The use of differential weighting and discounting in degree algorithms and their impact on classification inflation and equity: A further analysis

David O. Allen
University of the West of England, Bristol

Economics Working Paper Series
1803

The use of differential weighting and discounting in degree algorithms and their impact on classification inflation and equity: A further analysis

David O. Allen
University of the West of England, Bristol.

Version: 10/04/2018

This paper offers a critical analysis of how differential weighting and discounting in UK university degree algorithms are contributing to classification inflation. It shows that by their very nature the use of differential weighting and discounting favour one particular group of students whose mark attainment is relatively inconsistent. In this study, 20 different algorithms are applied to two sets of individual marks and those of a programme as a whole. In the case of the individual set of marks, the degree classification can be unaffected or can range from an upper second (62.25%) to a first (70.72%). For a programme including borderline adjustments the proportion of first class honours that could be awarded can range from 16% to 37%. In aggregate, 93 (44%) students in this sample (n = 211) receive a different degree classification depending on which of algorithm is applied or because of borderline adjustments.

Address for correspondence: David O. Allen, Bristol Business School, University of the West of England, Bristol, BS16 1QY, UK. Email: David.Allen@uwe.ac.uk

Acknowledgement: The author would like to thank Harvey Woolf for drawing their attention to the Student Assessment and Classifications Working Group (SACWG)'s 1997 findings, and to also thank Professor Don Webber (UWE) for their advice and guidance in this work

Background

The latest release of the Higher Education Statistics Agency (HESA) statistics show that 26% of graduates received first class degrees in 2016-17, compared to 18% in 2012/13. Similarly, the proportion who obtained a first or upper second has increased from 68% in 2012/13 to 75% in 2016/17. The media response has been predictable and varies from questioning the intrinsic value of current UK degrees, to calls for universities to halt grade inflation and restore public confidence in higher education.

In response, Andrew Wathey (2018), chair of the UK Standing Committee for Quality Assessment (UKSCQA) acknowledges that grade inflation is a long-standing phenomenon, “reaching back to the early 1990s at least”. He describes it difficult to disentangle genuine grade improvement from artificial grade inflation, but acknowledges that convergence of degree algorithms across the sector is one of the possible solutions.

Likewise, William Hammond (2018) from Universities UK (UUK) describes the challenge of distinguishing between improved attainment by grade improvement, rather than inflationary academic practices, noting that the latter “dominates the public narrative and needs to be addressed to maintain confidence in the sector.”

Concern about the heterogeneity in UK university degree algorithms is not new. The Consortium for Credit Accumulation and Transfer compiled information on degree award algorithms for a number of UK HE institutions and showed that there was a wide variation in algorithms across institutions (Armstrong, *et al*, 1998). Yorke *et al,* (2008) also found considerable variation across 35 UK universities and raised concerns on whether universities where abusing their cherished autonomy.

The ASKe/Weston Manor group and the Student Assessment and Classifications Working Group (SACWG) have also presented their own research (and that of others) to various common select committee meetings (see Memorandum 9 (2009) and 16 (2009) respectively, see also Parliament. House of Commons, 2008-09, pages 114-116). Both have made far reaching and forceful recommendations. In their submission the ASKe group stated; " ... if the UK degree is to continue to be seen as an academic benchmark standard, and to maintain its reputation internationally, and with employers, parents and students, there is a need for a complete "root and branch" change in assessment processes and practices".

Over this period, various Burgess reports (2004, 2007 and 2012), sponsored by UUK looked at how universities calculate and/or describe a student's final degree achievements, and questioned whether the traditional degree classification was fit for purpose. In its final report the UUK recommended that the classification should remain and complement the Higher Education Achievement Report which some commentators found disappointing.

This article seeks to disentangle the issues surrounding degree algorithms and to enhance our understanding of the issues. Section 1 explains just how degree algorithms work, as most lecturers (let alone students) do not understand their own institution's degree algorithm. The article also distinguishes between 'grade inflation' and 'classification inflation'.

Section 2 lists the algorithm used by 45 UK universities and illustrates the diversity in current practice. Section 3 applies 20 algorithms to two sets of individual students' marks and identifies two perspectives on potential equity: one from within a cohort of students and the other between universities.

Section four offers a literature review of those studies (notably Wolf and Turner 1997, and Yorke *et al*, 2008) that have attempted to measure the impact of this diversity in practice. This literature review provides the basis for section 5, which has three parts: (a) measures of the impact of different algorithms on the award profile of a programme, (b) estimates of the number of students likely to be affected by this diversity. It also illustrates what is actually happening to the distribution of degree marks when different algorithms are applied. Section 5(c) estimates the impact of borderline adjustments that earlier studies have not estimated.

Section 6 also has three sections: (a) explains why and how the degree outcomes for students can change (b) decomposes the impact of differential weighting and discounting for two group of different students, (c) offers a comparative analysis of the consistency in mark attainment between these two sets of students. This section is followed by brief concluding remarks and policy recommendations.

1: Understanding the mechanics (or theory) of classification inflation

Most UK degrees are made up of a number of modules and each module can have different credits e.g. 10, 15, 30, 20, 40, 60 etc. Degree programmes have different credit structures with some based on multiples of 15 credits (e.g. 15 /30 / 60) others based on multiples of 10 credits (e.g. 10 / 20 / 40). Typically, students study 120 credits a year or 360 credits in total. Translating module outcomes into a final degree classification (1st, U2, L2, 3rd)[1] starts by calculating the weighted average for each year of study used in the calculation.

In calculating the weighted average mark the percentage module marks are multiplied by their credits, these weighted values are then added together; finally, this total is divided by the value of credits use in the calculation. An example of a weighted average for one year of study is shown in [1] below.

Module 1 Module 2 Module 3 Module 4
((72% x 15) + (70% x 15) + (57% x 30) + (70% x 60)) /120 = 67.00% [1]

Discounting and classification inflation

Many universities use a ‘best of’ approach where the lowest module marks are discarded or discounted before the weighted average mark is calculated. Thus, in the case of the best 90 credits, 30 credits would be discounted per year, for the best 100 credits, 20 credits would be discounted per year.

The total credits used after discounting per year of study, varies across institutions but using the best 100 credits in each year is common across the sector, with only a few using the best 90 credits (see section 2). It is also the case that some universities apply different discounts to each year, for instance using all the best credits in year three (120 credit), and only the best 60 credits from year two (a total 180 credits are being used in the algorithm).

Discounting can require the splitting of credit value in order to achieve the desired total of credits. Using example [1] above, the weighted average for the best 100 credits and the best 90 credits are as follows:

<u>Best 100 credits</u> (where module 3 credits are reduced from 30 credits to 10 credits)
Module 1 Module 2 Module 3 Module 4
((72% x 15) + (70% x 15) + (57% x <u>10</u>) + (70% x 60)) /100 = 69.00% [2]

<u>Best 90 credits</u> (where module 3 is discarded)
Module 1 Module 2 Module 3 Module 4
((72% x 15) + (70% x 15) + (-----------) + (70% x 60)) / 90 = 70.33% [3]

The difference between those algorithms that discount and those that do not will become greater as the discounted module marks get lower (and vice versa), likewise, discounting can only improve the overall degree mark average. It follows that “If only the worst, outlying marks are omitted, it is possible that this would lead to grade inflation” (Universities UK-GuildHE, 2017, p.37, hereafter UUK-Guild), or, more correctly, *classification inflation*, in that the module grades are not being ‘inflated’ but instead the weighted average is.

[1] Traditionally, the decile boundaries for each classification are
3rd = 40% to 49%, L2 or 2:2 = 50% to 59%, U2 or 2:1 = 60% to 69% 1st = 70% and above

Differential weighting and classification inflation
Another source of classification inflation comes from which years are included in the algorithms (e.g. 1^{st}, 2^{nd} and 3^{rd} or only $2^{nd,}$ and 3^{rd}) and how these yearly averages are subsequently weighted to arrive at the degree classification mark. In terms of counting years, only a small proportion of UK universities use all years of study, most use years two and three, with a few using only the final year. The relative weight of each counting year can vary significantly; for example in the case of algorithms using only year two and year three marks, the range is from 40/60 to 20/80.

The usual justification for the higher weighting is that it captures the student's *exit velocity* or the standard that the student is performing at as they graduate from university. The notion of an exit velocity comes from the wide-spread belief that the student's marks generally improve from year two to year three.

In calculating the degree classification, each yearly average (however determined) is multiplied by its weight and the products added together. Below are two examples of how the degree classification is arrived at using year two (Y2) and three (Y3) marks:

Y2: (68% x 15) + (68% x 15) + (64% x 30) + (66% x 60) / 120 = 66.0% [4]
Y3: (74% x 15) + (70% x 15) + (64% x 30) + (74% x 60) / 120 = 71.0% [5]

Y2/Y3 Weighted **50:50** (equal weighting – no discounting)

(Y2) 66.0 x 0.5 = 33.0 [6]
(Y3) 71.0 x 0.5 = 35.5 [7]
Added together = **68.5** *Classification*: Upper second (U2 or 2:1) [8]

Y2/Y3 Weighted **20:80** (differential weighting – no discounting)

(Y2) 66.0 x 0.2 = 13.2 [9]
(Y3) 71.0 x 0.8 = 56.8 [10]
Added together = **70.0** *Classification*: First (1st) [11]

In this example the classification has increased, not because the module marks are artificially higher (i.e. grade inflation), but because there is a greater weighting on the year three marks. Applying differential weightings will not have a big effect on classification inflation when the year two and three marks are similar, but the combined effect after discounting can be large if marks are distinctly different across counting years. At an individual level, that one set of marks can result in two different degree classifications (in [8] or [11]) is counter to notions of equity.

Borderline marks and classification inflation
Most UK degree algorithms take the degree 'average' to either one or two decimal points e.g. 69.5% or 69.45%. This process generates borderline marks where the exam board has to determine what classification is awarded[2].

One typical example is a simple rule whereby marks equal to or less than 0.5% below a classification boundary are awarded the higher classification 'automatically' and confirmed

[2] A vivid description of the way exam boards deal with borderline marks comes from an opinion piece in the Guardian (16/07/2012) written by Jonathan Wolff (2012) - professor of philosophy at University College London

by the exam board, thus a 1st does not start at 70%, it starts at 69.5%. This lowering of the threshold is in effect devaluing the 'face value' of the degree classifications boundaries and can only lead to classification inflation.

In addition, most universities will grant an uplift in classification using the *preponderance principle* which looks at the proportion of marks obtained by the student in each classification band. For example, degree marks within a given range below a classification boundary (e.g. 68.5% - 69.49%) might be granted uplift in classification (from an upper second to a first) if the student has 60 credits in the higher boundary in their final year. In [8] the mark of 68.5% might be classed as a first because there are 60 credits in the first class boundary (70 %+) at year 3 (see [5]).

Alternatively, the requirement might be that the student has 120 credits (plus) in the higher boundary across all years of study used in the calculation (e.g. year two and three). If applied to [8], the student would not see an uplift in their classification in that they only have 90 credits in total with marks above 70% (see [4] and [5]). The complexity and diversity of these borderline rules cannot be understated, suffice to say these adjustments can have a significant impact on individual student's life chances, classification inflation and the award profile for a given programme.

2: UK degree algorithms in practice

Table 1 below offers a selection of the algorithms used by 42 UK universities, ranging from using all years of study with each year equally weighted (No.1), to using only the best 90 credits from years two and three (No. 20). There are no universities currently identified against algorithms 15 to 18, but they are included for completeness and later analysis.

