UKSCQA
UK Standing Committee for Quality Assessment

DEGREE ALGORITHM PRACTICE IN 2020:

RESEARCH REPORT

Universities UK

Distinction and Diversity in Higher Education
GuildHE

QAA

CONTENTS

EXECUTIVE SUMMARY

In May 2019, the UK Standing Committee for Quality Assessment (UKSCQA) published a statement of intent on degree classification in which it reaffirmed the sector's commitment to fairness, transparency and reliability in academic standards.

To support the sector in delivering against this statement, Universities UK (UUK), GuildHE, and the Quality Assurance Agency for Higher Education (QAA) have worked together on behalf of the UKSCQA to understand in more detail the practices and methods used in the UK higher education sector to classify degrees, and the principles and rationales driving these approaches. This report provides an overview of the current review processes and sector practice in algorithm design, based on survey data and workshop discussions with providers across the UK. It offers a basis for new recommendations and thinking on effective models of algorithm design.

KEY FINDING

The main findings of the sector-wide survey were as follows:

- **58% of respondents** were either currently conducting a review of their degree algorithms or had conducted a review since the academic year 2017–18
- During algorithm reviews, the area most frequently changed by institutions was the approach to **borderline policies**, followed by weighting, re-assessment and rounding procedures.
- For **73% of respondents** the method of calculating a degree classification was through a weighted arithmetic mean of percentage marks.
- Approximately **30% of institutions** ran individual students' marks through multiple algorithms for the purpose of calculating their degree classification, awarding the higher overall mark or classification returned by the different algorithms.
- Approximately **a third of respondents** had some form of discounting policy, in which not all marks were used in the calculation of the degree classification, although credits (for passing the module) would still be required for an award.
- The most common weighting that institutions used, (**in three quarters of cases)** showed an emphasis towards exit velocity, placing greater weight on the final year of study.
- **85% of institutions** had a borderline policy in which a student's classification is reviewed with the potential to be increased in accordance with set criteria and/or exam board approval.
- **91%** of institutions published their algorithms online although only eight institutions supplemented their published algorithm guidance with an explanatory rationale for the use of the particular algorithm.

The findings show that there are many common practices and approaches across the sector. However, some differences remain. This research report draws on engagement with the sector to understand where and when different approaches may be necessary and where there is scope for greater convergence and simplification.

1: CONTEXT AND METHOD

CONTEXT

In 2017, UUK and GuildHE published the report *Understanding degree algorithms*, which for the first time informed both the sector and policymakers on how algorithms are composed, the decisions and elements involved in their use, and how they can vary within and between institutions.[1] While principles of algorithm design do not feature explicitly in the 2018 Quality Code as part of the continued commitment to quality assurance and maintaining rigour in academic standards, algorithms have been identified as an area of interest within the context of the increasing proportion of upper degrees. As research for the UKSCQA (2018) showed, between 2007–08 and 2016–17, the proportion of upper degrees (first and upper-second awards) increased from 61% to 75%. In 2017–18 and 2018–19, the proportion was 76% (HESA, 2020).

While acknowledging that improvements and investments in teaching and learning have played a part in this, alongside students' continued hard work, concerns have been voiced about 'unexplained' increases and potential 'grade inflation' (Bachan, 2018; OfS, 2019). In response, through the UKSCQA, the sector made a commitment in a *statement of intent* in May 2019 to protect the transparency, reliability and fairness of the degree classification. Further research into the impact of different algorithms on degree classification was one of four central elements of the statement. Analysis by David Allen (2018a, 2018b) had demonstrated how different approaches in discounting and weighting could have a sizeable effect on the proportion of students gaining an upper degree. Further research by Allen (2020) highlights the diversity of sector practice and finds that students with the same set of marks can receive a different classification depending on the institution they attended. Allen (2020) argues that this risks inhibiting comparative analysis of degree outcomes and accurate measurement of attainment gaps.

The statement of intent therefore outlined that UK higher education institutions will continue to protect the value of degrees by reviewing and explaining how their process for calculating final classifications:

- fully reflects student attainment against learning criteria
- protects the integrity of classification boundary conventions
- maintains comparability of qualifications in the sector and over time

The **UKSCQA** proposed to undertake new research to map current approaches and decisions in algorithm practice and design and to assess progress and change since the 2017 report. The work featured in this report was undertaken in 2020 by **UUK**, **GuildHE** and the **QAA**. This report is the first output of this research and provides a descriptive picture of the methods used in UK higher education to classify degrees in 2020 and the decisions behind them.

1. Undertaken on behalf of the UKSCQA and funded by the Higher Education Funding Council for England (Hefce).

METHOD

Institutions with degree awarding powers were surveyed between January 16 and February 21, 2020.

In total, 69 responses were received (57 from institutions based in England, seven in Scotland, four in Wales, one in Northern Ireland). There was a relatively proportional mix of responses from institutions of different sizes. This is detailed further in Table 1 and shows a range of providers, some with subject specialisms, some with a large portfolio of courses, and others that validate undergraduate programmes delivered by other providers.

TABLE 1: INSTITUTIONAL PROFILE OF SURVEY RESPONDENTS

NUMBER OF UNDERGRADUATE STUDENTS	NUMBER OF INSTITUTIONS
Under 5,000	13
Between 5,000 and 9,999	14
Between 10,000 and 14,999	16
Between 15,000 and 19,999	9
20,000 or more	14
NUMBER OF SUBJECTS BY JACS 3.0 CODES[2]	**NUMBER OF INSTITUTIONS**
0–24	23
25–49	10
50–74	19
75–99	1
100+	3
NUMBER OF UNDERGRADUATE PROGRAMMES VALIDATED FOR OTHER PROVIDERS	**NUMBER OF INSTITUTIONS**
0	23
1–24	24
25–49	6
50–74	3
75–99	2
100+	2

Four workshops were subsequently delivered in March 2020 to provide additional context for the survey responses and to indicate areas for further exploration. Sessions took place in Glasgow, Cardiff, Leeds and London, with a total of 127 people attending. Participants were from a range of seniority, bringing technical, operational and strategic knowledge of algorithm practice and design. Some of the typical staff positions that attended included academic registrars, heads of assessment and quality assurance officers. The workshops were designed to be interactive and practical, with sessions on the principles of algorithm design, testing rules and approaches and finally an exploratory session on future-proofing algorithms. These were discussed in the context of a presentation of a preliminary analysis of the survey.

2. JACS 3.0 principal subject codes were introduced by HESA in 2012/13.

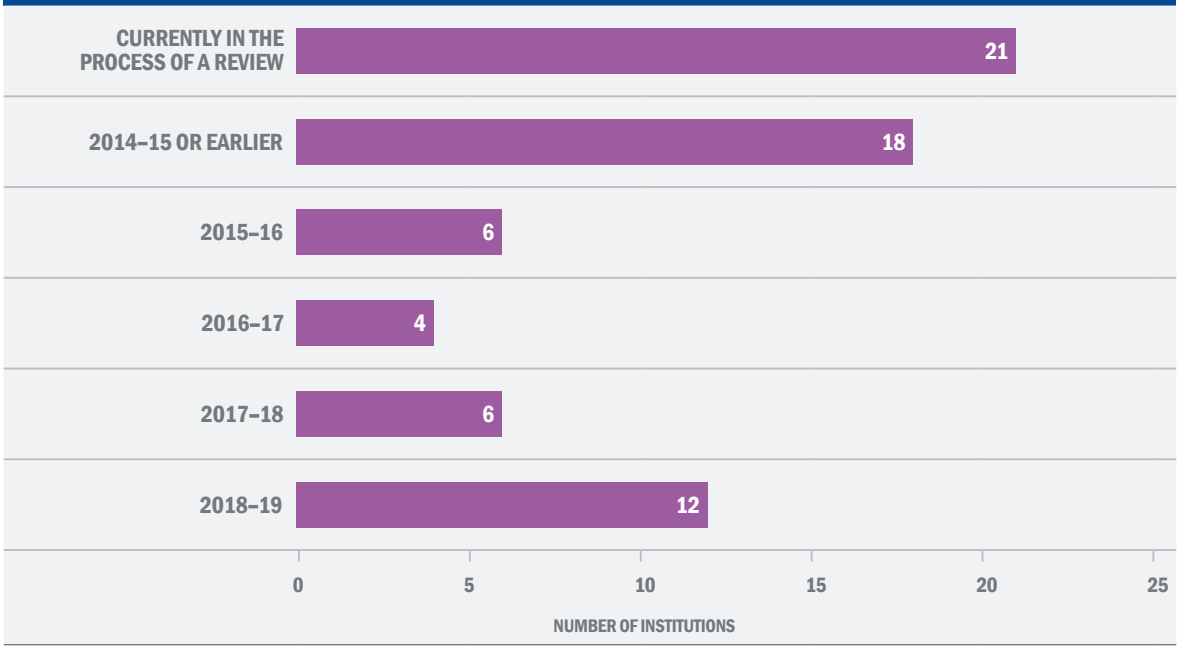
2: ALGORITHM REVIEWS

This section looks at the review processes that institutions take towards their algorithms. It covers when an algorithm was last reviewed, the reasons for the review and the outcomes of the review.

