GRADUATE OUTCOMES - SUMMARY OF YEAR THREE SOC CODING ASSURANCE

HESA RESEARCH AND INSIGHT TEAM
2022-04-13

CONTENTS

1 INTRODUCTION AND PROCESS

1.1 THE QUALITY ASSURANCE OF SOC CODING

Each year, the Standard Occupational Classification (SOC) coding of Graduate Outcomes undergoes multiple rounds of quality assurance. This includes the review of provider feedback, consistency checking of the data by the coding supplier (Oblong) and a series of checks carried out by HESA to identify non-random anomalies in the dataset.

Graduate Outcomes' SOC coding uses the most up-to-date classification system that is available at the time of coding. The review used the three SOC 2020 volumes from the Office for National Statistics (ONS), which included the coding indexes, as well as the CASCOT coding tool. Job title, job duties and employer details were all considered in assessing whether records were assigned the correct SOC code. HESA also referred to any coding guidance received previously from the ONS and sought further advice on any areas of uncertainty.

1.2 APPROACH TO REVIEWING HE PROVIDER FEEDBACK

This is the third year of the feedback process which received a total of 2,108 queries from 37 higher education (HE) providers. The level of feedback received is lower than year two where 3,048 queries were received from 55 HE providers. To allow providers time to review their data after the closure of the final cohort, the deadline for feedback was 7 January 2022. We'd like to thank all of the HE providers who took the time to supply feedback to HESA.

In the first stage of the process, feedback from HE providers was sent to HESA using the SOC feedback template (available on the HESA website) which ensures the correct level of detail is provided. Feedback is then collated into a single repository with provider names excluded to maintain neutrality. This is the same process used for year two of the survey (details of which can be found in the year two summary of the SOC coding assurance report), and continues to ensure the process is as efficient as possible. As with previous years, each query is assessed across the entire dataset, regardless of the provider who raised it.

Following an initial review by HESA, each query is assigned to one of the following categories and has the corresponding action:

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, across the entire dataset.

To aid further understanding of each outcome group, we have provided examples from the provider feedback in section 3. We hope you find this useful.

1.3 ADDITIONAL QUALITY CHECKS

Alongside the HE provider feedback review, Oblong were also carrying out their in-depth consistency checking exercise across the entire dataset. This helped to ensure a consistent application of coding across the year. Once complete, a further set of assessments were carried out on the updated SOC data, which attempted to identify non-random anomalies in the dataset. Detail on some of these checks is included in this report. For example, 0001 (uncodable) records in the dataset were examined to ensure codes had been assigned where possible. Some checks were also carried out based on various comparisons of records by SOC major group, activity type and subject groupings. More detail on the outcomes of these assessments can be found in section 2.2.

2 OUTCOMES

2.1 RESULTS OF HE PROVIDER FEEDBACK

A summary of the feedback and outcomes is shown in the table below. As with previous years, many providers raised similar queries or returned multiple rows of feedback for an occupation group, therefore the issues raised with Oblong were grouped by occupation. It is also worth noting that although some of the individual queries raised by providers may not be examples of miscoding in themselves, the entire occupation group was checked and where a systemic issue or inconsistency was identified, it was raised with Oblong.

	Number of occupation groups reviewed by Oblong
Systemic	6
Inconsistent	34
Non-systemic/Not actionable	16

The results from this year's assessment highlight a continued reduction in the number of issues identified as a result of HE provider feedback. In year one, 66 issues were identified as either inconsistent or systemic, reducing to 42 in year two and 40 in year three. The number of systemic issues identified this year is far lower, with only six systemic issues resulting from the process, compared to 12 last year.

The six systemic issues raised with Oblong in the year three HE provider feedback process were as follows:
- Social care worker
- Location Marshal
- Wealth manager
- Performance analyst
- Maintenance planner
- Farmer

The coding supplier was carrying out consistency checks on the entire collection as the HE provider feedback process was ongoing, and it is likely that some of the systemic or inconsistent

issues would have been rectified in their assurance work anyway. This simultaneous quality checking is required to ensure that provider feedback is reviewed and incorporated into the dataset so that the additional quality checks and final data delivery can occur in a timely way. This means that as providers were basing their assessment on earlier versions of the dataset, several queries had already been resolved in later iterations of the 'provisional' or 'raw' dataset. These were marked as not actionable queries.

2.2 ADDITIONAL QUALITY CHECKS

This section outlines some of the additional quality checks that HESA carry out as part of the process and the outcomes from these checks.

2.2.1 0001 (uncodable) records

An assessment of 0001 (uncodable) records was carried out by HESA to determine if there were any additional records that could have been assigned a code. In total, 40 records were returned to Oblong to be checked again as a result of this exercise, after HESA determined that they may have an applicable code in the indexes. However, many of these records were sent with the knowledge that they may still be deemed as uncodable if there was not sufficient information available to code. Of these 40 records, 20 were successfully assigned a code, which also led to further refinement of the coding process for one occupation.