Table 1: A selection of university algorithms (undergraduate, non-clinical awards)

ALL CREDITS AND ALL YEARS USED			
	Years Used	**Weights**	**UNIVERSITY**
1	Y1/Y2/Y3	EVEN	**London School of Economics**
2	Y1/Y2/Y3	10/30/60	**Warwick, QU Belfast, Queen Mary**
NO DISCOUNTING ALL CREDITS USED [240 in total]			
3	Y2/Y3	50 - 50	**Oxford Brookes, Newcastle, Bishop Grosseteste**
4	Y2/Y3	40 - 60	**UAE, Kent, York, Leicester, Essex, Durham**
5	Y2/Y3	33 - 67	**Reading, Exeter, Manchester Uni, Sheffield, Buckingham [New]**
6	Y2/Y3	30 - 70	**Nottingham Trent, Liverpool, Cardiff, Bournemouth, Harper Adams, Staffordshire**
7	Y2/Y3	25 - 75	**Aston, Manchester Met**
8	Y2/Y3	20 - 80	**Birmingham, Greenwich**
DISCOUNTED BEST 100 credits per level [200 credits in total]			
9	Y2/Y3	50 - 50	**Canterbury Christ Church, Portsmouth**
10	Y2/Y3	40 - 60	**Bath Spa, Northumbria***
11	Y2/Y3	33 - 67	**University of Wales Trinity Saint David***
12	Y2/Y3	30 - 70	**Cardiff Metropolitan, Southampton Solent**
13	Y2/Y3	25 - 75	**Arts University Bournemouth, Leeds Beckett, Sheffield Hallam**
14	Y2/Y3	20 - 80	**Bradford, Derby**
DISCOUNTED BEST 90 credits per level [180 credits in total]			
15	Y2/Y3	50 - 50	
16	Y2/Y3	40 - 60	
17	Y2/Y3	33 - 67	
18	Y2/Y3	30 - 70	
19	Y2/Y3	25 - 75	**Hertfordshire**
20	Y2/Y3	20 - 80	**East London**

Northumbria University and the University of Wales Trinity St David (both starred [*] in table 1) use more than one algorithm. In these universities (and others besides) the initial degree classification is based on the best 100 credits in year 3 and compared to the alternative calculation as shown in table 1(i.e. algorithms 10 and 11 respectively), the student's classification is based on which ever calculation has the highest mark. The justification for this two-pronged approach is summed up in Northumbria University's explanation to its students:

> "We use whichever benefits you most and to recognise 'exit velocity'– i.e. if you get better results in your final year (Level 6) your classification will be based on Level 6 marks alone" (Northumbria University, slide 2)

Coventry University also applies a similar multiple-rule approach, it too starts with the best 100 credits in year three, but the alternative calculation is based on the average of the best 220 credits across years two and three.

Finally, a number of universities employ a variety of algorithms *within* the institution, Bristol University is a case in point: the Faculty of Arts use a 40:60 weighting and its Engineering Faculty use a 25:75 weighting (year two and three respectively in both cases).

Table 1 is not an exhaustive list nor does it represent the actual distribution of algorithms across the sector. In this regard, the UUK-Guild (2017) project found that 84% of the 98 universities surveyed used year two and year three weightings in the range 50/50 to 20/80, although it was not clear which universities in this group combined differential weighting with discounting (algorithms 9 to 14 in table 1) .

3: Two perspectives on Equity and classification inflation

In section 1, equations [8] and [11] illustrated that differences in differential weighting can lead to different degree marks, which are further exaggerated by the application of discounting (see [2] and [3]). The potential for inequality comes from the purpose (or logic) underpinning any particular algorithm that involves discounting and/or differential weighting and the type of attainment behaviour it accommodates (either by design or by coincidence). In this context, attainment behaviour simply refers to the pattern of yearly marks, which can be generally consistent or inconsistent. In this respect, it is worth briefly revisiting these two practices.

Discounting

In the pure sense, the purpose of discounting would be to recognise consistent performance by removing outliers from either end of the range of performance, but as the UKK/GHE report notes "this does not appear to translate into practice" (p.37). In reality, discounting only the lowest marks accommodates those students whose marks in any one year (and for whatever reason) are not consistent in a *negative* way. Yet the student with consistent marks will not benefit from discounting.

Differential weighting and exit velocity

Likewise, higher weighting on the marks for the final year is a response to the notion of exit velocity. Put another way, exit velocity captures the notion that marks between year two and three are inconsistent but generally in a *positive* way. However, for the student whose yearly marks are consistent, differential weighting is of little consequence.

These two points are demonstrated in equations 12 to 21, which are applied to notional marks for Student X.

<u>*Best 100 credits*</u> (in each year module 3 credits are reduced from 30 credits to 10 credits)

	Module 1	Module 2	Module 3	Module 4		
Year 2	(68% x 15) +	(63% x 15) +	(61% x 30) +	(67% x 60) /120	= 65.13%	[12]
Year 2	(68% x 15) +	(63% x 15) +	(61% x <u>10</u>) +	(67% x 60) /<u>100</u>	= **65.95%**	[13]
Year 3	(65% x 15) +	(68% x 15) +	(66% x 30) +	(68% x 60) /120	= 67.13%	[14]
Year 3	(63% x 15) +	(68% x 15) +	(66% x <u>10</u>) +	(68% x 60) /<u>100</u>	= **67.35%**	[15]

Y2/Y3 Weighted **50:50** (equal weighting)

(Y2) 65.95 x 0.5 = 32.98 [16]
(Y3) 67.35 x 0.5 = 33.68 [17]
Added together = **66.95** Classification: Upper second (U2 or 2:1) [18]

Y2/Y3 Weighted **20:80** (differential weighting)

(Y2) 65.95 x 0.2 = 13.19 [19]
(Y3) 67.35 x 0.8 = 53.88 [20]
Added together = **67.07** Classification: Upper second (U2 or 2:1) [21]

Student X's module marks and years two and three average marks are very consistent. The initial difference between year two and three marks is 2.0 percentage points, which is reduced to 1.4 after discounting (equations 12 to 15). Discounting has made little difference to the net weighted average because the lowest mark in each year is consistent with all the other marks in that particular year. Furthermore, applying a differential weighting (equations 16 to 21) to these discounted marks also makes little difference to the degree classification because year two and three average marks are so similar.

We can test this supposition by applying different algorithms to two sets of marks. Table 2 compares two students [A] and [B] from the same university, studying the same course; the table includes all their marks and credits across all years of study and includes 'calculated' degree marks based on algorithms Nos. 1 to 20.

Both students demonstrate exit velocity; in particular, in year three Student A has improved on their year two average marks by 25.3 marks; however, their year two marks were much lower than their year one marks (50.0% compared to 63.5%). Student B on the other hand has seen their marks improve year on year where their year three average mark (69.6%) is 13.5 marks higher than the year two marks. Interestingly, if exit velocity referred to the difference between year one and year three Student B's mark attainment improved by 19 marks (69.6 – 50.6), which compares to Student A's improvement of 11.8 marks (75.3 – 63.5). In terms of consistency, for Student B the standard deviation of their marks (across all years of study and for years two and three only) is marginally lower than that for Student A.

Two Perspectives on equity

(i) Within a given cohort

The marks for students A and B are clearly not the same – but they are not substantially different. Whether one student is 'better' than the other is dependent on the criteria used. Here the algorithms distinguish between the two students. If their university was using algorithm No.1 (all years equally weighted), Student B would have the lower degree mark

(58.8%) because their year one average mark is lower compared to Student A. However, if the degree calculation is the unadjusted average of years two and three – algorithm No. 3, then both students have nearly identical degree marks (62.6% for A and 62.9% for B) with both achieving a 'low' U2.

The similarity in the calculated degree mark starts to disappear if their university uses any of the algorithms from No. 4 upwards. As a comparison, in algorithm No. 8 Student A's calculated degree mark is now 70.2% and they would be awarded a first, Student B on the other hand would see their mark only increasing to 66.9% thereby achieving a high upper second. Indeed, regardless of which algorithm is applied (excluding algorithm No. 1); the upper second classification would apply to all of Student B's calculated degree mark (subject to any borderline adjustments).

Table 2: Different algorithms applied to two students (A and B)

Module	CREDITS	A	B
Module 1	15	80	57
Module 2	30	54	48
Module 3	30	63	48
Module 4	15	65	52
Module 5	15	57	52
Module 6	15	72	52
Year 1 Average		*63.5*	*50.6*
Module 7	30	40	49
Module 8	30	47	64
Module 9	15	50	40
Module 10	15	56	52
Module 11	15	60	63
Module 12	15	60	68
Year 2 Average		*50.0*	*56.1*
Module 13	15	75	80
Module 14	15	71	70
Module 15	30	81	64
Module 16	30	74	67
Module 17	15	75	68
Module 18	15	71	77
Year 3 Average		*75.3*	*69.6*
Year 3 Credits > 70%		120	45
Standard Deviation of marks			
All years - based on No. 1		12.39	11.38
rs 2 & 3 - based on No. 3		10.28	8.32

ALL CREDITS AND ALL YEARS USED			Degree Mark	
Algorithm	Years Used	Weights	Student A	Student B
1	Y1/Y2/Y3	EVEN	62.9	58.8
2	Y1/Y2/Y3	10/30/60	66.5	63.7
[1] NO DISCOUNTING ALL CREDITS USED [240 in total]				
3	Y2/Y3	50 - 50	62.6	62.9
4	Y2/Y3	40 - 60	65.2	64.2
5	Y2/Y3	33 - 67	66.9	65.2
6	Y2/Y3	30 - 70	67.7	65.6
7	Y2/Y3	25 - 75	68.9	66.3
8	Y2/Y3	20 - 80	70.2	66.9
[2] DISCOUNTED [Best 100 per level - 200 credits in total]				
9	Y2/Y3	50 - 50	64.1	64.8
10	Y2/Y3	40 - 60	66.5	66.0
11	Y2/Y3	33 - 67	68.1	66.8
12	Y2/Y3	30 - 70	68.9	67.2
13	Y2/Y3	25 - 75	70.1	67.8
14	Y2/Y3	20 - 80	71.3	68.4
[3] DISCOUNTED [Best 90 per level - 180 credits in total]				
15	Y2/Y3	50 - 50	65.0	65.8
16	Y2/Y3	40 - 60	67.3	66.9
17	Y2/Y3	33 - 67	69.0	67.7
18	Y2/Y3	30 - 70	69.7	68.1
19	Y2/Y3	25 - 75	70.8	68.6
20	Y2/Y3	20 - 80	72.0	69.2

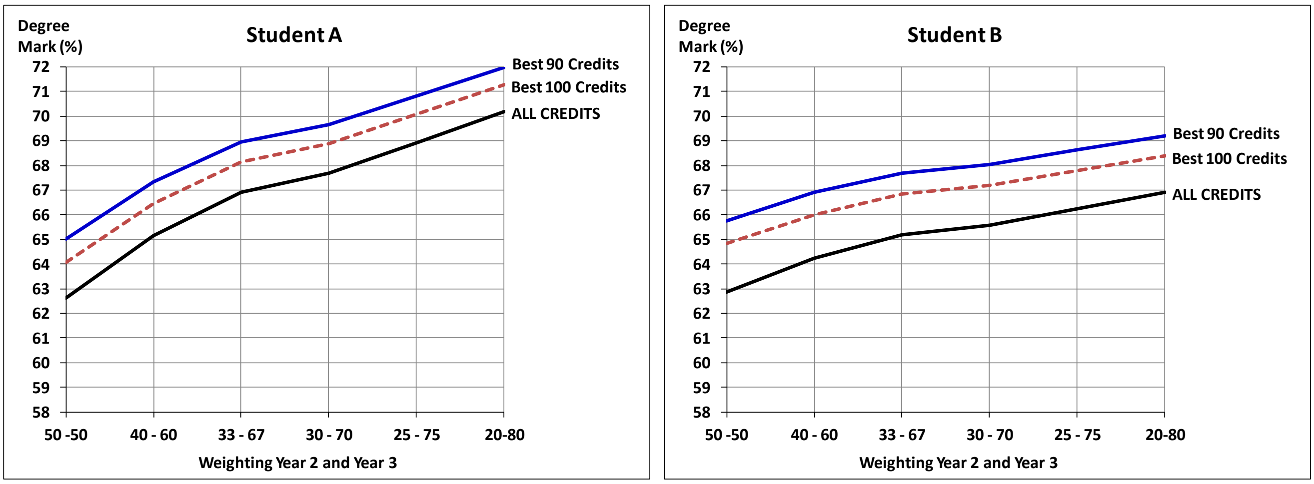
It is worth noting that given all their year three marks are equal to or greater than 70%, Student A would benefit most by the application of the preponderance principle in any borderline adjustment that could apply to algorithms Nos. 7, 11, 12, 17, and 18.

(ii) Between universities

An alternative perspective on potential inequality is to consider what classification a student would achieve had they gone to *another* university (all other things being equal). Thus, if Student A had undertaken their studies at Newcastle University, their efforts would be awarded with a low upper second with a mark of 62.6% (see table 1, algorithm No.3), however, had they gone to Hertfordshire University instead their efforts would have seen a first being awarded with a mark of 70.8% (see table 1 algorithm No.19).