FREQUENCY

Institutions were asked in which academic year they last concluded a review of their algorithm award process. As **Figure 1** illustrates, 58% of respondents were either currently conducting a review or had done so from the academic year 2017–18. The year 2017 is a useful benchmark, as this was the publication year of the previous report *Understanding degree algorithms* (UUK, 2017): discussions with sector representatives at workshops and through interviews suggest that the report prompted many internal discussions. However, just over a quarter (27%) of providers had not conducted a review since at least 2014–15 or potentially earlier.

FIGURE 1: YEAR IN WHICH ALGORITHM AWARD PROCESS WAS LAST REVIEWED



SCOPE

The majority (59%) of reviews took place as part of a wider review of academic practices and regulations. The remaining 41% were reviews focused on degree classification algorithms.

The most regularly cited reasons for undertaking a review were said to be in response to:

- a regular and established cycle of reviews of regulations (n=20)
- UKSCQA's research on degree classifications and the statement of intent (n=17)
- the report *Understanding degree algorithms* (UUK, 2017) (n=16)
- internal concerns about institutional grade profiles (n=13)
- a recommendation or request from external examiners (n=11)
- wider modular or framework changes (n=9)

Among those consulted during the most recent review by an institution, out of the 62 responses, the most frequently mentioned were academic staff (n=61), non-academic staff (n=60) and student leaders (n=56). Consultation among the wider student body was more partial (n=22). Beyond external examiners (n=31) the remit of consultations often remained internal, with external stakeholders often not surveyed for feedback. Professional statutory and regulatory bodies (PSRBs) were formally consulted in 10 cases and employers were consulted in four. It was reported that PSRBs were not always concerned with classification, but would prioritise the requirements for teaching, assessment and an award.

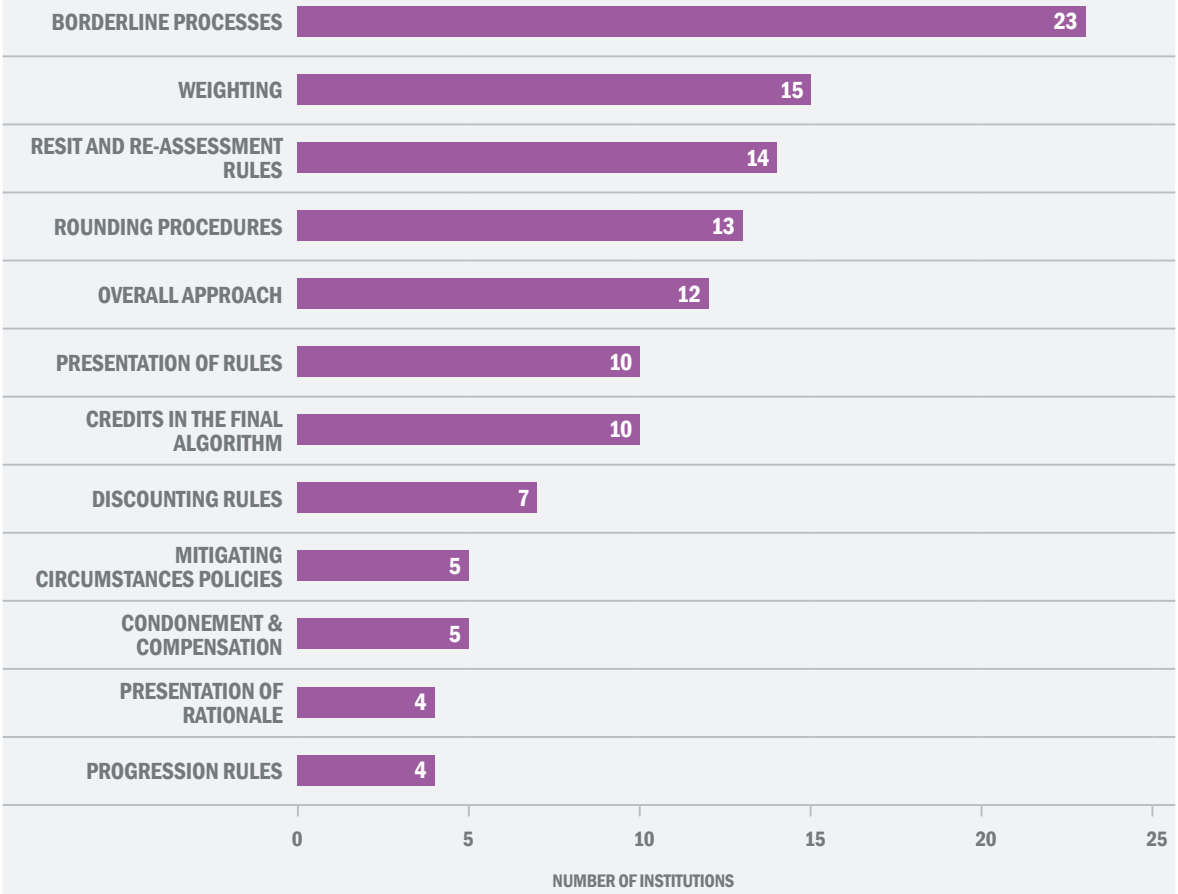
RECENT CHANGES

Following the latest review, institutions were asked whether any changes to classification practices had been made or were planned.

Over a quarter (29%) of respondents were currently considering changes, with many citing that reviews were being considered in the context of their institution's Degree Outcome Statement.[3] A further 22% had not made any changes during their last review. No relationship was identified between those providers that had concluded a review more recently and those that had made changes to their algorithm.

Figure 2 shows that for the remaining 49% that did make changes to their policies, this included a wide range of adjustments. The element of algorithm design that was changed most often by respondents was the method by which borderline policies were handled. Other prominent areas included modifying weighting, rounding and re-assessment rules. Twelve institutions indicated that their last review had led to a change in their overall algorithm approach. Of these, half moved to adopt an aggregate percentage of marks, four moved to a combination of aggregate marks and grade profile, and two moved to a grade-point average.

FIGURE 2: AREAS AFFECTED BY RECENT CHANGES TO ALGORITHM DESIGN



3. For providers in England and Wales, the Degree Outcome Statement is part of institutions' statement of intent activities, due for publication in 2019–20.

Further to this, respondents were asked what their reasons were for making changes in the areas shown in **Figure 2**. In general, there was a move towards the standardisation of practice across the institution (n=14), with many institutions stating a desire to create both parity for students and simplicity in design and implementation. There was also more broadly a move away from academic discretion, which was perceived to be more susceptible to inconsistencies (n=12). This trend towards a rules-based approach was also noted in UUK's 2017 report. Many institutions also made decisions to be in line with sector practice (n=11). The open comments and workshop discussions revealed that some institutions had looked at the 2017 survey results and realised they were in the minority of providers who had a particular practice. Respondents indicated that this sparked internal conversations within their institution, and while not exclusively the case, institutions that had recently achieved degree-awarding powers also reported a form of sector referencing by identifying similar institutions – in subject mix and student profile – to inform their chosen design.

In a smaller number of cases, other reasons for changing algorithms included to support gaps in attainment activity, to better reflect learner needs and, from a pedagogical perspective, to better classify learning outcomes.

MODELLING ALGORITHMS

Institutions overwhelmingly (96%) responded that they had modelled the impact that algorithm changes might have on their classification of undergraduate degrees. This took the form of quantitative analysis to establish whether proposed changes would make a material difference to overall results and the results of particular student groups. This often drew upon past years of student performance. Twelve institutions provided details of their modelling work against student outcomes. Six institutions expected a reduction in the number of upper degrees awarded of approximately 5 percentage points, with a more significant decrease in the number of first class degrees awarded. Five institutions expected the levels to remain as they were, and only one predicted a minor increase of an additional five students achieving upper awards as a result of the changes being implemented.

