICs, the annual turnover of 2,365 (24%) enterprises was below £250,000.

For the vast majority of these enterprises, 5,670 (58.7%) had registered turnover below £2 million, with only 90 enterprises in total having registered turnover of over £50 million.

Notes for: Introduction

1. The Inter-Departmental Business Register (IDBR) is a comprehensive list of UK businesses that is used by government for statistical purposes. It provides the main sampling frame for business surveys carried out by both ONS and other government departments. It is also an important data source for analyses of business activity.

2 . Quality assurance of administrative data (QAAD) assessment

2.1 UK Statistics Authority QAAD toolkit

The assessment of our administrative data sources has been carried out in accordance with the UK Statistics Authority Quality Assurance of Administrative Data (QAAD) toolkit.

Each administrative data source investigated has been evaluated according to the toolkit's risk and profile matrix (Table 1) reflecting the level of risk to data quality and the public interest profile of the statistics (Table 1).

Table 1: UK Statistics Authority quality assurance of administrative data (QAAD) risk and profile matrix

Level of risk of Quality concerns	Public interest profile		
	Lower	Medium	Higher
Low	Statistics of lower quality concern and lower public interest [A1]	Statistics of low quality concern and medium public interest [A1/A2]	Statistics of a low quality concern and higher public interest [A1/A2]
Medium	Statistics of medium quality concern and lower public interest [A1/A2]	Statistics of medium quality concern and medium public interest [A2]	Statistics of medium quality concern and higher public interest [A2/A3]
Higher	Statistics of higher quality concern and lower public interest [A1/A2/A3]	Statistics of higher quality concern and medium public interest [A3]	Statistics of higher quality concern and higher public interest [A3]

Source: Office for National Statistics

The toolkit outlines four specific areas for assurance and the rest of this report will focus on these areas in turn. These are:

- operational context and administrative data collection

- communication with data supply partners

- quality assurance principles, standards and checks applied by data suppliers

- producer's quality assurance investigations and documentation

In the assurance of our data source we have chosen to give a separate risk and profile matrix score (Table 1) for each of the four areas of assurance. This will allow us to focus our investigatory efforts on areas of particular risk or interest to our users (Table 2).

2.2 Assessment and justification against the QAAD risk and profile matrix

Table 2: QAAD risk and profile matrix assessment of administrative data used to measure education

	Low A1	Medium A2	High A3
Operational context and administrative data collection	A1		
Communication with data supply partners	A1		
Quality assurance principles, standards and checks by data supplier	A1		
Producers Quality Assurance Investigations and Documentation	A1		

Source: Office for National Statistics

The risk of quality concern and public interest profile has been set as "low" due to the small contribution that the independent schools data for SIC 85.2 and 85.3 feed into Index of Services (0.05%) and gross domestic product (GDP, 0.04%).

As such, a score of A1 (basic assurance) is deemed appropriate for this data source.

All scoring was carried out by the National Accounts and Economic Statistics (NAES) Group based on the level of risk of the data and the interest of our users. Results for each area of assurance for education are shown in Table 2. If you feel that this report does not adequately provide this level of assurance or you have any other feedback, please contact stoi.development@ons.gsi.gov.uk with your concerns.

3 . Areas of quality assurance of administrative data (QAAD)

3.1 Operational context and administrative data collection (risk QAAD matrix score: A1)

This relates to the need for statistical producers to gain an understanding of the environment and processes in which the administrative data are being complied and the factors that might increase the risks to the quality of the administrative data.

The process detailed in this section is summarised graphically in Figure 1 in the Annex.

The Independent Schools Council (ISC) is a limited company whose work is carried out by a small number of staff working in London. They output a series of articles and analyses to their website throughout the year, and produce statistical and economic reports on their membership. They are assisted by contributions from expert advisory groups in specialist areas and are tasked by their members to protect and promote the sector in everything they do. Their work is directed by a board, which comprises individuals nominated for appointment by each of their member associations.

The ISC brings together seven associations and four affiliated ones of independent schools, their heads, bursars and governors. Through these associations they represent over 1,200 independent schools in the UK and overseas. They educate over 500,000 children each year.

Schools become members of ISC by registering to become a member of one or more national associations. It is a requirement of membership of these associations that the school makes a submission for the annual census. The only exception to this is schools in Northern Ireland that are a member of one of the associations, and so are displayed on their school search page, but do not make a submission. Note that membership includes some schools in the Isle of Man and Channel Islands.

The type of school in their membership is varied – school sizes can range from under 25 pupils to over 2,000 and cover senior, junior and mixed age ranges. Schools can have boarders or be day only; they can also be single sex or coeducational. Some of the schools cater specifically for pupils with special needs.

To return data schools use an online data collection system, which is password protected. ISC provide a guidance document that gives details of how to complete all the forms. Non-responders are identified and chased directly by ISC, as well as being pressured by their association to respond. Nearly every year ISC obtain a 100% response and other years are near 100%. Quality checks are further detailed in section 3.3.

Pupils in ISC schools account for around 80% of the total number of pupils in independent schools in the UK. The UK independent sector as a whole educates around 625,000 children in around 2,600 schools.

3.2 Communication with data supply partners (risk QAAD matrix score: A1)

This relates to the need to maintain effective relationships with suppliers (through written agreements such as service level agreements or memoranda of understanding), which include change management processes and the consideration of statistical needs when changes are being made to relevant administrative systems.

The process detailed in this section is summarised graphically in Figure 2 in the Annex.

