

REPORT OF A BREACH OF THE CODE OF PRACTICE FOR STATISTICS

1. Core Information [\[guidance\]](#)

Title and link to statistical output	Widening Participation: UK Performance Indicators 2018/19
Name of producer organisation	Higher Education Statistics Agency
Name and contact details of person dealing with report	Rebecca Mantle, Official Statistics Manager rebecca.mantle@hesa.ac.uk 01242 211148
Link to published statement about the breach (if relevant)	N/A
Date of breach report	19 February 2020

2. Circumstances of breach [\[guidance\]](#)

Relevant principle(s) and practice(s)	Practice T3.6: Statistics should be released to all users at 9.30am on a weekday. Practice Q3.1: Statistics should be produced to a level of quality that meets users' needs.
Date of occurrence of breach	13 February 2020

Give an account of what happened including roles of persons involved, dates, times etc

On the morning we were due to release the Widening Participation: UK Performance Indicators 2018/19, HESA's official statistics team understood that some data were missing from tables due to be published (specifically data for some higher education providers). This was brought to the team's attention by an individual who had pre-release access to the data 24 hours prior to publication. The individuals who produced the release were able to rerun the data into the tables before they were published at the scheduled time of 09:30, but the csv source data files that we supply alongside the interactive tables were not available until 10:10.

Additional to this, one of the charts published within the release was not updated with correct data until 09:44 on the day of release.

Later the same day a user contacted the official statistics team to inform us that a value of "#ERROR!" was displaying in some published tables. Affected tables were corrected as soon as possible and a note was added to inform users:

Note: The significance marker for UKPIs Table T1, T2a & T2b for the academic years 2015/16 to 2017/18 was showing incorrectly as "#ERROR!" when "+" should have been displayed. Tables were corrected and updated by 15:40 on 13 Feb 2020.

Following these issues, the team did a further check of the data to ensure there were no remaining issues. Further investigations following the identification of these errors were undertaken to understand how they emerged.

In terms of the production process of the data release, individual HE provider data for HESA's UK Performance Indicators are previewed with providers in advance of publication of the data for QA purposes. The data created for that preview underwent a rigorous quality assurance process and accurate data was disseminated. The final part of the production process involves code to manipulate the data into a format suitable for publishing on. This

code was linking to a provider lookup list that was out of date and hence meant that some HE providers were missing from what was due to be published. Due to the rush to update the data on the morning of the release, the “#ERROR!” values crept into three tables unnoticed by the team. On investigation into this, the errors occurred because of manual copying and pasting of data from one software programme to another, leading to misrecognition of symbols by the software.

3. Impact of the breach [\[guidance\]](#)

Provide details of the impact of the breach both inside the producer body and externally

In relation to the csv source files: Prior to adding these to the release, users of the data could interact and see all of the data within the publication but would not have been able to download full source data until 40 minutes after publication.

In relation to the incorrect chart: Google analytics showed that between 09:30 and 09:44 no-one clicked into that chart nor interacted with filters on the chart.

In relation to the “#ERROR!” displaying in tables: The affected tables default to showing data for the latest academic year. Due to the fact that the error appeared in earlier years of the data only, we can be sure that only those individuals who used the academic year filter or who downloaded full source data would have seen the “#ERROR!” values. Google analytics showed that between 09:30 and 15:40 there were 79 unique interactions with the year filters across the affected tables. It also shows us that there were 4 unique downloads of the source data during this time.

4. Corrective actions (taken or planned) to prevent re-occurrence [\[guidance\]](#)

Describe the short-term actions made to redress the situation and the longer term changes to procedures etc

In addition to correcting the data as soon as possible and alerting users to the update to tables, we are looking to build further quality assurance into the process for producing our UK Performance Indicators. Specifically, this would involve a more thorough check of the data in its final format for publication. Rather than relying on spot-checks alone, we will run an independent full check of the dataset in its final format and test that process with previous years' data to ensure it works as we would need it to do before implementing for the next release of data.

In addition to this, the provider lookup list being referred to within existing processes will in future be based on dynamic code rather than be hard coded into the process, thereby making the code more robust.

Finally, we are looking to adapt our processes so that we need never rely on manual copying and pasting of data. We will do this by ensuring our automated workflows are efficient and repeatable so that they can be rerun whenever required.