3: DEGREE AWARD CLASSIFICATION AND RULES

This section describes the approach institutions reported taking towards their degree classification. It also provides a breakdown of reported variation within an institution, and where individual students have multiple algorithms applied to their results.

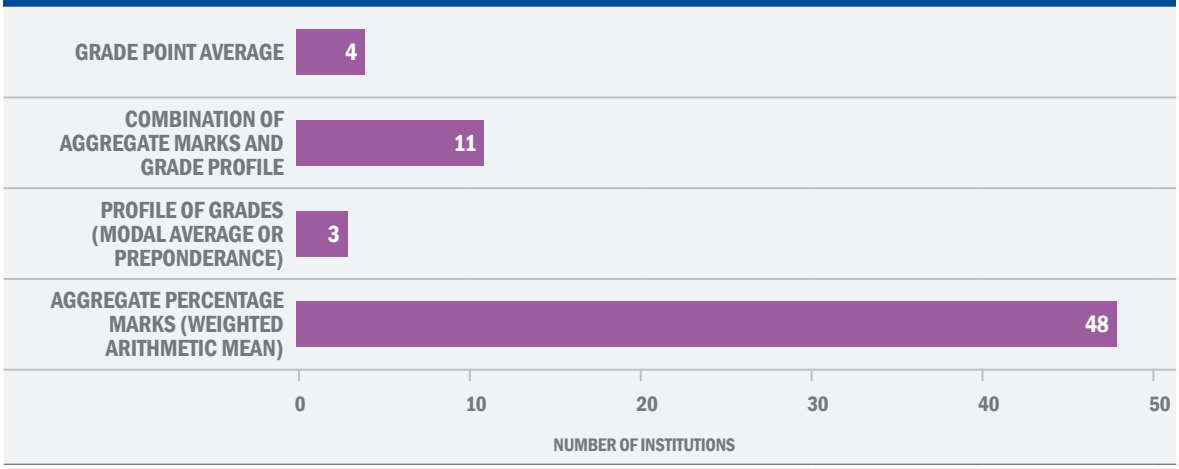
CLASSIFICATION PRACTICES

Institutions can take a number of different approaches to classify and calculate their degree awards.

This overall approach determines the framework within which related rules and policies operate. **Figure 3** details the overarching approach to classification practices among respondents.

- The most common approach to degree classification used by survey respondents was an aggregate percentage of marks calculated through a weighted arithmetic mean. This method was used by **48** (**73%**) of the respondents. The results of eligible modules are totalled and divided by the number of modules, producing a mean mark.
- Three institutions used grade profiles. This involves the use of the preponderance principle, which focuses on the classification awarded rather than the degree mark. It calculates the proportion of marks above a given boundary. For example, if a student achieved **45%** of their marks at a first-class and **55%** of their marks at an upper-second class, then the student's award classification would be an upper-second class degree.
- Eleven institutions used a combination of aggregate marks and grade profile. For example, to achieve an upper-second class degree, a student would need to achieve a weighted average mark of **60%**. Alternatively, they could also achieve a majority of credits above the grade boundary at the stipulated levels.
- Four institutions used a grade-point average (GPA). Under this approach, a student's marks are placed on a decimal scale, for example from **0 to 4.25**, based on an average of assessment grades.

FIGURE 3: APPROACH TO DEGREE CLASSIFICATION



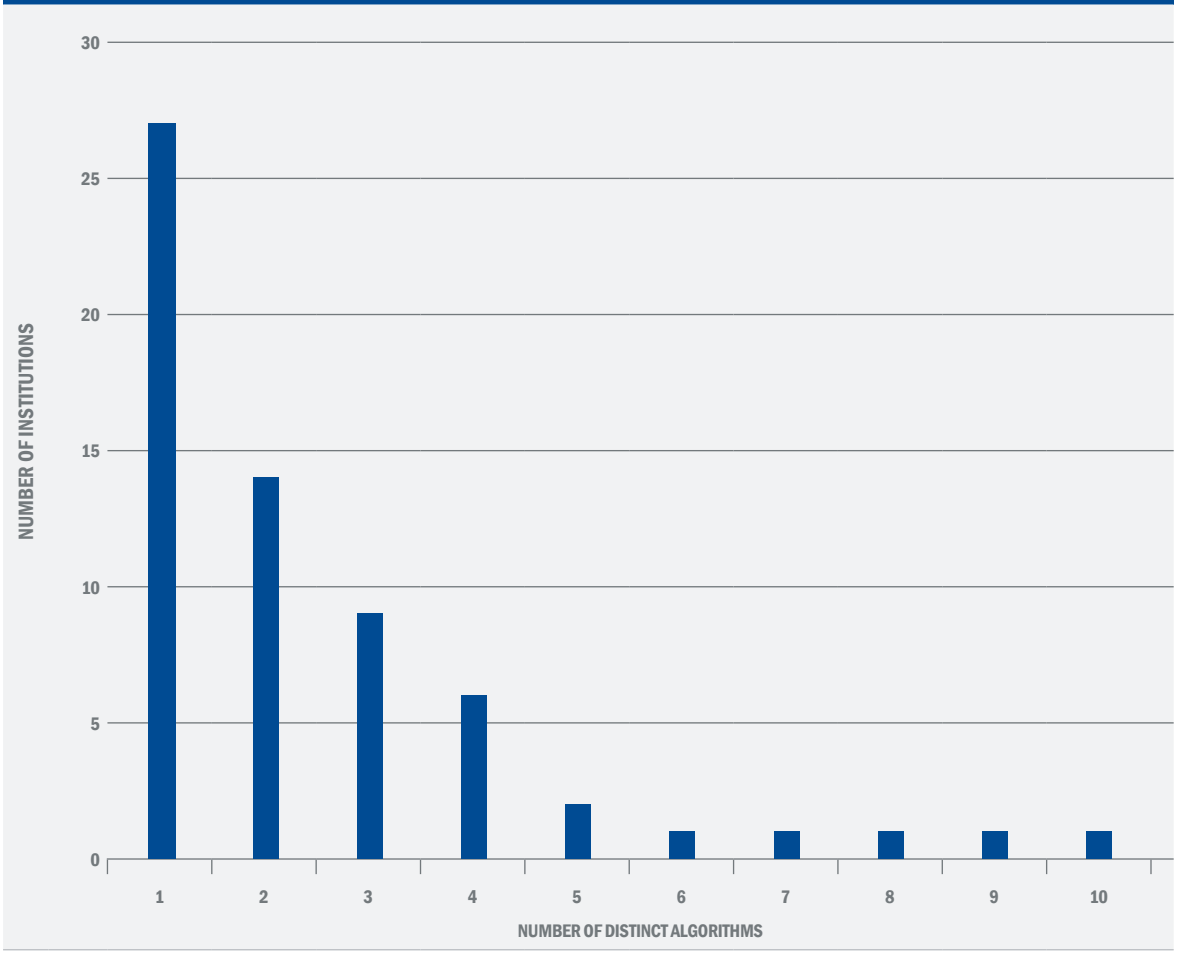
While not all module marks necessarily contribute to the final degree classification (for example where discounting is permitted), all institutions reported setting a standard pass mark that must be passed or condoned (ie fall within a specified range) in all credits required for an award. Across institutions, there was uniformity in setting the standard pass mark for degree awards at 40%. Variance from this was present in some institutions with integrated master's programmes, where the standard pass mark was raised to 50%, in line with many institutions' approach to postgraduate degrees. Most respondents (n=60) operated on a percentage-based marking and classification system; however, other approaches were adopted by the remaining institutions, corresponding to different approaches in marking practice.

In other cases, a numerical point scale was used, with scores ranging between 0 and 21. This practice often clustered classifications as sets of three numbers, for example, 19–21 for a first classification. Grade scales, distinct from other uses of scales in that they do not directly correspond to percentage points, were used in some cases. Finally, level descriptors provided narratives that referenced different competencies of award outcomes, for example, the Scottish Credit and Qualifications Framework.[4]

INSTITUTIONAL ALGORITHMS

Respondents were asked to consider the total number of algorithms and methods in place across all degree awards and validated programmes. Just under half (43%; n=27) of institutions used a single degree algorithm across all undergraduate programmes. Seven institutions reported using five or more algorithms across their degree programmes (see **Figure 4**).