2.2.2 Same activity consistency checking

In cases where graduates selected two instances of employment, but marked that these were the same activity, a code will be assigned separately to both instances of employment. Although a difference in employment type can alter a code, instances where the assigned codes did not match were checked for inconsistencies and changed where appropriate. Instances with one coded record and one 0001 (uncodable) record were returned to Oblong for checking, leading to an additional set of records being assigned a code.

2.2.3 SOC major group by subject

The distribution of SOC major groups was assessed by subject, using the Common Aggregation Hierarchy level 1 (CAH level 1). This was used to assess large subject-occupation groupings with a view to identify unexpected trends. For example, these checks would flag an error if a large number of medicine and dentistry graduates were coded as major group 7 under sales and customer service occupations.

There were no issues identified as a cause for any unexpected trends, as the occupations within the groups checked were coded correctly.

2.2.4 Distribution of SOC major groups

The distribution of SOC major groups across survey years was checked to ensure that there were no major differences that may be cause for concern. Overall distributions were similar and offered reassurance that there were no major discrepancies.

3 FURTHER DETAIL - EXAMPLES BY CODING GROUP

We're committed to ensuring this process is as transparent as possible. These reports provide an opportunity for HESA to demonstrate our comprehensive process and aim to support providers in carrying out their review in an efficient way. To facilitate a better understanding of the outcome groups, we have provided an example for each group, taken directly from the queries raised by the sector.

3.1 Systemic: Performance Analyst

This year's process saw a noticeable reduction in the number of systemic issues identified. Equally, the issues that were identified generally impacted smaller occupation groups. The issues were not always relevant to previous years of coding and in some cases were due to changes in the indexes which did not directly match the occupation but led to closer matches being identified for certain groups.

Coding of performance analysts was raised by the sector in provider feedback. This job title has a direct match in the indexes under Data Analysts (3544), which is in the minor group Business Associate Professionals, and is the correct code to apply in some cases. However, sports related performance analyst roles are more appropriately matched to Sports Coaches, Instructors and Officials (3432). This is particularly relevant when graduates are working to enhance or monitor performance and is in line with the task description in the SOC index 'monitors and analyses technique and performance and determines how future improvements can be made'.

Due to the presence of a direct, unique match between the job title in the data and the indexes, and certain broad tasks arguably being loosely applicable to some of the sports related performance analyst's duties, they had been coded as Data Analysts and there were no performance analysts assigned to Sports Coaches, Instructors and Officials. It is worth noting that 'Performance Analyst' is not a job title listed under the unit group Sports Coaches, Instructors and Officials in the index. However, once the occupation group was analysed, it became clear that 3432 was a more appropriate code in some cases and that there was a systemic issue in the coding of sports related performance analysts. This was also clear when comparing the other occupation groups assigned to the two codes. The group was raised with Oblong, with guidance to consider the specific duties of 3432 when coding this occupation. Some examples of the duties relevant to this systemic issue are below:

- "Analyse football matches, film games and analyse footage, report back to coaches to help improve performance."
- "Performance analysis. Performance coaching. Coach development. Filming and analytics."

Examples of other performance analyst duties that remained in Data Analysts (3544) are provided below:
- "Monitor compliance and performance. This involves database development, dashboard development and statistical analysis."
- "Analyse data, give feedback to stake holders, monitor performance, evaluation, training for new recruits, presenting new ideas, analysing existing products."

3.2 Inconsistent: Energy Adviser

Occupations can be marked as inconsistent for a number of reasons. Commonly with inconsistent issues an occupation group is split across two or more valid codes depending on job duties, and some records have not been coded into the most appropriate code and therefore need to be moved from one valid code to another.

Energy advisers are an example of an inconsistent issue that was identified in HE provider feedback. Energy advisers are generally coded into Estimators, Valuers and Assessors (3541) if duties are appropriate or Sales Related Occupations N.E.C. (7129) in some cases if the role is predominantly a sales adviser role. Some graduates with similar duties to those in 3541 had been coded into 7129, where 3541 was determined to be the more appropriate code. The issue was raised with Oblong, who amended the occupation group as appropriate.

3.3 Non-systemic: Creative Decorator

As expected, many of the non-systemic issues found in HE provider feedback were fixed in Oblong's consistency checks. Where they weren't, the fact that they were non-systemic meant that no further action was taken.

An instance that was raised in HE provider feedback and was determined to be non-systemic was a graduate with the job title of creative decorator. The graduate had been coded into Painters and decorators (5323), however, the graduate's job duties described that their job was 'making wreaths'. The HE provider had suggested a code of Design Occupations N.E.C. (3429), however upon an assessment of the indexes it was determined that the graduate should have been assigned a code of Florists (5443), which according to the index includes the task 'designs and makes up wreaths, bouquets, posies, corsages, headdresses and button holes using appropriate flowers, foliage, frame and trimmings'. Although this was the only instance of a 'creative decorator', the role was assessed against graduates with 'decorator' in their job title to ensure there were no systemic or inconsistent issues across the entire occupation group and it was determined that this was a non-systemic instance of miscoding.

