



Department  
for Education

# Quality assurance of Education Statistics

June 2017

# Contents

Summary	3
Expiry or review date	3
Who is this publication for?	3
Introduction	4
Production of a statistical outputs	4
Stages of Production	6
Project planning	6
Data collection	6
Input Validation	7
Data delivery	7
Data transformation	8
Data tables	8
Publication of statistics	8
Prepare briefing	9
Post publication	9
Other useful guidance	10

## Summary

The Department for Education (DfE) is committed to producing accurate, timely, high quality official statistics publications that take into account user needs and which are produced and disseminated in accordance with the UK Statistics Authority's (UKSA) Code of Practice. This note sets out the principles currently being embedded within the DfE to ensure we produce quality assured statistics.

## Expiry or review date

This guidance will be reviewed before December 2018.

## Who is this publication for?

This guidance is for:

- Users of DfE statistics

## Introduction

This document aims to provide an overview of the principles currently being embedded within the Department for Education (DfE) to ensure the quality of our statistics. It outlines the general steps involved in the statistics production process and the quality assurance processes involved at each stage. Where escalation or higher-level sign off is required, this has been noted. The document has been broken down into phases of the production process. Not all steps are required for each release. These processes are cyclical as part of any release schedule and form the basis of continual improvement. Each of the stages below are covered later in this document.

## Production of a statistical outputs



This document summarizes how producers of education statistics manage data quality through the identification and correction of each type of data error throughout the production process. Quality assurance checks should be run at each stage, this enables sources of error to be detected and corrected more efficiently. Errors can be introduced through respondent gaps, respondent error, coding errors, transformation errors and the escalation of existing or new errors through data processing.

The aims of quality assurance are to:

- to provide high quality statistics that meet user needs
- to establish public trust in the final outputs
- to deliver value for money products
- to protect confidential personal or commercially sensitive information.

Quality assurance should use resources that are proportionate to the task concerned. Too little resource – in terms of either staff or technology will increase the likelihood of errors in the data. Any error has the potential to undermine public confidence in our data, influence policy decisions and impact on budget allocations. Too much resource is not cost effective for the general public. Spending too much time quality assuring one-step of the process may introduce unnecessary risks to quality later in the process as pre-announced publication deadlines need to be met. The production process has been broken down into phases to highlight the main risks in each step and the appropriate checks to mitigate these.

# Stages of Production

## Project planning

Effective planning is key to any quality assurance process. It is important to know what the main risks to quality are and which of these may have increased from the last production. Assigning roles to tasks and identifying where sign off is required is key to ensuring these risks are managed effectively through the process.

This document is not intended to be a guide to project management. However, it is important to ensure that attention to the quality of the output (whether that is in the form of data, a report on key findings or secondary analysis) is factored in across all stages of the plan. For example, customer requirements need to be determined to ensure that at the end of the process the output is relevant and meets user needs. It is important to ensure that quality assurance documents, guidance notes and process plans are also updated and that lessons learned from previous productions are factored in to the process. Risk and issue logs need to take account of new data items, revisions and system changes so that the necessary checks are in place to ensure the quality of the output. Time for robust quality assurance needs to be a key feature in the plan including an allowance for independent checking where possible.

## Data collection

Data collection is often carried out by an external organisation, such as an executive agency, other government departments or private firms. It is important as part of this process that the checks undertaken meet our needs. Many of the checks undertaken at this stage are regarding the process of data collection and extraction from a data repository.

Checks should ensure that the data definitions for collection are consistent with previous years. Extraction code to group or repackage data should be scrutinised to ensure consistency with previous years. Any inconsistencies (which could be with other similar systems) should be reviewed i.e. to determine whether they affect the time series or to provide an explanation of the differences between two publications. Other changes to the collection process should be identified and assessed in the same way.

The raw dataset resulting from this collection should be checked for missing rows, columns and individual cells. Any omissions should be explained or addressed as relevant. The dataset should also be checked to ensure that rules of entry are met i.e. do totals match the sum of components, do totals match across sections as relevant and are other data conditions (including FTE not exceeding staff numbers) being met.

Any issues identified at this stage need to be reported to the supplier and the responsible statistician as soon as possible.

## **Input Validation**

Data input is often carried out by an external organisation, such as an executive agency, other government departments or private firms. It is important as part of this process that the checks undertaken meet our needs. Whether data is collected and input (internally or externally), robust validation checks should be in place to ensure data is submitted in the correct format without error. Much of this is likely to be automated with an 'error' occurring if a value is outside the expected range; however, there may be anomalies in situations, which are not usual but can occasionally occur. These queries may need to be dealt with on a case-by-case basis. Internal/external validation codes should be checked prior to data collection. It is important to collect information on errors and queries that occur so that users can understand the limitations of the data and so that processes and checks can be improved for the next collection.