FIGURE 4: NUMBER OF ALGORITHMS FOR AN INSTITUTION



4. For further information see https://scqf.org.uk/the-framework/about-the-framework/

For institutions with multiple algorithms or methods in place, these were often specific to programmes that sat outside a 'standard' fully class-based honours degree. For example, 20 respondents stated that integrated master's provision called for a distinct algorithm to account for the additional and higher level of study. Another reason cited in 17 responses was the stipulated requirements from PSRBs, which sometimes diverged from an institution's standard algorithm.

From the survey responses, 37% of a total of 152 distinct classification algorithms were reported as being partially or completely determined by PSRBs. Further discussions through workshops, however, indicated that while PSRBs held a particular interest in professional competence and the awarding of a degree, the classification element – whether the student receives a first or an upper-second, for example – was not always a priority.

Students who undertook time abroad during their study or a year in industry were also noted as requiring an algorithm to account for additional modes of assessment and years of study. This was reported in 10 cases.

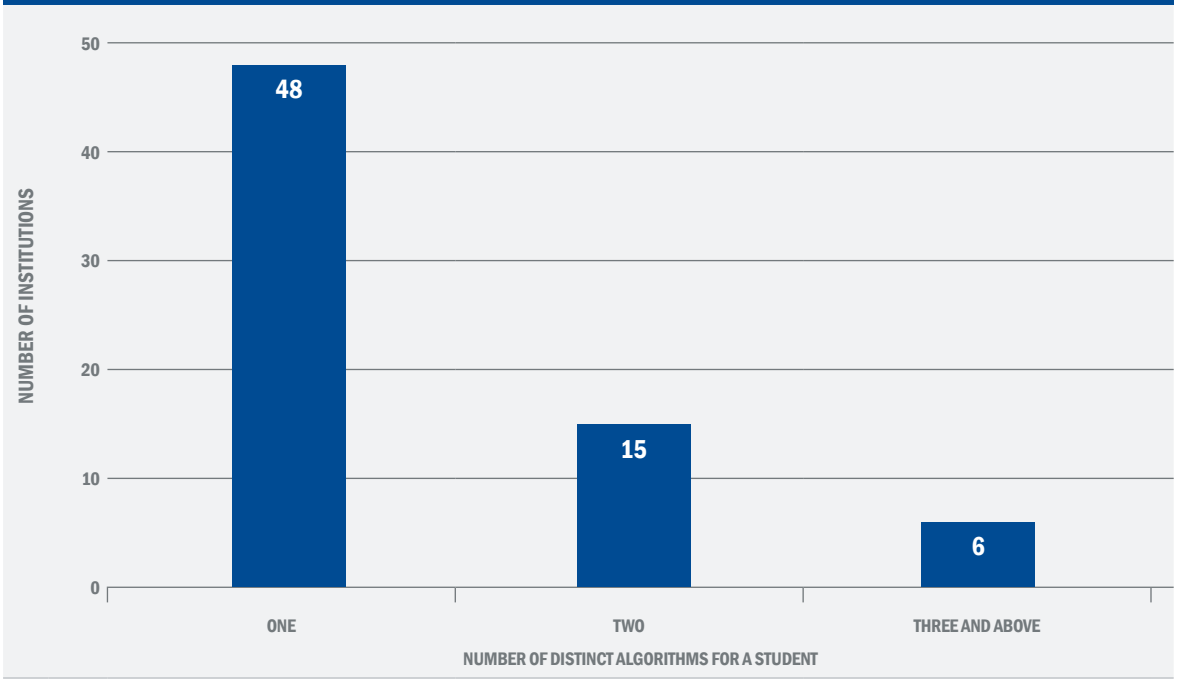
Other less common reasons from survey respondents included:

- to account for variations in methods of assessment across different programmes
- to acknowledge variations in intended learning outcomes across subjects and disciplines
- to offer provision for students who entered study either through a top-up course or with direct entry
- to match sector norms for specific subjects and disciplines that might differ from institutional practice
- to reflect historical conventions within disciplines

MULTIPLE ALGORITHMS FOR INDIVIDUAL STUDENTS

Institutions were asked what the maximum number of possible algorithms were for an individual student. This was explicitly in reference to the initial classification of an award before the further consideration of an exam board or borderline policy. **Figure 5** shows that 70% (n=48) of institutions applied a single algorithm for the calculation of individual degree classifications. Of the 21 that used multiple algorithms, 15 used two methods and six used three or more.

FIGURE 5: NUMBER OF DISTINCT ALGORITHMS FOR AN INDIVIDUAL



Where the calculation used multiple algorithms, the majority of approaches awarded the highest classification arising from the formulas.[5] The method of employing multiple algorithms to a student most often included applying different weightings between level 5/9[6] and level 6/10, commonly at a ratio of 1:1 and 1:2. Those with three algorithms followed a similar approach, in which the student's classification was calculated across three weightings. Institutions that operated with a combination of aggregate mean and modal grades as their principal approach interpreted the question with mixed responses. Many saw their combined approach as a singular algorithm in itself, while two noted that it could be interpreted as an additional methodological approach.

Some workshop participants suggested that having two algorithms could be useful to measure both consistency of performance and exit velocity in the final level of study. Notably, participants discussed how a second algorithm could pre-empt some of the policies covering discounting and borderline arrangements.

- Discounting often reduces the number of credits required at level 5/9. Some delegates noted that this serves a similar purpose to a second algorithm that gives extra weighting to the credits at level 6/10.
- Borderline rules and multiple algorithms often consider both consistency and the privileging of final-level grades. In some instances, borderline policies consider modal and mean averages, which is also a practice that some institutions use within multiple algorithms.

When change is made to an algorithm, it is likely to take many years to be seen in degree outcome awards. Survey respondents consistently did not alter a student's algorithm mid-way through a programme. The Competition and Markets Authority (CMA, 2015) guidance outlines how changes to a course as advertised may be in breach of consumer protection law, and that the timing, communication and scale of a change require careful consideration so as not to disadvantage a student. This is relevant to algorithm design, where students may behave differently depending on how they understand the classification of their degree. To facilitate changes in policy, some institutions implemented a transition phase, where both the new and old algorithms were applied. In such situations, the algorithm with the higher classification was awarded, since it was considered that no student should be detrimentally affected during the transition period.

Members noted there was potential for duplication, although the individual composition of policies in institutions often accounts for this. It was also felt that two algorithms had the potential to be confusing for students and could inhibit their ability to know 'where they are and where they need to go'. Finally, there were concerns about whether two algorithms for an individual student could clearly link to learning outcomes and pedagogy design.

5. A single exception to this arose in an instance where students were provided with a choice: either an algorithm that provided accreditation but a potentially lower classification, or one that did not achieve the accreditation requirements but did enable a potentially higher classification outcome.

6. Where levels are referenced in this report, the first number references levels for England, Wales and Northern Ireland, as detailed in the Framework for Higher Education Qualifications (FHEQ), while the second number references levels in Scotland, as detailed in the Scottish Credit and Qualifications Framework (SCQF).

4: MODULE RE-ASSESSMENT

The following section provides an overview of institutional approaches to the re-assessment of modules in relation to their degree-awarding policies.

Policies on mitigating circumstances are not included here since the outcomes depend on the nature of individual cases. Additionally, this section does not account for the practice of condonement, where a module failure is accepted as not disqualifying a student from the target award. This is about re-assessment within the degree classification.

To account for the different terms used across the sector – something that itself may require further investigation. These two definitions are provided for the purposes of this report:

- resit or re-assessment is the process of repeating a component of modular credit, such as a written exam or practical demonstration
- re-take or repeat has a broader meaning, referring to repeating an entire module (with all components), an entire year or a semester

All but two of the 68 responses to the question had a standard policy that currently allowed re-assessment opportunities. The two that did not allow re-assessments made exception only where it was required by a PSRB for the award of a final classification. Respondents were asked if there was a limit to the number of times a module or unit could be subject to re-assessment through resitting or repeating. As shown in **Table 2**, 95% of responses stated that there was a limit, with 31 institutions allowing one further attempt, 22 institutions two further attempts and eight institutions allowing three further attempts.