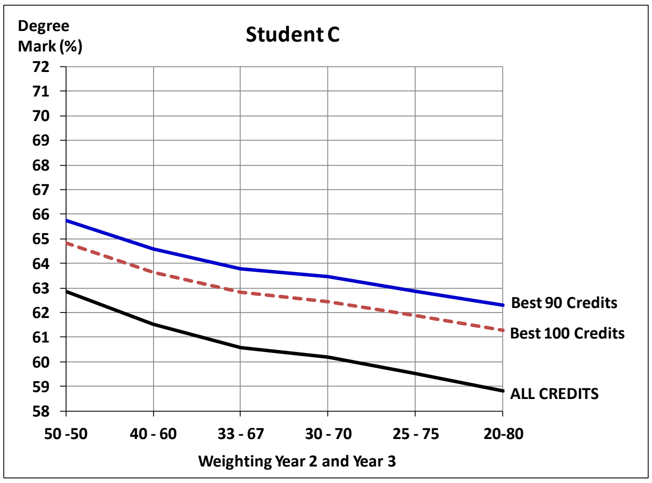
The differences in the calculated degree marks between Student A and B are illustrated in Figure 1 which plots the progression of degree marks for each set of algorithms (3 to 8 - All credits, 9 to14 – Best 100 credits and 15 to 20 – Best 90 credits) as the weighting for year three increases (from 50/50 to 20/80). The impact of differential weighting explains the slope of the lines that trace out the progression of the calculated degree marks, the bigger the difference between year two and three marks, the steeper the line. Discounting on the other hand shifts each series up as the number of lower module marks is increasingly discarded.

Figure 1: Degree mark under different algorithms for Student A and B



In addition, for the student whose marks in the final year are lower compared to those in their second year, differential weighting will see the overall calculated degree mark decline as we move through the algorithms. In figure 2 the year two and three module marks for Student B are switched around to create a possible mark profile for a notional student (Student C). Here the higher weighting on year three marks further penalises Student C's fall in academic attainment (the compassionate might wonder whether such students have suffered enough already).

Figure 2: Degree marks where year three marks are less than year two marks.



Based on table 1 and figure 1 the apparent inequality in potential degree outcomes comes from the manner in which differential weighting and discounting favours more one group of students, principally those whose module marks are relatively more inconsistent within years and across years. These potential differences are counter to notions of equity and fairness and for individual students the differences in degree algorithms across the sector could have profound implications for their life chances. Yorke *et al*, (2008) voice similar concerns about equity in their review of the assessment regulations across 35 UK universities. They note that, "Such variability, deriving from a system in which institutional autonomy is to the fore, raises a question about equitability in the treatment of students" (p 157). It also raises questions about the relationship between autonomy and the sector's wider responsibility to the society it serves.

4: Current research on the impact of different degree algorithms

Research in the use of different algorithms for the purpose of classifying degree outcomes is sparse. For the UK there have been only a few studies in the last twenty or so years where the principle source of this research is from members of the Student Assessment Classification Working Group (SACWC) based at Anglia Ruskin University.

One of the earliest studies is from Woolf and Turner (1997). The methodology employed is similar to that adopted here, namely the use of the student's actual marks, which are then "processed through the classification algorithms of other universities" to create "constructed classifications" (p.8). The sample covers students from five different universities, and it was the actual degree classifications awarded in the home institution that were compared to the constructed classification.

While the sample size is not discussed, their results indicated that 15% of the sample would have been given a different class of degree "if their results marks had been obtained elsewhere" (Woolf and Tuner, 1997, p.11). They go on to say:

> "Even if, for whatever reason, only 15% of students are potentially able to argue that they might have achieved a different class of at another institution, that represents around 30,000 of the 1995 honours graduates in the United Kingdom" (p.12).

Furthermore, most of the students affected by the difference in algorithms where classified as borderline by one or more of the institutions included in the sample. As result, Woolf and Turner (1997) conclude that criteria used for resolving borderlines were likely to be a "key factor in determining the differences the allocation of a class by different universities" (p.11).

A follow-on study by members of SACWG comes from Yorke *et al*, (2004) which uses marks from two 'new universities'; University A has 1,390 students and University B has 832 students. Where University A has a credit structure based on multiples of 15 credits, University B multiples of 20 credits.

The students in both universities study a range of subjects however these are not explicitly listed although ancillary analysis looks at the impact when the subjects are grouped using the Biglan (1973) hard/soft – applied/pure classification. The purpose of the Yorke *et al*, (2004) study is to measure the impact on the degree classification by:

(1) Varying the weightings on year two and three
[University A & B = 50/50, 40/60, 33/67, 30/70, 25/75, 20/80, 0/100]

(2) Progressively dropping (discounting) module marks from the degree algorithm.
[University A = all modules (240 credits) minus the lowest 15, 30, 45, 60 credits]
[University B = all modules (240 credits) minus the lowest 10, 20, 30, 40 credits]

There are three significant difference between the Yorke *et al,* (2004) analysis and the analysis conducted here, firstly effects of varying the weighting on years two and three and discounting are estimated separately. From table 1 it is likely that a significant proportion of UK universities do both simultaneously which has a greater impact on the variation in potential degree outcomes (see table 2). For this reason the analysis conducted here looks at a much wider range of algorithms progressing from those that only apply differential weighting (in table 2, algorithms Nos. 4 to 8) to those that combine this with discounting (in table 2, algorithms Nos. 9 to 20)

Secondly, the majority of UK algorithms that discount do so by year – e.g. the best 100 credits from year two and the best 100 credits from year three, which means each year's counting marks make an even contribution to the degree classification calculation. In Yorke *et al,* (2004) discounting 60 credits across all years of study could in some instances, result year two contributing only a third of the marks to the calculation if these lowest marks occurred in year two, for this reason the this study discounts evenly per year.

Lastly, in the Yorke *et al*, (2004) the impact of borderline decisions were not estimated but they are explored in this analysis (see section 5(c)).

Table 3 highlights the results from Yorke *et al*, (2004). They confirm that discarding modules with the lowest marks can only increase a student's degree classification, such that 32.7% of these students would see a change in their hypothetical classification. Increasing the weighting (from 50/50 to 0/100) had less of an effect with 21.3% of students seeing a change in their potential classification.

Importantly Yorke *et al,* (2004) also show that for some students, increasing the weighting on year three marks only serves to decrease the classification they could hypothetically receive; in this case, 4% of students would see a drop in their classification. Like Student C in figure 2, their year three average was lower than that achieved in year two (section 6 covers this effect in more detail).

Table 3: Results from Yorke *et al*, (2004)

[1] Discounting the lowest marks from 10 credits to 60 credits				
	Credits Discounted	Change in Classification	Number of Students	% of Students
University A	**up to 60 credits**	(increase)	**489**	**35.2**
University B	**up to 60 credits**	(increase)	**238**	**28.6**
		Total	**727**	**32.7**
[2] Increasing the weighting on year 3 from 50/50 to 0/100				
	Y2/Y3 weighting	Change in Classification	Number of Students	% of Students
University A	**0/100**	(increase)	**247**	**17.8**
		(decrease)	**62**	**4.5**
University B	**0/100**	(increase)	**138**	**16.6**
		(decrease)	**28**	**3.4**
		Total	**475**	**21.3**
A (n = 1,390)	**B** (n = 832)			
Adapted from Yorke et al, (2007) Tables 7a/b and Tables 8a/b pages 408 and 409				

Given the different methodologies, direct comparisons between Yorke *et al,* (2004) and Woolf, and Turner (1997) would be misleading, but it is likely that the proportion of students who could have achieved a higher classification from another university has increased between 1995 and 2004. Yorke *et al,* (2004) also note "that the notion of the honours degree classification is considerably less robust than its supporters would prefer" (p. 411).

A final and recent study comes from Sinclair *et al,* (2017) a poster for the UK Radiological and Radiation Oncology Congress. They run 4 different algorithms using the marks of 50 students. Their results include borderline adjustments, which confound any direct comparisons, but again, the degree outcomes change significantly depending on which algorithm is applied. In their concluding remarks Sinclair *et al,* (2017) note that "Students with the same final score will get a different award classification dependent on the algorithm used and Universities using these models may benefit from an improved contribution to ranking performance".

The implication on performance ranking is reasonable; however only the Complete University Guide (CUG) explicitly reports degree classifications in its metrics while the Guardian League table uses degree classifications indirectly in their value added calculations (Turnball, S. (2018). The variation in the way UK universities calculate their degree outcomes and the risk of classification inflation does however present significant problems for those researchers who might want to use degree outcomes in any comparative study. In this context, it is notable that the Teaching Excellence Framework (TEF) does not use degree classifications in its metrics[3].

Section 5: Measuring the impact of diverse algorithms

Here we apply the 20 algorithms listed in tables 1 and 2 to the aggregated marks of three cohorts of students on the same degree in order to illustrate how the proportion of 1st, U2, L2 and 3rds change under these different algorithms.

The sample

Like Wolf and Turner (1997) and Yorke *et al,* (2014) the set of module marks come from a medium sized degree course in an established subject delivered in a large English university (call it 'BA (Hons) ABC'). These marks exclude students who may have failed a module in their final year and does not take into account any borderline adjustments. Given that degree outcomes can and do change year on year, the simulation looks at three consecutive years of graduation data (Cohorts A, B and C) and all simulations use this aggregated data (n = 211).

Table 4 lists the average marks and standard deviations per year and the overall average (per cohort). Included is the distribution of marks using algorithm No.1, batched in the decile increments used by the traditional degree classifications (3rd, L2, U2, and 1st). Figures 3 (a) to (d) plot the distribution of degree marks batched in 2.5 percentage increments. Figures 3(a) to (c) cover the three cohorts (A, B, and C) and figure (3d) shows the distribution for all cohorts combined.

---

[3] In TEF 3 (the subject based TEF), it is intended to measure for 'grade inflation'. In the summary policy document the measure is described as "the proportion of firsts, 2:1s and other grades as a percentage of all classified degrees at that provider 1, 2, 3 and 10 years before the year of assessment. It will be collected by a mandatory declaration from all providers that apply for assessment". (Teaching Excellence Framework, 2017, p.10). That is to say, TEF 3 will be measuring classification inflation, not grade inflation.

Using this smaller increment (2.5%) allows us to see those marks that would be classified as 'high' U2s, which equals 21.8% across the aggregated data. They also closely match the structure of the grade point average (GPA) proposed by the Higher Education Academy.

Table 4: Annual average and mark distribution BA (Hons) ABC

Graduating cohort		Y1	Y2	Y3	Overall	Mark Distribution				
						40-49.5	50-59.5	60-69.5	70-79.5	80+
Cohort A	AVE	67.6	61.5	66.3	65.1	1	14	34	19	0
[n = 68]	SD	7.61	8.10	6.33	7.82	1.5%	20.6%	50.0%	27.9%	-
Cohort B	AVE	64.5	60.8	65.3	63.5	2	15	38	13	0
[n = 68]	SD	7.88	7.99	7.68	8.09	2.9%	22.1%	55.9%	19.1%	-
Cohort C	AVE	60.8	59.5	64.5	61.6	4	27	35	9	0
[n = 75]	SD	7.45	8.06	6.83	7.75	5.9%	39.7%	51.5%	13.2%	-
ALL Chorts	AVE	64.2	60.6	65.3	63.4	7	56	107	41	0
[n = 211]	SD	8.14	8.10	7.00	8.02	3.3%	26.5%	50.7%	19.4%	-

Figure 3: Distribution of degree marks, cohorts A, B and C and All Cohorts combined