3.4 Not an Issue: Ecologist

Many instances raised were determined to have no issues. It is worth noting that some of these are due to mismatches resulting from two sets of codes when graduates have two instances of employment.

Ecologists, assistant ecologists, and graduate ecologists were raised in HE provider feedback and were determined to have no issues with coding. All records were correctly assigned to the group Conservation Professionals (2151) which includes the job title ecologist in the indexes. It was requested that these be moved to Environment professionals (2152), which has many similarities, but having considered the presence of the job title in the indexes, as well as the job duties of the graduates, it was determined that Conservation Professionals was the best fit.
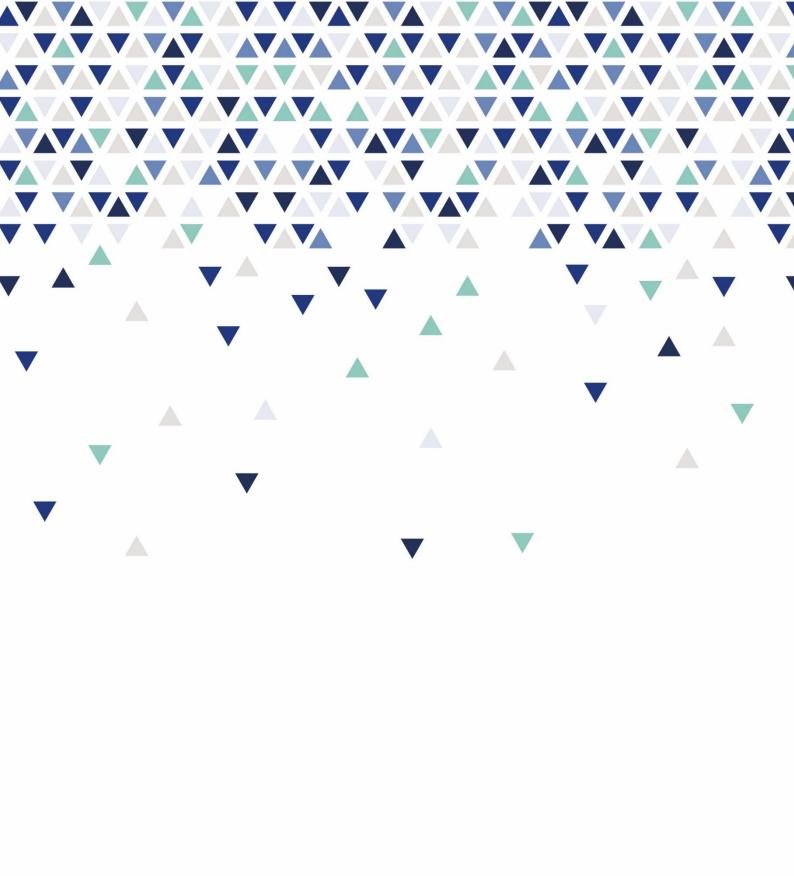
4 FEEDBACK TO HE PROVIDERS AND COMMON MISUNDERSTANDINGS

We have again included a section in this report to feed back to HE providers about the process of receiving their coding queries. We hope that this continues to be useful in communicating why certain decisions are made and promotes continuous improvement.

- The feedback provided to HESA via the template (available on the HESA website) should be for the relevant survey year only. All other queries should be raised separately and will be used to inform the coding of the next survey year.
- We ask that providers do not provide uncodable (0001) records, as these will be checked automatically at the end of the collection by HESA and Oblong.
- Graduates can have two distinct SOC codes, one for employment and the other for self-employment. Please provide the relevant job title and corresponding SOC code when submitting queries.
- We need sufficient evidence for a code to be changed, including reference to the SOC 2020 coding framework, as mentioned above. Please consider these additional points when explaining the reason why the change is requested:
  - Salary should not be the only justification for a change in code as this variable cannot be used reliably and consistently to determine SOC codes.
  - A degree being required for a role should not be the only justification for a change in code as it is not a reliable indicator of SOC according to the framework.
  - Employer name should not be the only justification for a change in code, job title and/or duties are the primary determinants.
  - Please refer to coding rules published by ONS to explain why some records have been coded in a certain way. For example, Assistant auditor will be coded as Auditor, but Auditor's assistant is an assistant role.
  - Information collected from sources outside the survey, such as a graduate's social media platforms, cannot be used as justification for a change.

5 NEXT STEPS

An assessment of the quality assurance process for year four (C20072) will be undertaken in the coming months, and the sector will be informed of the outcomes once this is complete. This will include details about our approach on provider feedback and enhanced internal quality assurance.

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