TABLE 2: NUMBER OF TIMES A MODULE CAN BE RE-ASSESSED

NUMBER OF RE-ASSESSMENTS	NUMBER OF INSTITUTIONS
No re-assessment	2
One	31
Two	22
Three	8
No Limit	3

The most common structure for re-assessment was as follows:

1. the initial assessment attempt
2. a resit opportunity of the assessment
3. a repeat opportunity of the whole module
4. a resit opportunity of the assessment in the repeated module

A majority (93%) of institutions allowed re-assessment for failed credits to take place in the final level of study. If a student resat or repeated a module, institutions applied rules to handle the 'new' mark within the algorithm (see **Table 3**). The rules centred on whether the new mark was capped or uncapped. A student whose mark was capped is normally only eligible for the pass mark of the module (most often 40%). Where a student resat a component of a module with multiple assessments, for example resitting one of two exams, 58% of institutions still capped the whole module mark. When a whole module was repeated, including all or the only assessment, 36% of institutions placed a cap on the module mark.

TABLE 3: ARE FIRST RE-ASSESSMENTS SUBJECT TO A CAP?

RE-ASSESSMENTS STAGE	NEW MARK RULE	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
Assessment Resit	Failed elements capped, affecting a proportion of module mark	24	35%
	Failed module capped, affecting entirety of module mark	40	58%
Module repeat	Module is capped	25	36%

Note: Table 3 shows the percentage of all respondents who selected each answer option (eg 35% refers to the number of respondents choosing this option from all 69 survey responses.)

Several institutions used re-assessment policies for credit purposes only, and the original mark was carried into the calculation of the classification.[7] This was to mitigate a resitting student having an advantage over other students' achievements on their first attempt, while still enabling them to receive the credit and demonstrate that they had met all the required learning outcomes for an award.

5: DEGREE CLASSIFICATION RULES

This section describes how rules are used in the calculation of a final degree classification.

The key areas identified were weightings, discounting and borderline provision, all of which are closely connected and have multiple interdependencies. This section details current practice in the sector. The workshop discussions have been integrated for additional context behind the reasons for variations in practice or the absence of particular policies.

WEIGHTINGS

Weighting refers to the proportion of academic levels that contribute to the final degree classification. Weighting policies can be applied to levels within an arithmetic mean or when assessing preponderance in a modal approach. Respondents were asked whether the weighting of an award varied across different programmes in their institution. A small majority (61%; n=42) of institutions stated that they have one algorithm with a single weighting applied across all programmes. A further five (7%) had different algorithms across programmes, but with the weighting within these always the same (the algorithm may vary on discounting or borderlines, for example). Finally, 32% (n=22) of institutions had different algorithms across programmes, which were composed of different weightings, among other things.

7. In one example, this was below the standard pass mark at 30. Further data was not present to consider sector-wide practice is this area.

Table 4 outlines the most common approach to weighting that institutions used. This does not capture all the weightings available, but outlines the most common or only weightings used by each provider. The use of levels is structured in reference to the Framework for Higher Education Qualifications (FHEQs) for England, Wales and Northern Ireland and comprises levels 4 to 6. In Scotland, institutions use four levels from 7 to 10, typically applying the algorithm to the final two levels of study, as in the rest of the UK. As there were no responses with a weighting at level 7 in Scotland, this has been excluded from **Table 4**. Additionally, the weightings have been categorised into several broad approaches.

TABLE 4: LEVEL OF STUDY WEIGHTINGS

	L4/L8	L5/L9	L6/L10	NUMBER OF INSTITUTIONS
Exit velocity	0	0	100	4
Emphasis on exit velocity	0	20	80	4
	0	25	75	10
	0	30	70	8
	0	33	67	11
	0	40	60	9
Level 5/9 and 6/10 split	0	50	50	6
Level 4/8 inclusion	10	30	60	4
	11	33	56	1
Even split	33	33	33	1

The first category of 'exit velocity' is underpinned by weighting all contributions to the final degree outcome from the final year.[8] This was prominent in programmes that often included a final performance, project or portfolio as the main form of assessment. The weighting therefore solely reflects the student's achievement at the end of their studies. While there was consensus that the final year should be more stretching, representing a higher level of learning and offering students the chance to demonstrate skills obtained, there were some views that the additional difficulty does not necessarily require an increased weighting too. Several participants also commented that there is a risk of the final level becoming 'too high stakes', and that their students, when consulted, favoured an algorithm over at least two levels.

The second category comprised three-quarters of the survey responses and covered an 'emphasis towards exit velocity'. This represents institutions with a weighting of between 20 and 40 at level 5/9 and between 60 and 80 at level 6/10. As reflected in the survey, workshop participants believed that this was the most suitable weighting for the majority of courses.

The reasons provided included that:

- it reflects the increased importance of level 6/10, while also attributing credits to the penultimate level to measure consistency
- it reflects the philosophy of algorithm design, which is concerned with measuring the value-added to students at the end of their period of study
- FHEQ level 6 and the Framework for Higher Education Qualifications Institutions in Scotland (FQHEIS) level 10 (QAA, 2019) convey increased performance criteria, which arguably justifies their additional weighting allocation

8. For part-time and flexible learning options, the final level of study may be completed over more than one year.

Six institutions provided equal weighting to levels 5/9 and 6/10. This weighting is used to measure consistency in performance across the final and penultimate level, viewing both levels as equally important for the final classification. While a higher weighting was not given in the final level, feedback suggested that the significance of the final level was already accounted for by virtue of its increased difficulty.

Notably, five respondents included a weighting at level 4/8 and just one used an even spread across three levels. In Scotland, where honours degrees typically comprise four years of study, no responding institution reported using level 7 marks to classify a degree. Removing level 4/8 meant that institutions could support student experimentation in module choice and provide transition support for students from disadvantaged backgrounds or who are returning to education. The main justification for the inclusion of level 4/8 was to incentivise student engagement. However, it was felt that other methods of engagement could be explored to support students; even when level 4/8 was not used for classification, students were still expected to pass the first year with individual marks included on transcripts.

Weighting variation

Workshop discussions on weightings also considered the level of variation in current practice and what the implications would be if there were less variation. Increased commonality in weightings was supported across the four workshops by the majority of participants. They noted that this could lead to greater transparency and comparability across institutions. Some participants also noted that changes to weightings might help to manage perceptions of grade inflation by reducing the number of variables in an algorithm that might be identified as being 'tweaked' to change the outcomes.

Where there was variation within the broad approach, for example in the emphasis on exit velocity, participants felt that the composition of that particular weighting was largely insignificant. This response from the survey captures the sentiment expressed by many at the workshops.

> "I am interested in the rationale for the Level 5/6 ratio: 40/60, 33/66, 30/70, 25/75 – I don't think anyone actually knows why they have chosen this precise ratio (and it mostly makes very little difference) **Survey respondent, small university, England**

Allen's (2018a) research evidences the potential impact that different weightings could have on the number of upper awards. The modelling, based on a sample of 211 real student profiles, shows that the higher a weighting is towards the final level, the more upper awards are produced. Where all credits are included at levels 5/9 and 6/10, a ratio of 40/60 compared to 20/80 produces just a 2-point difference in upper awards. This difference is slightly more pronounced when just the best 100 credits were used, with a difference of 4 points in upper awards.

DISCOUNTING

Minimum number of credits

Discounting is the practice of not counting module marks towards a final degree classification. Sector practice in this area can be understood by exploring the minimum number of credits that contribute towards the final degree classification at each level of study. If an institution weighted a level at zero, then in effect all of these credits would be discounted for the purposes of the degree classification, even if all have to be passed to gain the award. As level 4 (in England, Wales and Northern Ireland) and level 7/8 (in Scotland) were not counted in most instances, in effect this meant that most credit discounting took place at level 5/9 and level 6/10.

For 46 (68%) institutions, all 120 credits at each eligible level – those included in the classification algorithm – contributed towards the final award classification. In these 46 institutions, no credits were discounted. Ten institutions discounted 20 credits at each level, while other practice included discounting 15 or 30 credits at each level. This variation is in part determined by differences in curriculum and programme design, and different credit structures. A further nine institutions stated that their discounting arrangements varied depending on the programme and level of study.

Where there was variation at the level of study, this comprised the following practices:

- allowance of 20 credits to be discounted from either level 5/9 or level 6/10
- allowance of discounting at level 5/9 only and the full use of credits at level 6/10
- only discounting elective credits, as opposed to those that were attached to compulsory modules

The credit composition of some programmes may create barriers to implementing discounting in a uniform way. For example, a level may be composed of a few large, credit-bearing modules, each of more than 20 credits (the most common number of credits discounted). Discounting in these circumstances would require a higher proportion of credits to be removed from the award calculation. There was uncertainty about what a fair discounting policy would look like in these circumstances.