Figure 3(a): Cohort A

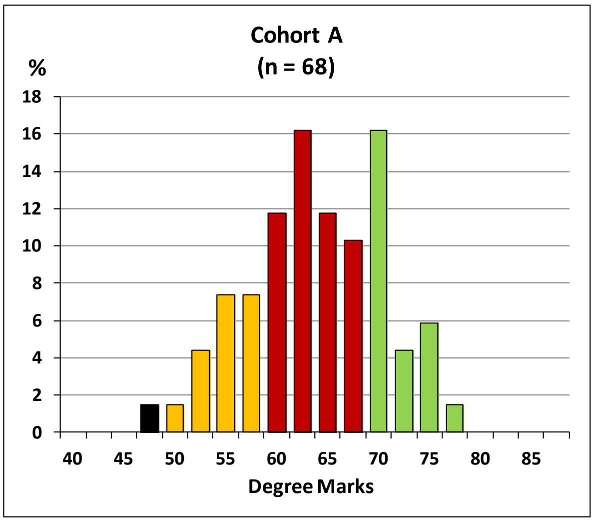


Figure 3(b): Cohort B

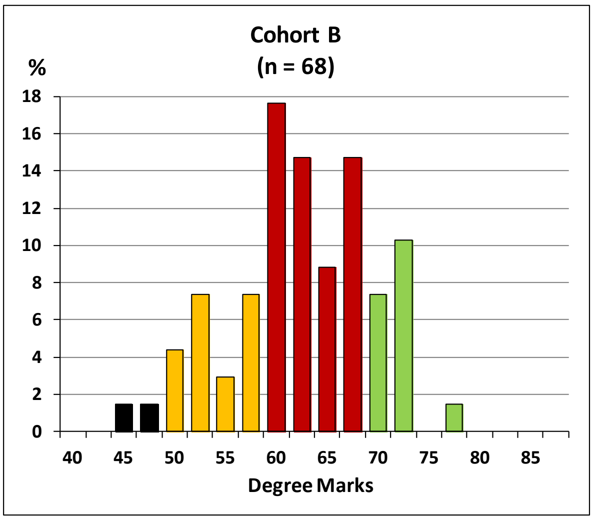


Figure 3 (c): Cohort C

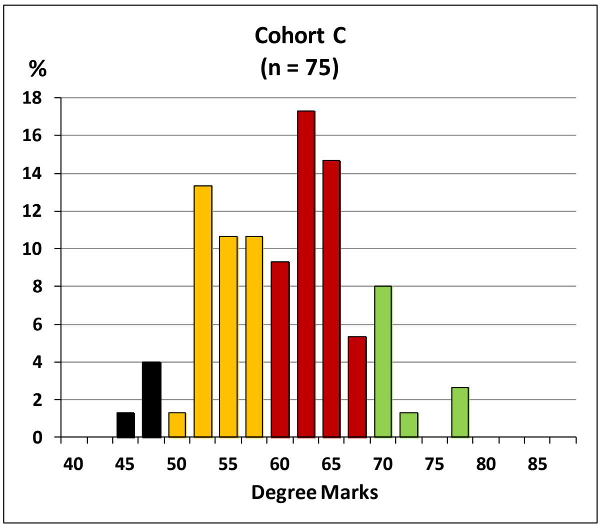
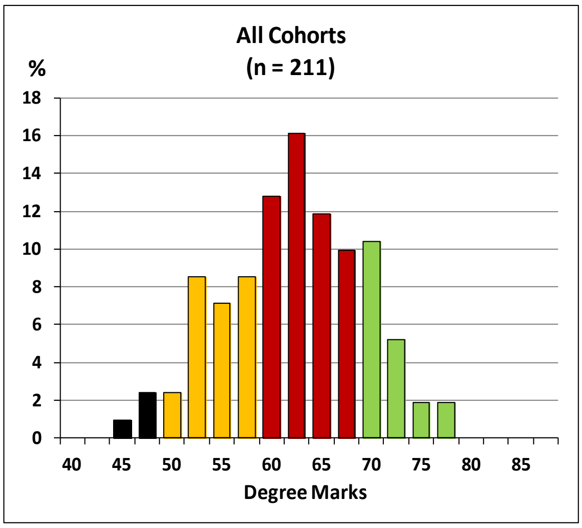


Figure 3(d): All Cohorts combined



The data and plots are included here to enable the reader to make their own judgements on whether this sample is representative or similar to data they may already be familiar with. It is pertinent to note that the performance of each graduating cohort is declining (i.e. there is no grade inflation).

In addition, and for each cohort, the year two average mark is lower than the year one marks. It is unknown if this is typical for the sector as a whole, but anecdotal evidence from external examiners suggests this dip in year two marks is a common phenomenon, one which might explain (in part) the sector's faith in the notion of exit velocity.

5(a): Measuring the impact of different algorithms

Methodology

The individual module marks (and their related credit) for all 211 students were entered into an excel spread sheet and the yearly averages computed. Separate columns are used to identify the best 100 credits and the best 90 credits for year two and three. Similarly, columns were created to calculate the average weighted marks used in each algorithm.

The initial set of results are reproduced in table 5 and illustrated in figure 7, and show that as we progress through the algorithms the number of thirds (3^{rd}) and lower seconds (L2) awards decline, while the number of upper second (U2) remains comparatively stable but firsts (1^{st}) increase.

The evidence for classification inflation comes from comparing the change in firsts: 35 when using algorithm No.3, to 54 when applying algorithm No. 8, to 67 when using algorithm No. 14, up to 72 when using algorithm No.20, or 37 additional firsts with no change in the individual module marks underpinning the degree calculations. The outside observer might wonder where these additional 37 firsts have come from.

Looking at table 5 and figure 7 we can get a sense of how classification inflation might have occurred in the UK if, over the years, a large number of universities gradually changed their algorithms from those based on all credits (for years two and three) to those using discounting and differential weighting.

In some respects, the prompt to make such changes may have come from Quality Assurance Agency (QAA) advice that universities should reduce the amount of variation across their programmes. In making these changes, it is very likely that the senior managers [4] of these universities looked around to see what their competitor set were doing and made changes to their algorithms but in ways that did not 'disadvantage' their students.

A case in point would be UCL who were recently advised by the QAA to "minimise the amount of variation and derogation across its programmes as soon as possible" (UCL, 2017, p.1). In response UCL set up a working group which used the same type of simulations conducted here and applied them to the different algorithms used *within* UCL programmes (33 in all) and those algorithms used by their competitor institutions (i.e. 19 Russell Group universities). UCL then benchmarked their "harmonised" scheme (p. 6) against this sector.

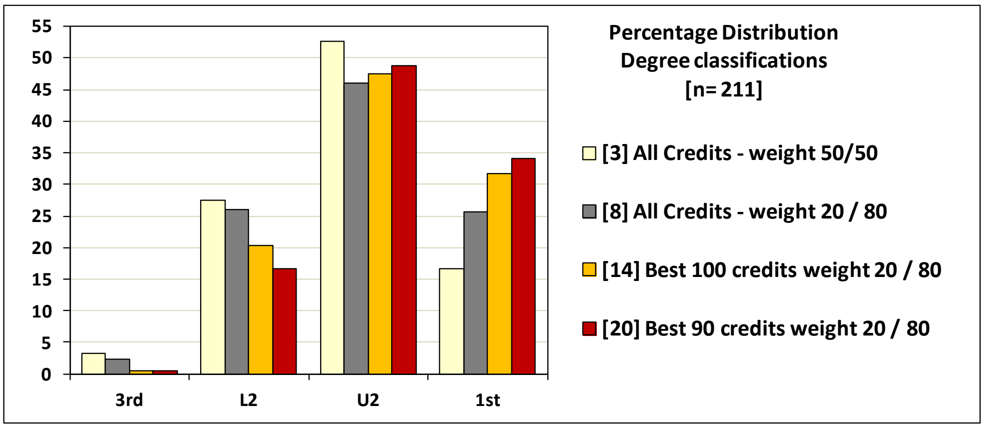
[4] As opposed to front line lecturers and teaching staff, who in the normal course of events, would not have been consulted directly on such matters.

This raises a significant question: if in ‘looking around’ and running similar simulations universities saw increasing disparities in the distribution of classification outcomes, why did nobody say something.

Table 5: Distribution of degree classification using different algorithms (n = 211) [5]

ALL CREDITS AND ALL YEARS USED [360 in total]							Percentage Distibution of Classifications				
	Years Used	Weights	3rd	L2	U2	1st	3rd	L2	U2	1st	U2 + 1st
1	Y1/Y2/Y3	EVEN	7	56	107	41	3.3	26.5	50.7	19.4	70.1
2	Y1/Y2/Y3	10/30/60	6	56	105	44	2.8	26.5	49.8	20.9	70.6
[1] NO DISCOUNTING ALL CREDITS USED [240 in total]											
	Years	Weights	3rd	L2	U2	1st	3rd	L2	U2	1st	U2 + 1st
3	Y2/Y3	50 - 50	7	58	111	35	3.3	27.5	52.6	16.6	69.2
4	Y2/Y3	40 - 60	6	57	108	40	2.8	27.0	51.2	19.0	70.1
5	Y2/Y3	33 - 67	5	57	106	43	2.4	27.0	50.2	20.4	70.6
6	Y2/Y3	30 - 70	5	57	103	46	2.4	27.0	48.8	21.8	70.6
7	Y2/Y3	25 - 75	5	56	102	48	2.4	26.5	48.3	22.7	71.1
8	Y2/Y3	20 - 80	5	55	97	54	2.4	26.1	46.0	25.6	71.6
DISCOUNTED BEST 100 credits per level [200 credits in total]											
	Years	Weights	3rd	L2	U2	1st	3rd	L2	U2	1st	U2 + 1st
9	Y2/Y3	50 - 50		55	103	53		26.1	48.8	25.1	73.9
10	Y2/Y3	40 - 60	1	52	102	56	0.5	24.6	48.3	26.5	74.9
11	Y2/Y3	33 - 67	1	48	103	59	0.5	22.7	48.8	28.0	76.8
12	Y2/Y3	30 - 70	1	46	102	62	0.5	21.8	48.3	29.4	77.7
13	Y2/Y3	25 - 75	1	44	102	64	0.5	20.9	48.3	30.3	78.7
14	Y2/Y3	20 - 80	1	43	100	67	0.5	20.4	47.4	31.8	79.1
DISCOUNTED BEST 90 credits per level [180 credits in total]											
	Years	Weights	3rd	L2	U2	1st	3rd	L2	U2	1st	U2 + 1st
15	Y2/Y3	50 - 50		49	101	61		23.2	47.9	28.9	76.8
16	Y2/Y3	40 - 60		45	100	66		21.3	47.4	31.3	78.7
17	Y2/Y3	33 - 67	1	42	101	67	0.5	19.9	47.9	31.8	79.6
18	Y2/Y3	30 - 70	1	40	101	69	0.5	19.0	47.9	32.7	80.6
19	Y2/Y3	25 - 75	1	38	101	71	0.5	18.0	47.9	33.6	81.5
20	Y2/Y3	20 - 80	1	35	103	72	0.5	16.6	48.8	34.1	82.9

Figure 7: Histogram for algorithms Nos. 3, 8, 14 and, 20



[5] These results are comparable to those shown in Allen (2018) tables 5 and 6. Where there are difference it is because of different rounding protocols when calculating the yearly averages .

5(b): Number of students affected by the application of different algorithms

Methodology

Estimating the number of students potentially affected by different algorithms is achieved by converting all the weighted averages (across all the algorithms used) into their respective classifications. The cut-off between each boundary is only 0.01% e.g. the range for a 3rd is 40% to 49.99 %, for a L2 it is 50% to 59.99% and so on.

An example is shown in table 6. Students with the same classification across a *given* range of algorithms are classified as 'No Change'; those students with one or more different classification are classified as 'Change'. Therefore, in table 6 students 2 and 3 would be classed as 'Change' across all algorithms, but 'No Change' across algorithms No. 3 to No. 20, whereas student 4 would be classified as 'No Change' across the whole range of algorithms.

The numbered algorithms (No.1 to No.20) are those found in table 5, algorithms No. 3 to No. 8 including 'Yorke (2004)' are the same algorithms tested in Yorke *et al* (2004). Finally, the distribution of classifications for each group ('Change' and 'No Change') across the range of algorithms is tabulated then compared (e.g. the tally for algorithm No. 3 could be compared with that of algorithm No. 20).

Table 6: Measuring student degree outcomes by algorithm

	All Years of Study		ALL CREDITS Y2/Y3						100 CREDITS Y2/Y3						90 CREDITS Y2/Y3						Yorke (2004)
Alg =>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Y3
1	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	U2
2	1ST	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2
3	U2	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST	1ST
4	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2
5	3rd	L2	3rd	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2	L2
6	L2	L2	L2	L2	L2	L2	L2	L2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	L2
7	L2	L2	L2	L2	L2	L2	U2	U2	L2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2
8	L2	L2	L2	L2	U2	U2	U2	U2	L2	L2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2	U2

No Change (No.3 to No.20)

Alg =>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	YORK
3rd																					
L2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
U2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3
1ST	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Change (No.3 to No.20)

Alg =>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	YORK
3rd	1		1																		
L2	3	4	3	4	3	3	2	2	3	2	1	1	1	1	1	1	1	1	1	1	2
U2					1	1	2	2	1	2	3	3	3	3	3	3	3	3	3	3	2
1ST																					

Comparison to Yorke et al, (2004)

The first set of results compares this paper's simulations to those of Yorke *et al*, (2004) for University A which has the same credit structure underlying the data used in this study.