## **Data delivery**

When in receipt of data from an external supplier. It is important to check that the data received appears, as the supplier/we would expect. It is possible that data has corrupted during the transfer or the wrong file was sent. It is important, therefore, to run some checks that ensure the file received looks as expected. Enough time should be available to enable any issues to be resolved in time for preparing the final outputs. Issues found at this early stage save costs in the end so it is worth factoring in enough time ensure this is done properly.

Further checks on the provider level data (usually institutions) should be carried out. For example, visual tools such as scatterplots should be used to identify any outliers in year on year trends. Follow up anything that looks odd and gather evidence for any justifiable reasons. Some of these may be important stories for the national trend analysis. Once again, ensure any validation rules, such as total checks have been met across all providers. Some checks on scenario testing and/or where some calculations/code are complex may be required at this stage.

Production should not proceed if there are any concerns at this point, and they need to be flagged to the responsible statistician as well as getting in touch with the supplier as soon as possible.

## Data transformation

This stage covers the process of converting data provided / collected into the tables used in the publication. When transforming data from one product to another it is possible that formulas and formatting introduce errors in the presentation of information. Manual transformations are at higher risk due to human error and so checks (including checking code and dual running if necessary) should be carried out to ensure any process has been carried out correctly. Automated processes should be checked for consistency against previous productions. Any changes to the data collection process may introduce errors due to changes in field orders, new coding sets etc... Therefore, it is important to run spot checks on the outputs to ensure the numbers produced are as expected.

## Data tables

The final data tables may form part of a publication alongside commentary or be available as supplementary tables. It is important to check the data appears sensible and in line with historic or other published data sources, particularly where new data items and/or new tables are being presented. Any differences should be investigated and explained where applicable. Tables that summarise the data in different ways should be compared to ensure figures match. Figures should also be compared to external sources where applicable. It is important to include the limitations of the data to ensure users draw appropriate conclusions i.e. if there are changes in measures across years - comparability over time may be limited.

Producing published tables often forms the end of a long process and often means limited time is available. It is important to ensure enough resource (including an independent reviewer where appropriate) is available to perform these checks as any errors are easily spotted by users and undermines their confidence in our data. Checks should include formatting, headings and footnotes as errors in these undermine the trust others place in our data. Suppression checks are also important as a breach of confidentiality undermines the trust the public place in providing their information and has possible legal ramifications.

## Publication of statistics

The final release of official statistics is often the only document many of our users see. Prior to final publication, it is important to ensure enough resource is available (including an independent reviewer where appropriate), to thoroughly check the document. It is worth stressing the importance of performing these checks as users easily spot any errors and undermines their confidence in our statistics. Checks should include ensuring the messages behind these accurate and up to date, any unusual changes been explained, that the methodology is explained, that trend descriptors, titles, labels are correct and timeframes are up to date. Checks should be in place to ensure that



percentages are calculated correctly, numbers in the text match the tables, and a spell check has been run on the data in English (UK).

## **Prepare briefing**

Some basic checks on any briefing material needs to be undertaken: such as the correct security markings, ensuring only those with pre-release access receive information early and are aware they should not share with others. For press lines and Q&A and press statements: ensure all trends and numbers used here are consistent with the official statistics publication and that the wording is clear including any definitions about what the data means so lines are not misinterpreted. Any press office re drafting of lines should be cleared by the responsible statistician

## **Post publication**

It is important to ensure that publication itself occurs as planned and that no corruption has occurred. It is important to run a lessons learned exercise and share the findings with other team leaders, ensuring any issues identified are assigned appropriate actions, which are allocated enough time to address before the next release. User feedback should be collected and used to identify areas for future improvement. In addition collating information via PQs and FOIs may prove useful in developing the future content of official statistics. For annual publications, this may be collected throughout the year.

## Other useful guidance

For assistance in producing statistical release templates: <http://style.ons.gov.uk/>

For general analysis guidance, see the aqua book:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/416478/aqua\\_book\\_final\\_web.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/416478/aqua_book_final_web.pdf)



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](http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3)

email [psi@nationalarchives.gsi.gov.uk](mailto: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](http://www.education.gov.uk/contactus)

download [www.gov.uk/government/publications](http://www.gov.uk/government/publications)

Reference: DFE-00158-2017



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



Like us on Facebook:  
[facebook.com/educationgovuk](https://facebook.com/educationgovuk)