



GRADUATE OUTCOMES - A SUMMARY OF THE YEAR TWO SOC CODING ASSURANCE

HESA RESEARCH AND INSIGHT TEAM
2021-04-29



HESA Communications Team

HESA

95 Promenade

Cheltenham

GL50 1HZ

E communications@hesa.ac.uk

T +44 (0) 1242 388 513

W www.hesa.ac.uk

Higher Education Statistics Agency Limited is a company limited by guarantee, registered in England at 95 Promenade, Cheltenham, GL50 1HZ. Registered No. 02766993. Registered Charity No. 1039709. Certified to ISO 27001. The members are Universities UK and GuildHE.

CONTENTS

1- Introduction.....	3
1.1- The quality assurance of SOC coding.....	3
1.2- Approach to higher education provider feedback	3
1.3- Additional quality assurance checks	4
2- Methodology/ Process	4
2.1- Process for reviewing occupation groups	4
3- Results/Outcomes.....	5
3.1- Results of higher education provider feedback.....	5
3.2- Additional quality checks	6
3.2.1- Update from SOC2010 to SOC2020	6
3.2.2- Detailed quality checking exercises	6
4- Discussion/Problems encountered.....	7
4.1- Feedback for higher education Providers and Common Misunderstandings.....	7
4.2- Next Steps.....	7

1- INTRODUCTION

1.1- THE QUALITY ASSURANCE OF SOC CODING

An effective assurance process is required for the Standard Occupational Classification (SOC) coding of Graduate Outcomes, helping to ensure that the dataset is as high-quality and consistent as possible. Steps are taken to ensure the prevention of errors and Graduate Outcomes uses the most up-to-date classification system to align with the current labour market. However, due to the complex and subjective nature of SOC coding, as well as the need for the timely return of information, detection and correction of errors is required.

There are a number of key stages involved in the quality assurance of the SOC coding for Graduate Outcomes. This includes the review of provider feedback, which runs alongside the consistency checking completed by the coding suppliers Oblong, as well as several other checks designed to identify non-random anomalies in the dataset.

1.2- APPROACH TO HIGHER EDUCATION PROVIDER FEEDBACK

Overall, 3048 queries were received from 55 higher education (HE) providers this year, compared to year one when 2500 queries were received from 90 HE providers. The deadline for feedback was 6 January 2021, allowing HE providers some time to review their graduates after the final cohort had closed.

Initially, feedback was received from HE providers in the SOC feedback template, and a master log was compiled by HESA detailing HUSID/FEPUSID, job title, job duties, assigned SOC code and proposed SOC code, a justification for the change, HESA's response to the query and HESA's comments on the request. This made the process far more streamlined than it was previously when HE providers supplied feedback in any format, which had to be manually compiled into a log. The new process made it far easier to ensure occupation groups were not repeatedly checked, allowed queries to be grouped by their job titles for comparison purposes and meant that the query being raised was clearer. As before, this assessment does not take into account the name of the HE provider in order to locate systemic issues across the entire dataset.

After checking, queries were marked as one of the following categories.

Category	Description	Action
Systemic	Widespread errors that require a change in the coding process for an occupation group	Passed to Oblong for further review and amendment if necessary
Inconsistent	Where multiple records in an occupation group are coded inconsistently and randomly	Passed to Oblong for further review and amendment if necessary
Non-systemic	Isolated cases that are below the threshold for systemic*	Not passed to Oblong
Not actionable	No basis or evidence exists for the coding to be changed	Not passed to Oblong

*Systemic and inconsistent issues were generally considered to be occupation groups with at least five records and 10% of the sample (in that occupation) impacted.

The term 'inconsistent' encompasses a variety of coding issues, so some further clarity is below:

- Most graduates in this role are coded correctly, however some have been placed into a different code that is incorrect.
- Graduates are coded between two potentially correct codes, in a seemingly random way.
- There are a number of codes relevant to a role, assigned dependent on job duties, but some graduates have been coded incorrectly between the two.
- Some of the graduates are coded into a 'Not Elsewhere Classified' minor group that could potentially be coded into a more specific minor group within the same major group.

1.3- ADDITIONAL QUALITY ASSURANCE CHECKS

A set of additional checks have also been identified and completed. Some of these were discussed with the Graduate Outcomes Steering Group last year and the rest have emerged from our own review. These include additional quality assurance processes designed to aid in identifying non-random anomalies in the dataset. The checks included the review of 0001 (uncodable) records, an assessment of average salary by major group and comparisons of national, HE provider and subject level SOC groupings. Systemic issues identified in the HE provider feedback process were also reviewed. These checks were completed on the near-final version of the dataset received from Oblong following their consistency work across both years. They are highlighted in further detail in Section 3.

2- METHODOLOGY/ PROCESS

2.1- PROCESS FOR REVIEWING OCCUPATION GROUPS

To aid in the review the [CASCOT coding tool](#) was used alongside the three [SOC 2020 volumes](#) from ONS. These highlight the underlying principles of coding and include the ONS coding indexes in full. Comparisons were made between the job title, duties and employer details entered by the graduates for each occupation in question and the guidance available, helping to determine if the correct SOC code was assigned. As quality assurance is an ongoing process it was also important to refer to decisions made in previous reviews in relation to coding, and where there was uncertainty, queries were raised with Oblong. This year we also enhanced our engagement with ONS by consulting them on issues which required clarification on the correct interpretation of their coding indexes.