Table 7 shows the results of changing the weighting on year three from 50/50 to 0/100 (i.e. basing the degree classification solely on the year 3 average mark). There are 58 students with a change in degree classification, which represents 27.5% of the sample used here. This compares to the Yorke *et al* (2004) estimate of 22.3% for university A (see results [2]- table 3). The approximate proportions suggest a comparable computation is being used.

The portion of firsts changes from 16.6% in algorithm No. 3 to 29.4% when the year two average marks are removed from the algorithm. Likewise, the proportion of good honours (U2 + 1sts) changes from 69.2% to 75.4%. Interestingly the application of the Yorke *et al*, (2004) algorithm (year three only) resulted in only one student's potential degree classification to fall, from a first (1st) to an upper second (U2), a much lower proportion than that reported in the Yorke *et al*, (2004) study (see results [1], table 3).

Table 7: Comparison to Yorke *et al*, (2004) - increased weighting

Algorithm No. 3, Y2-Y3, All Credits. Weighted 50 / 50					
ALL COHORTS	Change	No Change	ALL	%	U2 + 1st
3rd	3	4	7	3.3%	
L2	20	38	58	27.5%	
U2	34	77	111	52.6%	69.2%
1ST	1	34	35	16.6%	
Total	58	153	211		
Percentage	27.5%	72.5%			
Increased weight on year three					
Yorke *et al*, (2004) Y3, All CREDITS Y3 Weighted 100%					
ALL COHORTS	Change	No Change	ALL	%	U2 + 1st
3rd	1	4	5	2.4%	
L2	9	38	47	22.3%	
U2	20	77	97	46.0%	75.4%
1ST	28	34	62	29.4%	
Total	58	153	211		
Percentage	27.5%	72.5%			

In Yorke *et al* (2004) since, discounting was not combined with increases in the weightings it would be difficult (if not misleading) to make a direct comparison, but this can be inferred by looking at the results from this study, which are discussed below.

The results of this study

Like table 5, table 8 shows the distribution of degree classifications between 'Change' and 'No Change' students as we move up through algorithms No.3, and No. 20 (not including algorithm 'Yorke (2004)'). That is, from an even weighting for year two and three with no discounting, to a 20/80 weighting for year two and three respectively and a total of 60 credits discounted.

In total 81 students (or 38.4%) [6] would see a change in degree classification (had they obtained their results in another institution) which compares to the Yorke *et al*, (2004) estimate of 35.2% for university A (see results [1]- table 3)

The difference between an increased weighting alone ([2] – algorithm No. 8), and weighting and discounting combined [3 and 4] are pertinent (see Difference - table 8). The impact of weighting alone is greatest in the upper seconds (U2) and firsts (1st). The impact of the increased weighting on year three suggests that within the 'Change' group there is a large

[6] The proportion of students whose degree classification can change varies significantly from cohort to cohort; the proportion of 'Change' students for each cohort is: cohort A – 33 (48.5%), cohort B – 21 (30.9%) and cohort C - 36.0%, see table 12. This variation appears to be associated with the variability of year two and three marks within each cohort and merits further research.

proportion of students whose year three mark are significantly better than their year two marks. In terms of discounting and weighting [3 and 4], the impact is greatest in the lower classifications (3rd and L2). Again, the magnitude of these changes tells us that the marks being discounted for each student are significantly lower than those not being discounted.

Table 8: Decomposition between weightings and discounting BA (Hons) ABC)

[1]	Algorithm No. 3, Y2-Y3, All Credits. Weighted 50 / 50					
ALL COHORTS	Change	No Change	ALL	%	U2 + 1st	
3rd	7	0	7	3.3%		
L2	30	28	58	27.5%		
U2	44	67	111	52.6%	69.2%	
1ST	0	35	35	16.6%		
Total	81	130	211			
Percentage	38.4%	61.6%				
Increased weight on year three						
[2]	Algorithm No. 8, Y2-Y3, Best 100 Credits. Weighted 20 / 80					Difference
ALL COHORTS	Change	No Change	ALL	%	U2 + 1st	[2] - [1]
3rd	5	0	5	2.4%		-2
L2	27	28	55	26.1%		-3
U2	30	67	97	46.0%	71.6%	-14
1ST	19	35	54	25.6%		19
Total	81	130	211			
Percentage	38.4%	61.6%				
Discounting - Best 100 credits Y2 and Y3						
[3]	Algorithm No. 14, Y2-Y3, Best 100 Credits. Weighted 20 / 80					
ALL COHORTS	Change	No Change	ALL	%	U2 + 1st	[3] - [2]
3rd	1	0	1	0.5%		-4
L2	15	28	43	20.4%		-12
U2	33	67	100	47.4%	79.1%	3
1ST	32	35	67	31.8%		13
Total	81	130	211			
Percentage	38.4%	61.6%				
Discounting - Best 90 credits Y2 and Y3						
[4]	Algorithm No. 20, Y2-Y3, Best 90 Credits. Weighted 20 / 80					
ALL COHORTS	Change	No Change	ALL	%	U2 + 1st	[4] - [3]
3rd	1	0	1	0.5%		0
L2	7	28	35	16.6%		-8
U2	36	67	103	48.8%	82.9%	3
1ST	37	35	72	34.1%		5
Total	81	130	211			
Percentage	38.4%	61.6%				

It is important to note that as the range of algorithms increase the number of 'Change' students will also increase; examples from this sample are tabulated below:

Range of algorithms	Number of algorithms	'Change' students	Percentage of of this sample
No. 3 to 8	6	36	17.0%
No. 3 to 8 + Yorke (2004)	7	58	27.5%
No. 3 to 20	18	81	38.4%
No. 1 to 20	20	116	54.9%

Therefore, while there is a theoretical certainty that applying different algorithms will result in a range of different degree classification outcomes, the magnitude will be related the total number of algorithms used in the analysis. This may in part explain differences between such studies. Otherwise, we could expect variation across universities, credit structures, and subject groupings and between years.

Illustrating the distribution of marks as different algorithms are applied
As shown earlier (e.g. figure 7) as we progress through the algorithms the number of thirds (3^{rd}) and lower seconds (L2) awards decline, the number of upper second (U2) remains comparatively stable but firsts (1^{st}) increase. Figures 8 and 9 decompose these changes by comparing the distribution of degree marks for algorithms No. 3 and No. 20, for each set of students 'Change' and 'No Change' (before borderline adjustments).

No Change (Figure 8)
For those in the 'No Change' group their degree marks increase but their classification does not (by default). For instance, the number of students with a lower second in the mark range 50 to 54.99 in algorithm No. 3 has fallen from 22 (i.e. 15+7) to 4 in algorithm No. 20, while those in the higher mark range 55 to 59.99 change from 6 to 24 (respectively).

The number of low upper second (mark range 60 to 64.99) in algorithm No. 3 has also fallen from 58 to 18 in algorithm 20, and the number of high upper seconds has increased from 9 (No. 3) to 49 (No. 20). The change in the algorithms is pushing down on the lower marks, which forces up the next batch of higher marks; like Student B in table 2, the starting position of a low U2 becomes a high U2.

Figure 8: 'No Change' distribution of marks, algorithm No. 3 and No. 20 compared

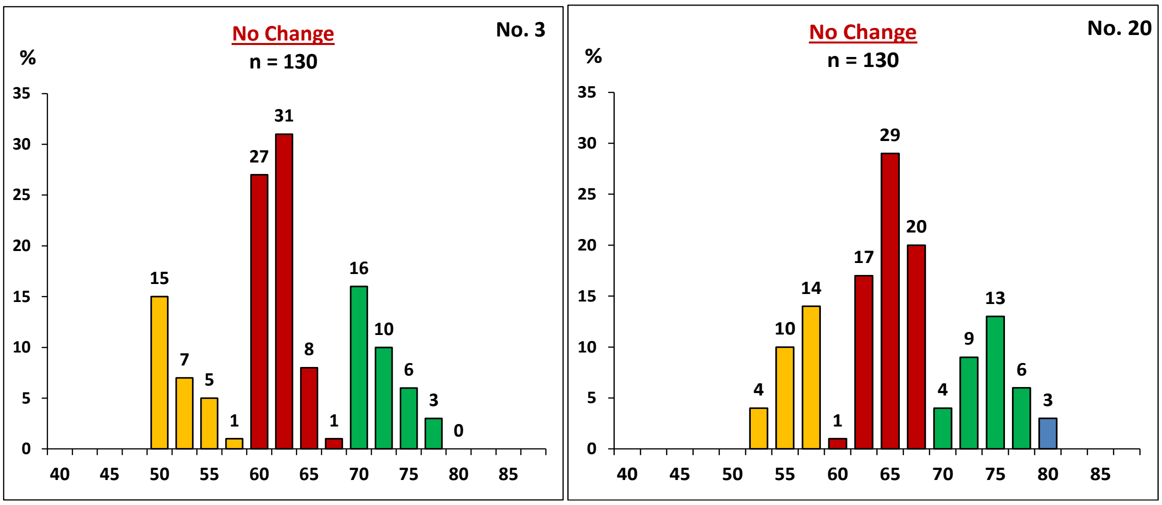
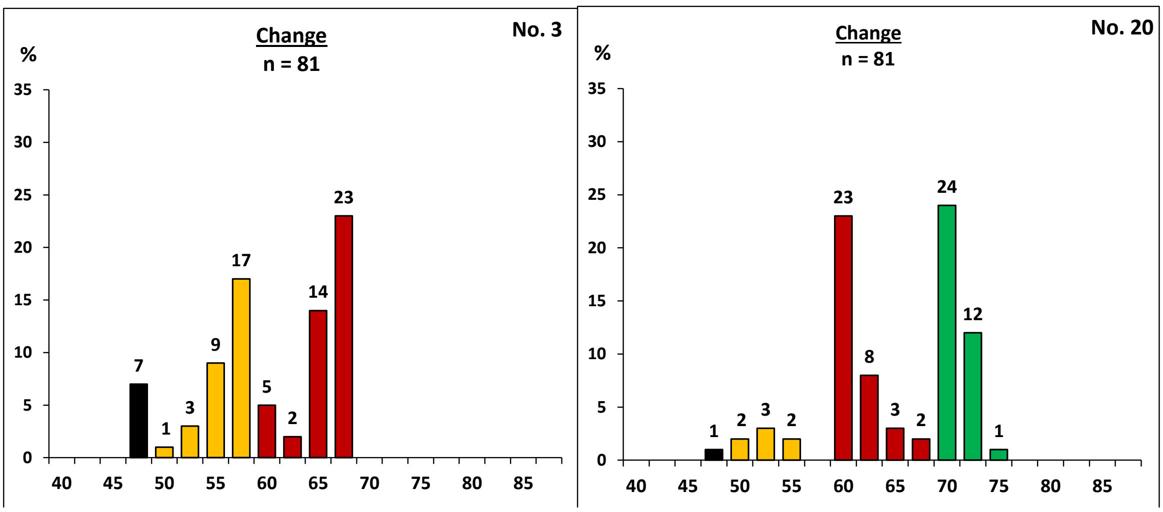


Figure 9: 'Change' distribution of marks, algorithm No. 3, and No. 20 compared



Change (Figure 9)
For those students whose classification could have been different it is the same story but the starting point is higher and closer to the classification boundary, and as a result, these students' classification tips over into the next classification. For example, there are 37 students with a high upper second in algorithm No.3, but this falls to only 5 in algorithm No.20, in the meantime the number of first has increased from zero in No. 3 to 37 in algorithm No. 20. The number of students in the mark range 40 to 59.99 has also fallen dramatically from 37 in algorithm No. 3 to just 8 in algorithm No. 20.

5(c): Estimating the impact of borderline adjustments

Methodology
Two decision rules are applied to the students' calculated degree average:

[1] Automatic uplift – degree marks equal to or less than 0.5% below a classification boundary are awarded the higher classification.

3rd at 49.5% = L2
L2 at 59.5% = U2
U2 at 69.5% = 1st

[2] Preponderance Principle – In each classification range, an uplift is awarded if the student has 60 credits in the higher boundary in their final year, the qualifying degree mark ranges are:

3rd: 48.50% to 49.45% and 60 Year 3 credits at 50% or above = L2
L2: 58.50% to 59.45% and 60 Year 3 credits at 60% or above = U2
U2: 68.50% to 69.45% and 60 Year 3 credits at 70% or above = 1st

In rule [2], a less forgiving criterion would be to raise the initial qualifying marks by 0.5%, e.g. changing 48.50% to 49.0% and so on. Alternatively, a more generous criterion would be to reduce the number of year three credits in the higher boundary to 40 credits[7].