Application of discounting

For the third of institutions employing discounting procedures, once the composition of discounting across the levels had been accounted for, there was uniformity in the credits that were selected for discounting. The method consisted of retaining the 'best' credits based on student marks, rather than discounting both high and low marks.

Discounting was seen to have a role in encouraging students to take challenging, unfamiliar and innovative modules. Delegates at the workshops noted that even if students were unaware of discounting, it would still inadvertently support them by not punishing acts of risk. There were also suggestions that discounting allows students to have a 'bad day', thus recognising life events and so helping the classification to reflect the higher levels of achievement a student has demonstrated across other modules. However, other participants suggested that mitigating circumstances policies already account for any legitimate reasons for a student's unusually poor performance, before the algorithm is applied.

Views on discounting were mixed, reflecting their partial adoption in the sector. As with the other algorithm elements, participants commented that discounting cannot be separated from other policies. However, if a policy were to exist, participants were more amenable to discounting at the lower levels of study. Where it might be deemed appropriate, participants agreed that there is a need for a clear rationale and a limit to the number and type of credits subject to discounting.

Those that were not supportive or had concerns about discounting policies provided the following reasons:

- it penalises consistent performance and is not a reflection of attainment
- it could affect the value of the degree, for example, a classification may imply that a student is of a particular standard in all their modules. If they achieved a lower mark in a technical module which was subsequently discounted, a mismatch between an employer's expectations and a student's ability could damage the course and the institution's reputation
- it could contribute to a fall in engagement with students, who are aware that a module may not count towards their final classification
- it increases the algorithmic complexity and makes explanations and clarity for students more difficult
- current discounting practice is always at the lower end of marks and so has inflationary potential

The inflationary potential of discounting on award results has been considered by Allen (2018a). As credits are discounted from the lowest marks, this creates an increase in upper awards. In a 50/50 weighting with no discounting, Allen's modelling produced 69% upper awards. However, if 100 credits are used at each level, then 74% of classifications become upper awards. This was consistent across the other weightings, so for example a weighting of 20/80 initially produced 72% of marks with an upper award, whereas if just the best 100 credits are used, this increases to 79% (Allen, 2018a).

BORDERLINE

Borderline provision refers to the process by which student degree classifications are reviewed with the potential to be moved across a classification boundary.

This decision is often taken through another form of algorithm or an examination board or both. Borderline decisions at the classification level are taken after weighting, discounting and other algorithm policies have been applied, but are separate from academic appeals; the student will not necessarily know at what stage their classification was confirmed.

In the 2019 statement of intent, the sector was called on to 'protect the integrity of classification boundary conventions', namely the overarching convention of: first (=>70), upper-second (60–69), lower-second (50–59), 3rd (40–49), pass (35–40). In total, 58 of 67 respondents considered borderline cases. The main reason identified in favour of borderline provision was to account for marking outliers. Many members expressed the view that marking was not an exact measurement and so a level of tolerance was seen as suitable to account for this. This was particularly prominent where a percentage-based system was used with a high level of granularity between classification boundaries. Associated with this was the concern that in a mean calculation, the number of inputs was not sufficiently large to create an average that would not be influenced by outlier performances.

Several reasons against the use of borderline provision were produced during the workshops.

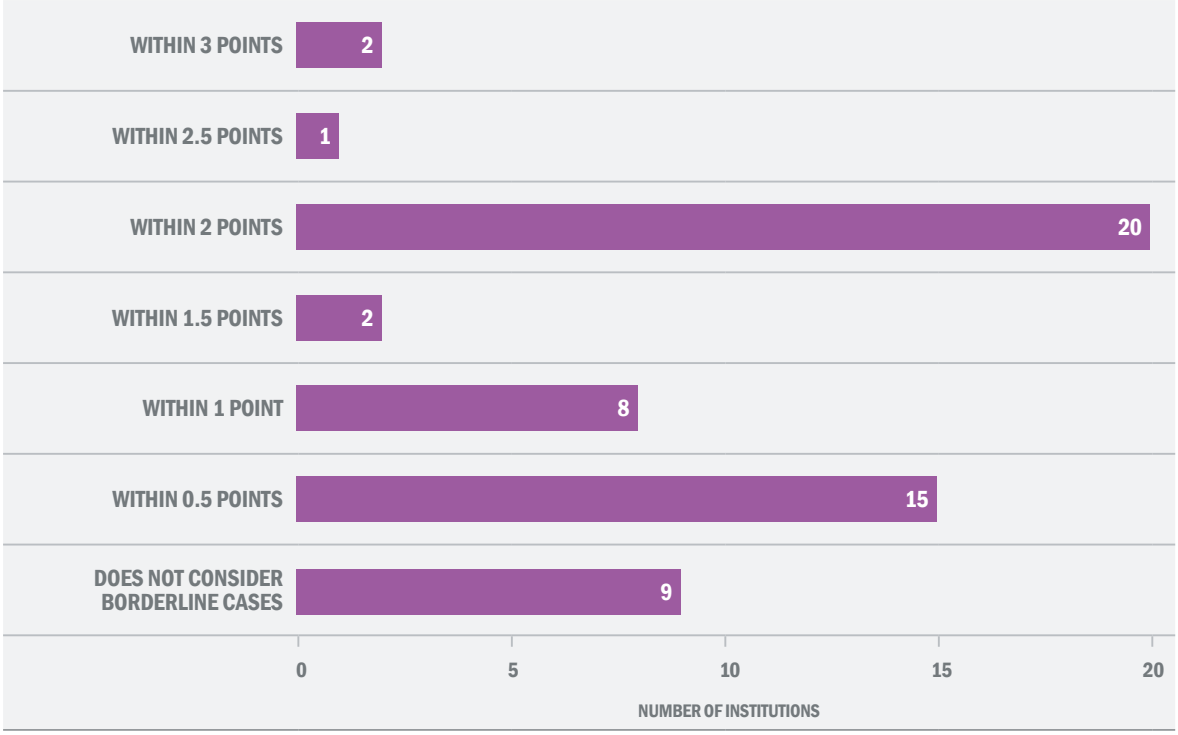
- Some members were concerned that borderline policies could re-write existing grade boundaries, so changing existing standards. This could create new standards of practice that artificially inflate outcomes because the boundaries are lower than first realised.
- There were also concerns that such policies could be unintentionally advantageous to students who are near to, but not working at a higher classification level.
- Workshop participants stated the need to consider all the factors that have been potentially favourable to students before the application of policies on borderline results. These include re-assessment, compensation, condonement, trailing, recognition of prior learning, second marking, discounting and rounding. The above notwithstanding, members commented that students have the opportunity to challenge a mark if they believe there has been a procedural error (as opposed to the challenge of academic judgement).
- It was noted that borderline policies were nearly always used to uplift classifications (for example, where a single poor performance has pulled a final mark down to just below the classification boundary), but never to downgrade a classification where a single high mark has pulled a student's average just above the classification boundary.

The overwhelming majority applied the same borderline arrangements to all students across their institution, irrespective of programme of study. Two institutions delegated responsibility for developing borderline policies to different departments of the university. For the majority of institutions, this took place in three stages, described below.

Stage 1: Eligibility

Stage 1 comprises the methods used by institutions to flag cases as eligible for borderline consideration. Institutions often stipulate the permissible extent of divergence from the grade boundary that qualifies a student for consideration. As shown in **Figure 6**, the two most common were an automatic rounding policy within 0.5 points (n=15) and an algorithm that flagged marks within 2 points of the grade boundary (n=20). The largest divergence from the boundary in the survey sample was within 3 points (n=2).