3- RESULTS/OUTCOMES

3.1- RESULTS OF HIGHER EDUCATION PROVIDER FEEDBACK

The table below provides a breakdown of the feedback from HE providers. Many queries were repeated, so the breakdown supplies the total rows of feedback in each category and the number of occupation groups that were raised with Oblong as a result.

	Total Rows of Feedback	Issues Raised with Oblong
Systemic	40	12
Inconsistent	97	30
Non-systemic/ Not actionable	2911	5
Total	3048	47

Although there were more queries received this year, with 3048 queries overall compared to 2500 in year one, there was a considerable improvement upon the numbers of issues identified. In year one, 66 occupations were confirmed as systemic or inconsistent compared to the 42 that were identified this year. At the time of this assessment Oblong were carrying out their own consistency checks on the entire collection. It is very likely that several inconsistencies identified through HE provider feedback would have also been picked up by Oblong's quality assurance processes.

It is also worth noting that 95.5% of all queries were deemed non-systemic or not actionable.

The 12 systemic issues identified from HE provider feedback process were raised with Oblong. Once recoded these groups were thoroughly checked in the wider dataset to ensure that there were no longer widespread issues. The systemic groups identified were as follows:

- Agent's Assistant
- Case workers for MPs
- Constituency Assistant
- DevOps Engineer
- Digital Learning Developer
- Occupational Health Technician
- Probation Service Officer
- Supply Chain Planner
- Teaching Assistant (with HLTA duties)
- Hearing Aid Audiologist
- Support Workers
- Research Technicians

3.2- ADDITIONAL QUALITY CHECKS

Quality checks were not limited to the assessment of HE provider feedback. There were areas identified throughout the year that were checked in the final dataset and additional exercises were completed that were designed to identify any further issues.

3.2.1- Update from SOC2010 to SOC2020

Year one was initially coded to SOC2010 prior to the release of the SOC2020 coding framework. A large-scale quality assurance project was completed on the dataset after it was coded to SOC2020. It considered areas in the framework with multiple possible new codes, codes that had changed between major groups and occupations with changes within the same major groups. The overall impact of the changes on the distributions of the datasets was included, as well as a thorough review of the most common codes and job titles impacted by the update. Any inconsistencies were raised with Oblong and were highlighted to be checked in the final dataset for both years to ensure they had been rectified.

3.2.2- Detailed quality checking exercises

There were a number of other exercises that took place to aid in ensuring data was of a high quality, both throughout the year and upon receipt of the final dataset.

Graduates who had selected two instances of employment, but selected that these were the same activity, were checked if the assigned codes did not match. Generally, differences were only present where the employment type (i.e., self-employed or paid work for an employer) impacted the code that should be assigned, causing small minor group or acceptable major group variations. The number of instances with issues was below systemic levels.

Average salary distribution was checked for graduates in full-time employment, by employment activity type and major group. There were no concerning differences or patterns between years or employment activities.

Additionally, the distribution of records in each major group across years was assessed. The distribution was incredibly similar across both years of the survey, with no areas of major change even when split by subject. Subject areas with slight differences were assessed to confirm that these were not caused by any coding issues and that these differences were plausible.

Further to this, an assessment of SOC major group split across subjects was completed for years one and two. The distributions were checked for each subject and major group to ensure that there were no areas of concern and that the distribution of codes is expected.

Finally, the entire set of uncodables was checked for year two and were deemed to be either correctly coded as 0001 or with issues below systemic levels.

4- DISCUSSION/PROBLEMS ENCOUNTERED

4.1- FEEDBACK FOR HIGHER EDUCATION PROVIDERS AND COMMON MISUNDERSTANDINGS

There are a number of issues that were widespread in HE provider feedback:

- Salary cannot be solely used to denote a 'professional role'. There are a wide range of factors that influence salary, including location, age of graduate, employer etc. It is not a reliable predictor of major group. An assessment was carried out on the overall salary distribution by SOC major groups at a national level in the additional quality checks described above and as requested by the sector.
- It seems sometimes an earlier stream of data had been taken by a HE provider and all 0001 records were included in their feedback, even though most had been coded since. All 0001 records are checked, so do not need to be included in feedback in future.
- A job cannot necessarily be coded differently just because a graduate has a degree or is highly qualified for a role, especially without further evidence. Equally, there is an assumption that if a degree is required an occupation must be in Major Groups 1, 2 or 3, which is not always the case.
- There are certain employers that HE providers argue should be coded differently regardless of job title and job duties. This is not an approach we recommend.
- There was an assumption that graduates had been provided two codes in error, when these were actually codes for the graduates with two different types of employment. This also led to the misunderstanding that one employment had been prioritised over the other based on this, which is not correct. There will sometimes appear to be two instances of the same job if a graduate says that their self-employed instance is the same as their employed instance.
- It seemed that some HE providers had linked their graduates to the wrong employment instance, for example if a graduate had been assigned a code for a portfolio and a paid work instance.
- Advice by ONS to ignore the prefixing word if it does not impact coding only refers to words that come before the occupation, and not words that come after it, which was not understood by some HE providers. For example, assistant auditor would be coded as an auditor, but auditor's assistant requires assistant to be used as an indexing word, and is therefore coded as an assistant role.
- Further guidance on the information used to determine coding can be found on the [HESA website](#).

4.2- NEXT STEPS

We will be assessing our work in this area and we will look to inform the sector about our plans in the coming months following publications of this summary.



HESA Communications Team

HESA
95 Promenade
Cheltenham
GL50 1HZ
E communications@hesa.ac.uk
T +44 (0) 1242 388 513
W www.hesa.ac.uk