The results are shown in table 9 which shows that as we move up the algorithms the frequency of these adjustments falls (from 27 to 13) and the need for such adjustments reduces as the algorithms increase the calculated degree mark. It is interesting to note that in algorithm No. 3 the *preponderance* decision is more frequent than the *automatic uplift*, and plays a bigger role in changing the outcomes in the lower classifications, but as we move up through the four algorithms the relative proportion of automatic uplifts increases.

The impact on the aggregate classification outcomes is not as big when applying higher weightings and/or discounting, nevertheless, as we progress through the algorithms the number of thirds (3rd) and lower seconds (L2) awards decline, while the number of upper seconds (U2) remains comparatively stable but firsts (1st) increase from 72 to 79 (37.4%).

Despite the modest impact on the overall classifications, at an individual level these uplifts can be a 'key factor' in determining their degree outcomes. Table 10 records the number of borderline adjustments and shows that in total 68 individual students (or 32.2% of the total sample) would have been granted an uplift somewhere along these four algorithms. It also shows significant variation between the three years – which might be linked to the change in

[7] The criteria in rule [1] are common across the UK higher education sector, however the specifics of the criteria in rule [2] are less so, although they are a fair representation of sector practice.

average marks (see table 6 where higher average marks (e.g. Cohort A) pushes the calculated degree marks closer to the borderline boundaries.

Table 9: Impact of borderline adjustments on selected algorithms (n = 211)

No. 3	3rd	L2	U2	1st	Borderline Adjustments
Classifications Before	**7**	**58**	**111**	**35**	
Loose through automatic uplift [-]	**0**	**3**	**5**		**8**
Loose through preponderance [-]	**4**	**6**	**9**		**19**
Gain through automatic uplift [+]		**0**	**3**	**5**	
Gain through preponderance [+]		**4**	**6**	**9**	
Classifications After	**3**	**53**	**106**	**49**	**27**
Difference	**-4**	**-5**	**-5**	**14**	

No. 8	3rd	L2	U2	1st	Borderline Adjustments
Classifications Before	**5**	**55**	**97**	**54**	
Loose through automatic uplift [-]	**0**	**11**	**4**		**15**
Loose through preponderance [-]	**3**	**1**	**8**		**12**
Gain through automatic uplift [+]		**0**	**11**	**4**	
Gain through preponderance [+]		**3**	**1**	**8**	
Classifications After	**2**	**46**	**97**	**66**	**27**
Difference	**-3**	**-9**	**0**	**12**	

No. 14	3rd	L2	U2	1st	Borderline Adjustments
Classifications Before	**1**	**43**	**100**	**67**	
Loose through automatic uplift [-]	**0**	**4**	**4**		**8**
Loose through preponderance [-]	**0**	**4**	**3**		**7**
Gain through automatic uplift [+]		**0**	**4**	**4**	
Gain through preponderance [+]		**0**	**4**	**3**	
Classifications After	**1**	**35**	**101**	**74**	**15**
Difference	**0**	**-8**	**1**	**7**	

No. 20	3rd	L2	U2	1st	Borderline Adjustments
Classifications Before	**1**	**35**	**103**	**72**	
Loose through automatic uplift [-]	**0**	**3**	**4**		**7**
Loose through preponderance [-]	**0**	**3**	**3**		**6**
Gain through automatic uplift [+]		**0**	**3**	**4**	
Gain through preponderance [+]		**0**	**3**	**3**	
Classifications After	**1**	**29**	**102**	**79**	**13**
Difference	0	-6	-1	7	

Table 10 also shows that some students can have their classification adjusted more than once, this is because the previous algorithm (e.g. No 3) increased the degree mark but only enough to allow rule [2] to be applied, the next algorithm (e.g. No. 8) 'tipped' the degree mark to enable rule [1] to be applied.

It is surprising that the 'Change' students, 56 in all (69% of this group) dominate these borderline decisions. For the 56 'Change' students there is no additional impact, for want of a better phrase they get the higher classification sooner, that is to say: as a result of the borderline adjustments they do not require further increases in weighting or discounting to get the next higher classification. For some of the 'No Change' students these borderline adjustments do make a difference, 12 of these students would see their classification elevated into the next boundary (which correlates with the modest increase in the classifications shown in table 10).

Table 10: Number of students with borderline adjustments

Year	ONCE	TWICE	Three times	Total
Cohort A - Change	22	4	0	28
Cohort A - No Change	1	0	1	
Cohort B - Change	10	3	0	16
Cohort B - No Change	2	1	0	
Cohort C - Change	15	2	0	24
Cohort C - No Change	5	2	0	
All Change	47	9	0	56
All No Change	8	3	1	12
Total	55	12	1	68

In total, including differential weighting, discounting and borderline adjustments, the number of students in this sample whose degree classification would differ from that initially generated by algorithm No. 3 increases to 93, or 44.% of the sample. In the style of Woolf and Tuner (1997), that represents around 182,309 of the 2016/17 honours graduates in the United Kingdom.

6: Comparative analysis 'Change' and 'No Change'

6(a): A closer look at the outcomes for Change students

The change in potential classification for each of the 81 students (batched by cohort), across algorithms Nos. 3, 8, 14, 15 and 20 and the related calculated degree marks are shown in table 11. The salient observation from these individual profiles is the calculated classifications can go down for many students in the 'Change' group.

For example student number 55 (cohort C) would see a fall in their potential degree classification if algorithm No.8 is applied in preference to algorithm No.3, their year three marks must be substantially lower than their year two marks – which is penalised by the higher weighting given to year three marks (i.e. 80% in algorithm no. 8). Their potential classification could however be restored to a lower second (L2), by using algorithms that discount (e.g. Nos. 14, 15, and 20).

Likewise, 19 students would see a lower classification if algorithm No. 15 is applied in preference to algorithm No. 14 [8] again, regardless of the higher number of discounted credits their year two marks are sufficiently low enough to pull down the calculated degree mark when equally weighted with their year three marks.

For these 19 students the average of their year two marks is 54.3% compared to year three average mark of 67.56%, the average difference between year two and three marks is therefore 13.24 marks. While this shows a big improvement on year two marks and reinforces the notion of exit velocity, we would have to ask why the year two marks are this low in the first place.

In total 22 students in the 'Change' group (27% of this group) who could see a lower degree classification being awarded depending on which university they attended and which algorithm that university uses, but for those affected much would depend on what borderline adjustments might be made.

[8] See students 4, 8, 9, 13, 19, 21, 23, 24, 25, 42, 49, 65, 66, 67, 69, 71, 73, 77 and 78

Table 11: Individual profiles for 'Change' students (n = 81)

COHORT A ['Change']										
St No.	Classification per Algorithm					Marks per algorithm				
	No.3	No.8	No.14	No.15	No.20	No.3	No.8	No.14	No.15	No.20
1	3rd	L2	L2	L2	L2	49.1	53.1	54.6	50.8	55.2
2	L2	L2	L2	U2	U2	57.6	58.2	59.5	60.1	60.1
3	L2	L2	L2	L2	U2	56.0	57.7	59.4	59.0	60.5
4	L2	L2	U2	L2	U2	55.3	59.1	60.6	56.8	60.8
5	L2	L2	U2	U2	U2	58.7	58.7	60.6	61.4	61.0
6	L2	U2	U2	U2	U2	59.9	60.3	61.3	62.3	61.9
7	L2	L2	U2	U2	U2	58.9	59.9	61.1	60.5	61.9
8	L2	L2	U2	L2	U2	56.3	59.7	61.3	58.4	62.1
9	L2	U2	U2	L2	U2	56.9	61.1	61.8	58.8	62.2
10	L2	L2	U2	U2	U2	58.8	59.6	61.5	61.6	62.4
11	L2	L2	U2	U2	U2	58.9	59.7	61.4	62.1	62.5
12	L2	U2	U2	U2	U2	58.0	60.6	62.2	60.4	62.7
13	L2	U2	U2	L2	U2	56.1	60.0	62.0	58.8	63.2
14	L2	U2	U2	U2	U2	59.7	60.9	62.6	62.2	63.7
15	L2	U2	U2	U2	U2	59.3	61.2	63.0	62.6	64.0
16	L2	U2	U2	U2	U2	59.4	61.0	64.0	63.1	66.0
17	U2	U2	U2	1ST	U2	66.9	65.7	68.2	70.3	69.5
18	U2	U2	U2	1ST	1ST	68.1	68.2	69.5	70.9	70.2
19	U2	U2	1ST	U2	1ST	64.2	69.2	70.3	66.4	70.7
20	U2	U2	U2	1ST	1ST	66.9	67.5	69.9	70.3	71.0
21	U2	U2	1ST	U2	1ST	66.4	68.8	70.9	69.4	71.6
22	U2	U2	1ST	1ST	1ST	68.4	69.3	70.9	70.6	71.6
23	U2	U2	1ST	U2	1ST	66.4	69.6	71.1	68.9	71.7
24	U2	U2	1ST	U2	1ST	66.6	69.1	71.5	69.2	71.9
25	U2	U2	1ST	U2	1ST	66.8	69.8	71.2	68.4	72.0
26	U2	1ST	1ST	1ST	1ST	68.6	70.0	71.4	70.7	72.2
27	U2	1ST	1ST	1ST	1ST	68.2	70.1	71.4	71.2	72.2
28	U2	1ST	1ST	1ST	1ST	69.0	71.0	72.0	71.1	72.8
29	U2	1ST	1ST	1ST	1ST	69.2	71.6	72.6	70.9	73.0
30	U2	U2	1ST	1ST	1ST	69.4	69.1	72.0	73.6	73.3
31	U2	1ST	1ST	1ST	1ST	68.4	71.1	73.2	71.7	73.9
32	U2	1ST	1ST	1ST	1ST	69.9	71.7	73.2	72.1	73.9
33	U2	1ST	1ST	1ST	1ST	69.5	72.4	73.9	72.5	75.1
					MEAN	62.8	64.7	66.4	65.4	67.2

COHORT B ['Change']										
St No.	Classification per Algorithm					Marks per algorithm				
	No.3	No.8	No.14	No.15	No.20	No.3	No.8	No.14	No.15	No.20
34	3rd	3rd	3rd	L2	3rd	49.1	47.5	48.4	51.0	48.9
35	3rd	3rd	L2	L2	L2	49.2	49.2	50.9	52.0	51.9
36	3rd	3rd	L2	L2	L2	49.0	49.3	53.3	53.2	54.5
37	3rd	3rd	L2	L2	L2	49.3	48.9	50.5	54.3	55.1
38	U2	L2	U2	U2	U2	60.0	59.1	60.3	62.7	61.1
39	U2	L2	U2	U2	U2	60.3	59.1	60.4	63.0	61.2
40	U2	L2	U2	U2	U2	60.5	59.5	60.9	62.8	61.8
41	U2	L2	U2	U2	U2	61.6	59.9	61.8	63.5	62.2
42	L2	L2	U2	L2	U2	53.1	59.8	62.2	55.3	62.9
43	L2	U2	U2	U2	U2	59.1	62.1	64.2	62.1	65.5
44	L2	U2	U2	U2	U2	59.7	63.6	64.8	62.1	65.5
45	U2	U2	U2	1ST	U2	68.5	66.2	68.0	70.9	69.1
46	U2	U2	U2	U2	1ST	65.4	67.2	69.5	68.3	71.0
47	U2	U2	1ST	1ST	1ST	67.6	68.2	70.3	71.0	71.6
48	U2	U2	1ST	1ST	1ST	69.3	69.6	70.9	71.4	71.8
49	U2	1ST	1ST	U2	1ST	65.1	70.9	71.7	66.9	72.2
50	U2	1ST	1ST	1ST	1ST	69.0	70.7	71.7	71.3	72.3
51	U2	1ST	1ST	1ST	1ST	68.6	71.1	72.7	71.6	73.3
52	U2	1ST	1ST	1ST	1ST	69.7	71.7	72.7	71.4	73.4
53	U2	1ST	1ST	1ST	1ST	68.3	71.3	73.4	70.8	74.3
54	U2	1ST	1ST	1ST	1ST	69.0	71.6	73.6	72.7	74.5
					MEAN	61.5	62.7	64.4	64.2	65.4