The Independent Schools Council (ISC) holds contact details for all school heads, who can update contact details via a "self-service" approach through the member zone on the ISC website. The ISC also update these details when they are supplied via the school's census submission. Once the school has accessed the census form and entered their contact details, the ISC use this person as the primary point of contact.

The ISC contact schools initially in December to remind them the census is approaching and send a guidance document that gives details of all the forms to complete and how to assign pupils to different categories where appropriate.

The ISC contact schools again just before census day to launch the census and inform them that the data entry form is available via the data portal. The ISC data portal is restricted with a school's individual login only communicated to the school head and passwords are changed at regular intervals.

Generally, the census is completed without the need for the intervention of the associations. However, the ISC do sometimes engage with associations to encourage their member schools to submit data when deadlines have been missed.

The National Accounts and Economic Statistics (NAES) Group has not contacted the ISC before as data have been freely available on the website. They deemed the ISC data source to have a very low proportional impact on the statistics that feed into the Index of Services (1.8%) and gross domestic product (GDP, 1.1%) and feel the level of assurance carried out throughout the process is sufficient and proportionate to its importance.

The next step for NAES going forward is to establish contact with ISC using the contact details acquired during the quality assurance of administrative data (QAAD) review.

3.3 Quality assurance principles, standards and checks by data supplier (risk QAAD matrix score: A1)

This relates to the validation checks and procedures undertaken by the data supplier, any process of audit of the operational system and any steps taken to determine the accuracy of the administrative data.

The process detailed in this section is summarised graphically in Figure 3 in the Annex.

Quality assurance of data submitted by the schools to the Independent Schools Council (ISC) occurs in three main stages.

Firstly, the census submission form has inbuilt automatic checks. When each page of data is submitted, scripts run in the background to check logical errors (for example, totals not entered when data for a particular category has been entered) and longitudinal errors (for example, large percentage change in pupil numbers from last year). Schools cannot submit their form until they have either corrected it or entered an explanation against the error; for example, in the case of large pupil number changes, they may be correct if the school has expanded significantly. Once schools have done this, the submitted form is sent for checking by the Research Team.

The ISC Research Team have an error checker tool, which is a more complex series of scripts that run on the data to cross-check between screens and check the logic of different sections of the data, for example, is the pupil teacher ratio within a normal range, or are there unusually low numbers of new pupils. All completed forms are also visually inspected screen by screen and compared with the previous year's submission for anomalies. Any errors found are added to the relevant screen and then the form is returned to the school for them to correct their data and resubmit. The school's submission will then be rechecked in its entirety by the ISC Research Team and may be returned with further issues to be corrected. This process continues until no errors are found and the form enters the "completed" status.

Once all schools have submitted, trends are analysed across the whole dataset and broken down by different school types as part of preparing the census publication. Any unusual trends are investigated and schools may be contacted again to clarify their submission if their figures appear to be skewing the data by being significantly different from schools in a similar category. There are some revisions due to late responding schools but they are very small.

3.4 Producers quality assurance investigations and documentation (risk QAAD matrix score: A1)

This relates to the quality assurance conducted by the statistical producer, including corroboration against other data sources.

The process detailed in this section is summarised graphically in Figure 4 in the Annex.

The National Accounts and Economic Statistics (NAES) Group download data from the annual report on the Independent Schools Council (ISC) website in April each year, using detailed desk instructions. Data is loaded into two systems, first Excel then CORD with no formal checks applied other than by eye comparison with previous data. If any discrepancy is spotted, NAES would investigate in CORD to discover the issue and take any required action. A simple calculation is run to aggregate the figure into a higher level SIC 85. For boarding pupils, the most recent figures for male and female in the totals column are added together and for day pupils, the most recent figures for male and female in the totals column are added together. The number of schools is simply the given value.

In the event of missing data, imputation is built into the system that forecasts a figure based on previous year's data. NAES confirmed that they compare the education figures, including revisions, with those from previous years – revisions occur every year.

They deemed the education data source to have a very low proportional impact on the independent school statistics that feed into the Index of Services (1.8%) and gross domestic product (GDP, 1.1%) and feel the level of assurance carried out throughout the process is sufficient and proportionate to its importance.

Revisions are made by the ISC, which NAES check for and update as required. NAES maintain a full audit log of changes and revisions made on a briefing document for each round.

Going forward, NAES aim to establish communication with the ISC Research Team to learn more about the data and processes.

4 . Summary

The overall flow of education statistics from suppliers to Office for National Statistics is summarised graphically in Figure 5 in the [Annex](#).

In investigating the administrative source for the activities of education, the National Accounts and Economic Statistics (NAES) Group consider the main strengths of the data for our purpose to be:

- detailed and extensive quality assurance by the Independent Schools Council (ISC)

- full census, with 100% response most years

- simple NAES processes

- ISC very knowledgeable and quick to answer queries

We believe current limitations of this data source are:

- not all private schools are part of the ISC census as schools are not forced to be in an association

- no communication between NAES and ISC, national associations or schools; this should be addressed by establishing links with the ISC

In constantly seeking to improve our data sources we will be taking steps to address these limitations and these will be communicated to users in the future quality assurance of administrative data (QAAD) report updates for this topic.

However, despite these slight limitations, based on the low risk of quality concerns and the small contribution that the independent school statistics feed into index of services (0.05%) and gross domestic product (0.04%), NAES consider this data source to fulfil the requirements of an A1 assurance rating.