FIGURE 6: MAXIMUM DIVERGENCE FROM A GRADE BOUNDARY TO QUALIFY FOR BORDERLINE CONSIDERATION



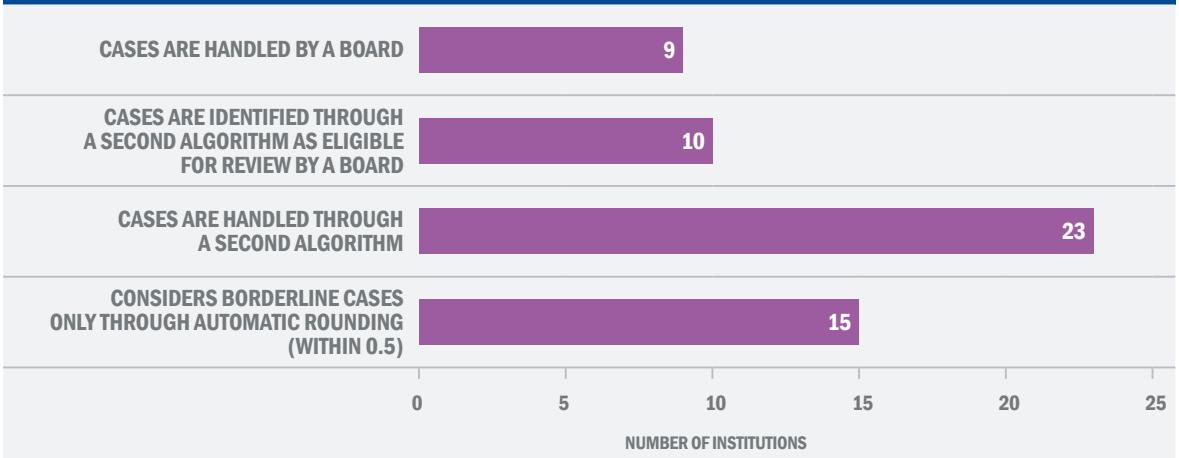
If a borderline policy were to exist, the majority said that they supported a limit on the maximum divergence of a mark to qualify for consideration. Most believed that 2% below the boundary would be an appropriate place for this. However, a 2% divergence is only relevant for numerically calculated algorithms: further work would need to consider what equivalent practice would look like for categorical and alpha-based marking schemes. Participants commented positively on the move away from academic discretion, noting that the use of additional algorithms helped to ensure consistency.

Workshop participants who used a points-based marking scale reported having fewer borderline cases. This marking process awards marks at staggered points, for example 62, 65, 68, with nothing in between. Further research would need to be conducted to confirm this effect and the reasons for it.

Stage 2: method

Stage 2 involves the method by which borderline cases are assessed. Most decisions by survey participants were made using an additional algorithm. The use of examination or assessment boards at some stage of the borderline process was present in 19 of 68 institutions who responded to the question (see **Figure 7**).

FIGURE 7: INSTITUTIONAL APPROACH TO BORDERLINE CASES



Stage 3: criteria

Stage 3 refers to the criteria used to make decisions concerning a student's degree classification. Where a second algorithm was used, institutions were evenly divided, with decisions based either on the proportion of credits awarded over the grade boundary (n=19) or the proportion of credits from the final level that were above the grade boundary (n=20). Three institutions looked at the mark awarded to a final dissertation, portfolio or performance to decide the borderline outcome. Finally, two institutions referred to the achievement of marks that were not included in the original calculation. This could involve referencing discounted or level 4/8 marks where these did not factor in the initial calculation.

Based on the workshop discussion, borderline rules often consider the highest level of achievement, while still showing a preference for some emphasis on exit velocity. However, it was acknowledged that this issue may have already been accounted for in discounting and weighting policies, so there was the potential for duplication of practice across these rules.

ROUNDING

A policy of automatic rounding within 0.5% of the boundary was supported by workshop participants and carried consensus as a fair and reasonable policy. Rounding also affects borderline rules, particularly in the event of a divergence that flags a case as eligible for consideration. For example, if there were a 2-point divergence, would this be set at 68% or 67.5%? Further research would be needed to establish which marks institutions use and at what point rounding is applied.

Currently, there is the opportunity for marks to be rounded twice in the calculation of a degree classification. Where a module comprises multiple assessments, the total module mark is likely to be decimalised and potentially rounded. However, rounding could also take place at the classification stage. Workshop participants agreed that rounding at multiple stages should be avoided, since this could have the accumulative effect of inflating a grade. Participants felt that it would be more appropriate to apply rounding only at the classification stage. However, there were technical limitations that would limit the adoption of this approach across the sector. Specifically, many student record systems were reported as requiring module marks to be entered as a whole number. This means that many institutions round at module level and the decimalised mark is not stored.

Feedback showed that participants felt it was important to exercise transparency in relation to how rounding is applied, for example in stating the number of decimal places used in automatic rounding. Many students will want all of their individual module marks to appear as whole numbers on their degree transcript. Providers need to be clear in their communications, since the classification is not necessarily based on whole numbers but rather more accurate ones. For example, the grade boundaries with rounding would be accurately displayed as 59.5% rather than 60%.

6: PRESENTATION AND PUBLICATION

Institutions were asked about the publication and presentation of their degree algorithms. Out of the institutions surveyed, 91% (n=63) published the technical rules of the algorithm(s) on their website.

The majority (80%) of those that did publish these rules located them within the website area of academic regulation rather than on student-facing web pages. This has implications for the CMA's guidance for institutions, which states: 'You should ensure that you draw students' attention to your terms and conditions, and any other rules and regulations, and make them accessible' (CMA, 2015:6). However, it remains unclear to what extent students are aware of the regulations to which they are subject.

Allen (2020) provides an overview of classification practices across the sector, albeit limited to information publicly available on institutions' websites. However, our survey reveals that only eight of our respondents supplemented their algorithm guidance with an explanation of why the algorithm was used. This is something we expect providers in England and Wales to be featuring in their degree outcome statements.

Half of respondents provided additional materials for students in the form of calculation guides, calculators and examples. One institution used a calculator that connected to its student record system, providing up-to-date information to students about their expected award outcomes. Student behaviour around an algorithm was a topic of conversation in the workshops. Participants noted that a clearer steer to students about their expected classification could result in increased effort among those near a borderline, although equally, it could lead to reduced effort among those in a mid-grade band. In these cases, the accuracy of the calculation assumes great importance, due to concerns that students may miscalculate their predicted award classification. Greater student awareness of how their award is calculated could potentially avoid appeals, as students know where they are.

The complexity of algorithms was something that affected staff as well. Some fed back that staff who joined institutions could find the adoption of new regulations difficult, even as specialists.

7: DISCUSSION

A significant theme of the workshop discussions and written feedback within the survey was the importance of the interaction between different elements of a degree algorithm.

A single principle, for example, might be achieved through different approaches, but if the approaches are used together, they may either exaggerate the effect or cancel each other out. The result will be that the core algorithm principle is not achieved. It is therefore crucial that elements of algorithm design are not studied in isolation, but in how they combine to most effectively and fairly classify a degree, in accordance with the intended learning outcomes.

Subjects of interest in both the survey and the workshops have several areas of significant crossover within algorithm practices, as described below.

ADDITIVE IMPORTANCE OF LEVEL 6/10

Many policies indicated the increased importance of level 6/10. When combined, there is the possibility that the accumulative impact of these policies creates an unintended 'high-stakes' assessment environment for students that places additional pressure on the final year.

This can manifest through:

- the weighting of either full exit velocity or emphasis on exit velocity
- a discounting policy that does not discount at the final level
- a borderline criterion that assesses achievement at the final level

However, it can also mean that a student who performs well in their final year is given a disproportionate advantage over a student who was a more consistent performer across their penultimate and final years. The former's success at level 6/10 is allowed to boost their final classification mark multiple times.

MULTIPLE ALGORITHMS

Several instances were identified where two distinct algorithms were applied to individual students. Arguably, this causes duplication of practice and a potential lack of clarity for students, postgraduate admissions officers and employers. The use of multiple algorithms also holds inflationary potential since they often increase an award rather than providing an average or opting for the lower mark.

This could be created by the use of:

- multiple algorithm weightings and/or methods for individuals
- a combined algorithm with different calculation criteria
- an additional algorithm for borderline candidates

RE-ASSESSMENT AND DISCOUNTING

Re-assessment procedures have a high potential to overlap with the use of discounting policies. When a module or element is re-assessed, many institutions reported using the original mark or a capped mark for the purposes of calculating the degree classification. However, when combined with a discounting policy this can unintentionally remove these marks from the classification. While re-assessments are still necessary for the purpose of the degree award, the effect of their presence in a classification calculation can be greatly reduced through a discounting policy.

DESIGN EMPHASIS

Finally, throughout the analysis and engagement with the sector, the reasons for particular algorithmic design could broadly be divided into three categories: pedagogical, institutional and learner oriented. While every institution used a combination of these, their relative institutional importance may explain the variation in the sector.