COHORT C ['Change']										
St No.	Classification per Algorithm					Marks per algorithm				
	No.3	No.8	No.14	No.15	No.20	No.3	No.8	No.14	No.15	No.20
55	L2	3rd	L2	L2	L2	50.1	49.5	51.2	53.2	52.1
56	3rd	L2	L2	L2	L2	48.8	51.4	53.2	51.0	54.0
57	3rd	L2	L2	L2	L2	49.2	53.1	54.1	50.6	54.7
58	L2	L2	L2	L2	U2	57.7	58.3	59.3	59.3	60.0
59	L2	L2	L2	L2	U2	53.2	57.1	58.9	55.9	60.1
60	L2	L2	L2	L2	U2	57.8	58.7	59.6	59.3	60.1
61	L2	L2	L2	U2	U2	57.6	56.6	58.8	60.8	60.2
62	L2	L2	L2	U2	U2	58.1	57.9	59.5	61.2	60.5
63	L2	L2	L2	U2	U2	57.2	58.6	59.9	60.4	60.8
64	U2	L2	U2	U2	U2	61.3	59.7	60.8	63.0	61.1
65	L2	L2	U2	L2	U2	56.3	58.3	60.9	59.8	61.7
66	L2	L2	U2	L2	U2	53.7	59.7	61.1	55.5	61.8
67	L2	L2	U2	L2	U2	56.7	59.8	61.2	59.5	62.0
68	L2	L2	U2	U2	U2	57.6	59.6	62.1	60.1	62.7
69	L2	U2	U2	L2	U2	56.2	60.4	62.3	58.8	62.8
70	U2	U2	U2	U2	1ST	66.6	66.6	70.0	69.7	70.7
71	U2	U2	1ST	U2	1ST	65.9	70.0	68.9	67.7	70.8
72	U2	U2	U2	U2	1ST	67.1	68.4	70.0	69.6	70.8
73	U2	U2	1ST	U2	1ST	66.3	68.4	70.2	69.3	71.1
74	U2	U2	1ST	1ST	1ST	69.8	69.8	70.6	71.3	71.1
75	U2	U2	1ST	1ST	1ST	68.6	68.7	70.8	71.3	71.5
76	U2	1ST	1ST	1ST	1ST	68.4	70.7	71.3	70.3	71.8
77	U2	1ST	1ST	U2	1ST	62.6	70.2	71.3	65.0	72.0
78	U2	1ST	1ST	U2	1ST	67.1	70.7	71.6	69.1	72.1
79	U2	1ST	1ST	1ST	1ST	67.6	70.3	71.7	70.0	72.5
80	U2	1ST	1ST	1ST	1ST	69.8	70.7	72.0	72.0	72.9
81	U2	1ST	1ST	1ST	1ST	67.2	70.2	72.6	70.8	74.1
					MEAN	60.7	62.7	64.2	63.1	65.0
					ALL COHORTS MEAN	61.7	63.5	65.1	64.3	66.0

Algorithm	Years	Credits	Weight
No. 3	Y2/Y3	240	50/50
No. 8	Y2/Y3	240	20/80
No. 14	Y2/Y3	100	20/80
No. 15	Y2/Y3	90	50/50
No. 20	Y2/Y3	90	20/80

6(b): Impact of differential weighting and discounting: 'Change' versus 'No Change'
Table 12 decomposes the impact of differential weighting and discounting between the two groups of students. The difference in overall marks between algorithm No. 20 and No. 3 (see Average Differences, column Overall / 20 - 3) is greatest for the 'No Change' group at 4.3 marks compared to 3.3 marks for the 'No Change' students. These differences when expressed as a percentage change in the mark for algorithm No. 3 become 6.9% and 5.1% respectively. That is to say, on average, the 'No Change' group benefit more from a greater weighting on year three and greater discounts on their lower marks.

To decompose the impact of increased weighting we deduct the mark derived using algorithm No. 8 (all credits, year two and three weighted 20/80) from algorithm No. 3 (all credits year two and three weighted 50/50) see Average Differences, column W / 8 - 3. For the 'Change' group the average difference is 1.7 marks, which represents 2.8% percentage increase on the mark from algorithm No. 3, for the 'No Change' group the average difference is 1.2 marks or a 1.9% increase on algorithm No. 3 mark.

To estimate the impact of discounting we deduct the calculated mark derived using algorithm No 8 from that derived using algorithm No. 20 (which discounts 60 credits from years two and three and uses the same weighting 20/80). For the 'Change' group the average difference is 2.52 marks, for the 'No Change' group the average difference is 1.21 marks, the contribution to the increase in marks between algorithm No. 3 and 20 is 4.1% and 3.2% respectively.

Table 12: Decomposing the impact of differential weighting and discounting

'Change'		Algorithm			Average Differences			% Increase No. 20 over No.3		
					Overall	W	D	Overall	W	D
Cohort	No.	No. 3	No. 8	No. 20	20 - 3	8 - 3	20 - 8	20 - 3	8 - 3	20 - 8
A	33	62.8	64.7	67.2	4.4	1.9	2.5	7.0%	3.0%	4.0%
B	21	61.5	62.7	65.4	3.9	1.2	2.8	6.4%	1.9%	4.5%
C	27	60.7	62.7	65.0	4.4	2.0	2.4	7.2%	3.3%	3.9%
ALL	81	61.7	63.5	66.0	4.3	1.7	2.5	6.9%	2.8%	4.1%

'No Change'		Algorithm			Average Differences			% Increase No. 20 over No.3		
					Overall	W	D	Overall	W	D
Cohort	No.	No. 3	No. 8	No. 20	20 - 3	8 - 3	20 - 8	20 - 3	8 - 3	20 - 8
A	35	64.9	65.9	67.9	3.0	0.9	2.0	4.6%	1.5%	3.1%
B	47	63.7	65.2	67.2	3.5	1.4	2.1	5.5%	2.2%	3.2%
C	48	62.8	64.0	66.0	3.3	1.2	2.1	5.2%	1.9%	3.3%
ALL	130	63.7	64.9	67.0	3.3	1.2	2.1	5.1%	1.9%	3.2%

No. = Number of students, W = Weighting and D = Discount

These differences can only be explained by differences in mark attainment, within years and across (counting) years between the two groups of students. For the 'Change' students the difference in marks within a year and between years must be relatively bigger than they are for the 'No Change' students.

Nevertheless, in both cases, there is no change in the underlying module marks that feed into the degree algorithms. The differences in mark attainment between the 'No Change' and 'Change' students are explored below.

6(c) Measures of consistency

It is remarkable that for the majority of students (62%) in this sample, applying all the algorithms (from No. 3 to No. 20) makes **no** difference to their degree classification (subject to any subsequent borderline adjustments). The supposition raised in section 3 is that differential weighting and discounting favours students whose module marks within a given year and, between counting years, are more varied or inconsistent. That is to say, students with more consistent marks are not greatly advantaged by the application of differential weighting or discounting.

Inspection of the sample used here shows there is a difference in consistency between the two groups of students ('Change' versus 'No Change') where consistency is the extent to which a student's year three marks are similar to their year three marks. Scatter plots of year two marks on year three marks are shown for both groups in figures 10(a) and 10(b). These plots show that for the 'No Change' students their year three marks are more closely associated with their year two marks ($R^2 = 0.69$) than those of the 'Change' students ($R^2 = 0.35$).

Figure 10: Scatter plots year two and year three

Figure 10(a): Change

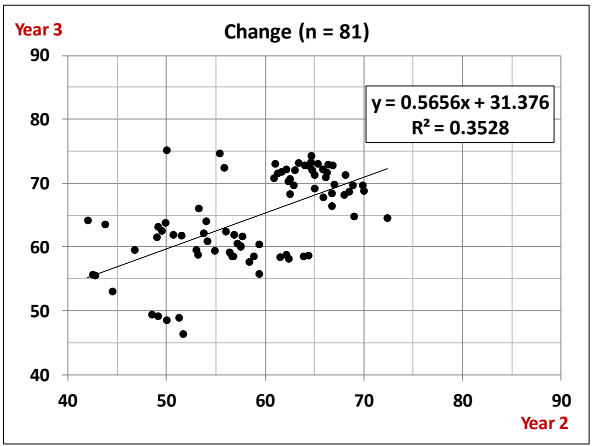
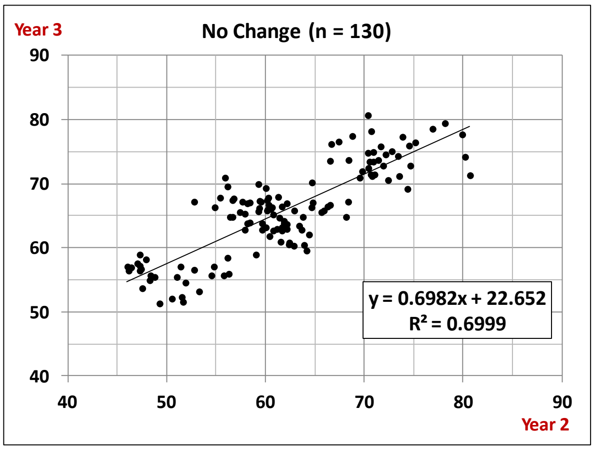


Figure 10(b): No Change



In addition, figures 11(a) and (b) shows the percentage distribution of the differences between year three marks and year two marks for the two sets of students. In figure 11(b), the spread of mark differences (Y3 -Y2) is large and the frequency appears random or at least haphazard. In figure 11(a), the distribution of differences almost approximates to a normal distribution.

If consistency is a smaller difference between year two and three then the 'No Change' students are more consistent in mark attainment than the 'Change' students. In the 'No Change' group 52.3% of the students had a year three mark that was between 0 and 5% greater than their year two mark, for the 'Change' group this porpotion was lower at 32.1%. Where year three marks are 6% or more above the year two marks, the porportions are 'No Change' 40.0% and 'Change' 55.6% respectively.

In terms of discounting, consistency would be a smaller spread in module marks for each counting year. Table 14 lists the average spread in marks for each cohort (A, B and C) and between the two groups. The outcome here is mixed, in year two the 'No Change' group has the higher average difference (or spread) in marks at 21.2 marks compared to 19.7 marks for the 'Change' group.

Figure 11: Distribution of the difference in Year 3 and Year 2 Marks

Figure 11(a): No Change

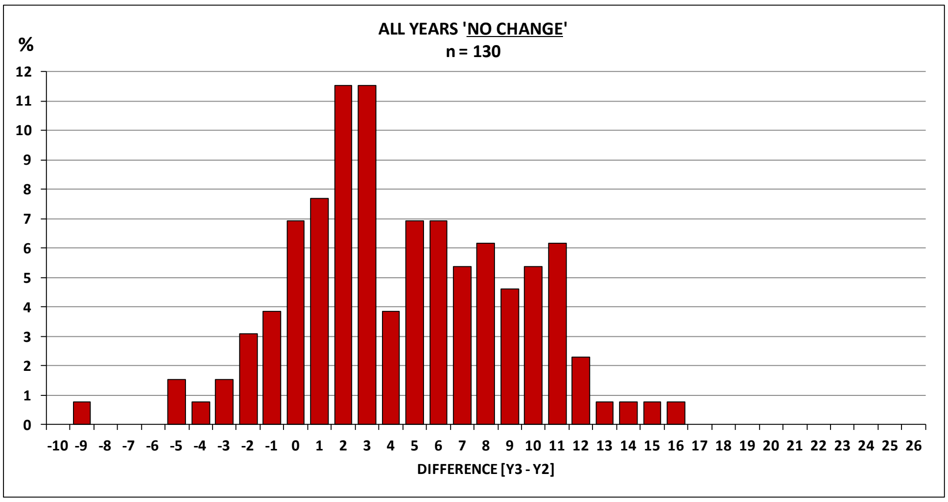


Figure 11(b): 'Change'

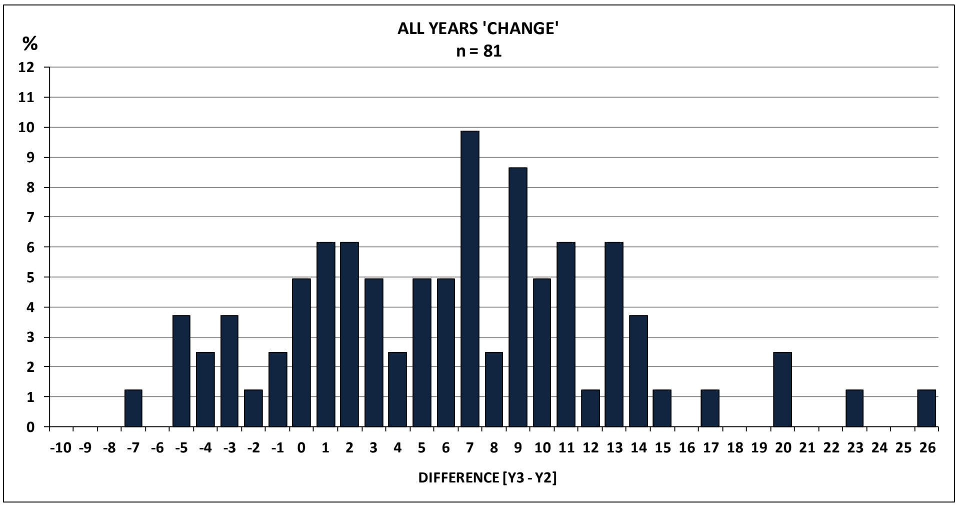


Table 13: Average spread in module mark (highest versus lowest mark)

CHANGE (n = 81)						
	Year 2			Year 3		
	High	Low	Diff	High	Low	Diff
A	69.9	49.5	20.4	75.0	56.9	18.1
B	70.3	50.4	19.9	72.8	52.1	20.7
C	67.4	48.7	18.7	71.8	54.7	17.0
Total	69.2	49.5	19.7	73.4	54.9	18.4
NO CHANGE (n = 130)						
	Year 2			Year 3		
	High	Low	Diff	High	Low	Diff
A	75.1	53.9	21.2	75.7	59.3	16.5
B	74.5	51.6	22.9	73.3	58.6	14.7
C	71.2	51.7	19.5	72.1	57.6	14.5
Total	73.4	52.2	21.2	73.5	58.4	15.1

In year 3 the average difference in marks for the 'Change' group 18.4 marks, 3.3 marks greater than that for the 'No Change' group (15.1). This greater difference the net year 3 marks (after discounting) combined with differential weighting contributes to classification inflation within the 'Change' group.