- **Pedagogy**: an algorithm matches programme design, subject needs or PSRB requirements.
- **Institutional**: an algorithm reflects an institution's mission, context and regulations.
- **Learner**: an algorithm considers the learners' circumstances both in regard to supporting students and the clarity with which it can be communicated.

SECTOR-WIDE PRINCIPLES

The survey asked respondents to comment where there was a desire to explore particular areas in more detail from the perspective of sector-wide principles. The main areas that were mentioned were borderline, discounting and weighting arrangements. This report has explored these policies and their use in detail. A move to greater parity and comparability across institutions and programmes is considered beneficial for learners' and external understanding, as well as institutional processes. Additionally, robust standards should be maintained, with an algorithm measuring achievement in assessment, not potential.

The following algorithm principles were supported both in the workshops and the survey responses by most institutions.

To be effective, an algorithm must:

1. provide an appropriate and reliable summary of a student's performance against the learning outcomes, reflecting the design, delivery and structure of a degree programme
2. fairly reflect a student's performance without unduly over-emphasising particular aspects, with consideration being taken at the design stage of how each element within a method of classification interacts with other elements
3. protect academic standards by adhering to the current conventions and national reference points used to define classification bands and boundaries
4. normally be reviewed at least every five years – or alongside national cyclical review timetables – to ensure algorithms remain relevant and appropriate, with input from across the provider, including students, academic and non-academic staff, and accrediting bodies
5. be designed and reviewed in a way that is mindful of the impact of different calculation approaches to classification for different groups of students
6. be communicated and explained clearly to students, both in how it works and why

BIBLIOGRAPHY

Allen D (2018a) 'The use of differential weighting and discounting in degree algorithms and their impact on classification inflation and equity: A further analysis', *University of the West of England – Economics Working Paper Series* No. 1803. At: www1.uwe.ac.uk/bl/research/bcef/publications.aspx

Allen D (2018b) 'Degree algorithms, grade inflation and equity: the UK higher education sector'. *University of the West of England – Economics Working Paper Series* No. 1801. At: www1.uwe.ac.uk/bl/research/bcef/publications.aspx

Allen D (2020) 'What can be done about degree algorithm variations?', *Wonkhe*, 20 February. At: https://wonkhe.com/blogs/what-can-be-done-about-degree-algorithm-variations-2/

Bachan R (2018) *The drivers of degree classifications - UKSCQA-commissioned analysis*. At: https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2018/drivers-of-degree-classifications.pdf

CMA (2015) *UK Higher Education Providers: advice on consumer protection law*. At: https://www.gov.uk/government/publications/higher-education-consumer-law-advice-for-providers

HESA (2020) *Figure 16 – Percentage of first degree qualifiers obtaining each classification 2014/15 to 2018/19*. At: https://www.hesa.ac.uk/data-and-analysis/sb255/figure-16

OfS (2018) *Analysis of degree classifications over time: changes in graduate attainment*. At: https://www.officeforstudents.org.uk/media/af3bfe43-5f87-4749-9a8b-ea94f3755976/ofs2018_54.pdf

QAA (2018) *The revised UK Quality Code for Higher Education*. At: https://www.qaa.ac.uk/quality-code

QAA (2019) *Framework for Higher Education Qualifications: Annex D: Outcome classification descriptions for FHEQ Level 6 and FQHEIS Level 10 degrees*. At: https://www.qaa.ac.uk/quality-code/qualifications-and-credit-frameworks

UKSCQA (2018) *Degree Classification: Transparent Consistent and Fair Academic Standards*. At: https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/degree-classification.aspx

UKSCQA (2019) *Degree Classification: Transparency, Reliability and Fairness – A Statement of Intent*. At: https://ukscqa.org.uk/wp-content/uploads/2017/09/Statement-of-intent-FINAL.pdf

UUK (2017) *Understanding degree algorithms*. At: www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2017/understanding-degreealgorithms.pdf

ANNEXE: DATA FROM SURVEY AND WORKSHOPS

YEAR ALGORITHM WAS LAST REVIEWED	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
Currently in the process	21	31%
2018–19	12	18%
2017–18	6	9%
2016–17	4	6%
2015–16	6	9%
2014–15 or earlier	18	27%

CLASSIFICATION PRACTICES	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
Aggregate percentage marks	48	73%
Combination of aggregate marks and grade profile	11	17%
Grade Point Average	4	6%
Profile of grades	4	4%

MAXIMUM NUMBER OF DISTINCT ALGORITHMS FOR AN INDIVIDUAL	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
One	48	70%
Two	15	22%
Three and above	6	9%

NUMBER OF RE-ASSESSMENTS	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
No re-assessment	2	3%
One	31	47%
Two	22	33%
Three	8	12%
No limit	3	5%

WEIGHTING PRACTICE	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
One algorithm which is used across all programmes	42	61%
Different algorithms across programmes but the weighting is always the same	5	7%
Different algorithms across programmes which use different weighting	22	32%

Percentages may not total 100 due to rounding

MOST COMMON WEIGHTING APPROACH (L4/8, L5/9, L6/10)	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
0/0/100	4	7%
0/20/80	4	7%
0/25/75	10	17%
0/30/70	8	14%
0/33/67	11	19%
0/40/60	9	16%
0/50/50	6	10%
10/30/60	4	7%
11/33/56	1	2%
33/33/33	1	2%

MINIMUM NUMBER OF CREDITS AT EACH LEVEL	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
120	46	69%
105	1	1%
100	10	15%
90	1	1%
It varies between programmes	2	3%
It varies between levels	7	10%

BORDERLINE APPROACH	NUMBER OF INSTITUTIONS	PERCENTAGE OF INSTITUTIONS
Does not consider borderline cases at all	10	15%
Only through automatic rounding (within 0.5)	15	22%
Cases are handled through a second algorithm	23	34%
Cases are identified through a second algorithm as eligible for review by a board	10	15%
Cases are handled by a board	9	13%

Percentages may not total 100 due to rounding

GLOSSARY

Across the sector, there is variance in terminology when describing specific aspects of algorithm design. The definitions below apply for the use of these terms in this report.

Borderline – the process by which student degree classifications are reviewed with potential to be increased, a decision often taken using an algorithm or an examination board.

Compensation – a way of mitigating poor performance in a module or modules, where poor performance is offset by considering the score against satisfactory performance in other modules.

Condonement – institutional acceptance that the failure of a module does not disqualify the student from eligibility for the target award.

Discounting – the practice of not counting module marks towards a final degree classification.

Preponderance principle – a modal calculation of the number of marks given above a given threshold.

Professional Statutory Regulatory Body (PSRB) – organisations with responsibility for regulating the standards of entry into professions.

Resit or reassessment – the process of repeating a component of modular credit, such as a written exam or practical demonstration.

Re-take or repeat – the process of repeating all components of an entire module, or an entire year or a semester.

Trailing credits – core, compulsory credits that have not been passed at the first attempt or re-assessment and are carried over and repeated in the next academic year.

Weighting – the weighting of academic levels to contribute to the final degree classification.

Universities UK is the collective voice of 137 universities in England, Scotland, Wales and Northern Ireland. Our mission is to create the conditions for UK universities to be the best in the world; maximising their positive impact locally, nationally and globally. Universities UK acts on behalf of universities, represented by their heads of institution.

This publication and accompanying research was conducted by Universities UK, GuildHE and the Quality Assurance Agency on behalf of the UK Standing Committee for Quality Assessment. The UK Standing Committee for Quality Assessment provides sector-led oversight of higher education quality assessment arrangements that continue to be shared across the UK. The committee has members drawn from regulated providers in England and Wales, publicly-funded universities and colleges in Scotland and Northern Ireland, and providers currently designated for student support by the Secretary of State in England. Student interests are represented by both the National Union of Students and individual student members.Membership is also drawn from the four UK higher education funding/ regulatory bodies, sector bodies and regulatory partners.

Authored by:

Charlotte Snelling, Policy Manager

Rowan Fisher, Policy Researcher

Woburn House
20 Tavistock Square
London WC1H 9HQ

+44 (0)20 7419 4111

info@universitiesuk.ac.uk

universitiesuk.ac.uk

@UniversitiesUK

ISBN 978-1-84036-451-4

July 2020

UKSCQA UK Standing Committee for Quality Assessment

Universities UK

Distinction and Diversity in Higher Education GuildHE

QAA