Finally, table 14 shows the average yearly marks (and their related standard deviations) for both sets of students. The marks for the 'No Change' students are generally higher, particularly in years two and three. The standard deviations across the cohorts are however more varied and do not provide the basis for any strong inference or conclusion.

Table 14: Yearly average marks and exit velocity

	COHORT A ['Change'] n = 33				COHORT A ['No Change'] n =35			
Year marks & Exit Velocity [Y3-Y2]	Y1	Y2	Y3	Y3 - Y2	Y1	Y2	Y3	Y3 - Y2
MEAN	66.8	59.6	66.0	6.3	68.3	63.4	66.5	3.16
STANDARD DEVIATION	7.74	6.62	5.49	4.77	7.40	8.91	7.03	5.25
	COHORT B ['Change'] n = 21				COHORT B ['No Change'] n = 47			
Year marks & Exit Velocity [Y3-Y2]	Y1	Y2	Y3	Y3 - Y2	Y1	Y2	Y3	Y3 - Y2
MEAN	62.5	59.5	63.5	3.9	65.4	61.4	66.1	4.76
STANDARD DEVIATION	8.33	7.84	8.91	7.94	7.49	8.01	6.88	4.35
	COHORT C ['Change'] n = 27				COHORT C ['No Change'] n = 48			
Year marks & Exit Velocity [Y3-Y2]	Y1	Y2	Y3	Y3 - Y2	Y1	Y2	Y3	Y3 - Y2
MEAN	59.0	57.4	64.0	6.6	61.7	60.8	64.8	4.09
STANDARD DEVIATION	7.40	7.85	6.95	6.94	7.36	7.91	6.66	3.92
	CHANGE n = 81				NO CHANGE n = 130			
Year marks & Exit Velocity [Y3-Y2]	Y1	Y2	Y3	Y3 - Y2	Y1	Y2	Y3	Y3 - Y2
MEAN	63.1	58.8	64.7	5.8	64.8	61.7	65.8	4.08
STANDARD DEVIATION	8.48	7.44	7.08	6.55	7.88	8.29	6.88	4.51

In the case of the 'Change' students, their 'exit velocity' or the difference between years two and three marks is higher (5.8 marks) than that of the 'No Change' students (4.08 marks), however, it appears to be a result of lower year two marks, which the majority of algorithms with a lower weighting on year two condone.

Given there is little or no research into exit velocity [9] the true intention for rewarding it by higher weightings on year three marks seems misjudged. Particularly if, for a large proportion of students, these higher year three marks are the result of students knuckling down and making up for their lower attainment in year two. Certainly, all university lecturers would applaud a student's recovery in mark attainment; however, it is likely they would not want this to be rewarded in preference to other students whose 'attainment behaviour' is equally laudable.

Conclusion

This paper has looked at a wide range of issues relating to the use of degree algorithms by UK universities. The results presented are comparable those found in earlier studies, the main findings being:

[9] One such study comes from Betteney (2015) who describes this lack of research as a 'surprising void'.

1. Discounting and differential weighting used together or in isolation artificially increase a student's degree mark, which in turn can increase their degree classification. That is they cause classification inflation.
2. Over time changes in University degree algorithm to those that use discounting and differential weighing are likely to explain an important part of the rise in the proportion of good honours in the last twenty or so years.
3. In aggregate, a high proportion of students (38%) could "argue that they might have been awarded a different class of degree at another institution." These potential differences are counter to notions of equity and fairness and could have profound implications for the life chances of adversely affected students

This paper provides additional evidence that:

4. Differential weighting and discounting benefits most those students who have a greater variation in module marks and yearly average marks. As such, the differences in the potential classification awarded can occur *between* students on the same programme. This again is counter to notions of equity and fair play.
5. When combined, differential weighting, discounting, and borderline adjustments can potentially increase the degree classification of up to 44% of students.
6. Borderline adjustments do not significantly change the overall award profile of a programme but individually they can change the degree classification for a large proportion of students (32% in this sample).
7. Finally, students who benefit most from discounting and differential weighting also feature prominently in borderline decisions.

It is deeply troubling that the current way UK universities calculate their students' degree classifications largely benefit only one set of students. It is also troubling that this outcome is justified based on the concept of exit velocity which is not fully understood but often dressed up as a pedagogical justification for a higher weighting on year three marks. It is also very likely that employers have not been consulted on the relevance of exit velocity to their specific staffing needs.

If the real reason for the use of discounting and differential weighting is to compensate for poorer attainment in year two and thereby improve (or inflate) degree outcomes, then perhaps it is time UK universities took a closer look at what is happening in the second year of their degree programmes.

Ethical implications aside, the practical question becomes - *why have such a diverse range of algorithms if, for the majority of students, their degree classification is unaffected by whatever algorithm is used.*

Finally, returning to Wathey's (2018) observation that convergence in algorithms might be a solution to grade inflation, based on the evidence here and if equity is a meaningful aspiration, any such algorithm should not involve discounting or differential weighting. The choice is then between algorithm No. 3 and algorithm No. 1, where algorithm No.1 returns a greater number of firsts.

References

Allen, D. O. (2018 - Jan) Degree algorithms, grade inflation and equity: the UK higher education sector, Bristol Centre for Economics and Finance, Working Paper 1803, http://www1.uwe.ac.uk/bl/research/bcef/publications.aspx

Armstrong,M., **Clarkson, P**. and **Noble, M**. (1998) Modularity and Credit Frameworks:The NUCCAT Survey and 1998 Conference Report. Newcastle uponTyne: Northern Universities Consortium for Credit Accumulation and Transfer.

Betteney, M.(2015) Exit velocity: Is there truth in the commonly-held belief that grades for undergraduate students tend to improve from Year 2 to Year 3? A case study
Compass: Journal of Learning and Teaching, Vol 6, No 10, 2015
https://journals.gre.ac.uk/index.php/compass/article/view/136
Accessed 24/02/2018

Biglan, A. (1973) The characteristics of subject matter in different scientific areas, Journal of Applied Psychology, 57(3), 195–203.
http://edshare.soton.ac.uk/15017/1/Biglan_-_1973_-_THE_CHARACTERISTICS_OF_SUBJECT_MATTER_IN_DIFFERENT_ACADEMIC_AREAS.pdf
Accessed 26/01/2018

Coventry University, Academic Regulations
http://www.coventry.ac.uk/Documents/Registry/Regulations%20Academic/2017-18/Academic-Regulations/FINAL-6-Academic-Regulations-for-the-UG-Curr-Frwk-2017.pdf
Accessed 02/04/2018

Hammond. W. (2018), Are today's degrees really first class? Wonkhe blog
http://wonkhe.com/blogs/are-todays-degrees-first-class/
Accessed 15/02/2017

HESA (2018), Higher Education Student Statistics: UK, 2016/17 - Qualifications achieved
https://www.hesa.ac.uk/news/11-01-2018/sfr247-higher-education-student-statistics/qualifications
Accessed 22/02/2018

Higher Education Academy, Guide on grade point average (GPA) for students
https://www.heacademy.ac.uk/system/files/resources/Guide%20on%20grade%20point%20average%20for%20students_0.pdf
Accessed 15/11/2017

House of Commons (2008-9), Innovation, Universities, Science and Skills Committee. Students and Universities, Eleventh Report of Session 2008–09 (HC 170-I, [Incorporating HC 370-i-iii]). London. The Stationary Office.
https://publications.parliament.uk/pa/cm200809/cmselect/cmdius/170/170i.pdf
Accessed 18/02/2018

Memorandum 9 (2009) Submission from ASKe[13]
https://publications.parliament.uk/pa/cm200809/cmselect/cmdius/170/170we16.htm
Accessed 01/02/2018

Memorandum 16 (2009), Submission from the Student Assessment and Classification Working Group (SACWG),
https://publications.parliament.uk/pa/cm200809/cmselect/cmdius/170/170we26.htm
Accessed 01/02/2018

Northumbria University, How your bachelors degree is classified [2016/17]
https://www.northumbria.ac.uk/-/media/corporate-website/new-sitecore-gallery/services/academic-registry/documents/qte/assessment/guidance-for-students/how-your-undergraduate-degree-results-are-classified---student-guidance.pdf?la=en&hash=57EA358DB631DD48BC4BEBA36FC124A13C2266C7
Accessed 02/04/2018

Student Assessment Classification Working Group
https://web.anglia.ac.uk/anet/faculties/alss/sacwg.phtml

Sinclair, N., Wright, C., Edwards, G. and Keane, P. (2017). Degree Classification: Does the Calculation Model Affect the Award?
http://researchopen.lsbu.ac.uk/1743/1/Sinclair%20et%20al%20UKRC%202017.pdf.
Accessed 09/01/2017

Teaching Excellence Framework: Lessons Learned. Summary policy document, September 2017,
http://dera.ioe.ac.uk/30059/1/TEF_Lessons_Learned_Summary_Policy_Document.pdf
Accessed 08/02/2018

Turnball, S. (2018), A Guide to UK League Tables in Higher Education, Higher Education Policy Institute (HEPI), Report 101.
http://www.hepi.ac.uk/wp-content/uploads/2018/01/HEPI-A-Guide-to-UK-League-Tables-in-Higher-Education-Report-101-EMBARGOED-4-JAN-2018.pdf
Accessed 03/04/2018

UCL, (2017) Discussion Paper A : Undergraduate Classification Regulations 2018/19 – unknown author, March 2017.
https://www.ucl.ac.uk/srs/academic-manual/documents/assessment_consultation_2016-17/UG_Classification_Discussion_Paper_March_2017.pdf.
Accessed 02/01/2018

University of Bristol, Academic Quality and Policy Office, Annex 12
http://www.bristol.ac.uk/academic-quality/assessment/annex/weightings/

Universities UK (2004) The Report of the Measuring and Recording Student Achievement Scoping Group (the "Burgess Report") London, Universities UK
http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/measuring-and-recording-student-achievement.aspx
Accessed 09/01/2018

Universities UK (2007) Beyond the honours degree classification: The Burgess Group final report
http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/beyond-the-honours-degree-classification-burgess-group.aspx
Accessed 09/01/2018

Universities UK (2012) Introducing Higher Education Achievement Report (HEAR)
http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/introducing-higher-education-achievement-report-hear.aspx
Accessed 09/01/2018

Universities UK-GuildHE, October 2017, Understanding Degree Algorithms
http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2017/understanding-degree-algorithms.pdf.
Accessed 28/11/2017

University of Wales Trinity St David, How degree classifications are calculated: Initial Degree http://www.uwtsd.ac.uk/media/uwtsd-website/content-assets/documents/academic-office/regulations-student-guides/How-Degree-Classifications-are-Calculated---Initial-Degrees.pdf
Accessed 02/04/2018

Wathey, A. (2018), Taking on grade inflation in UK higher education. Wonkhe blog
http://wonkhe.com/blogs/taking-on-grade-inflation-in-uk-higher-education/
Accessed 15/02/2017

Wolff, J. (2012) 'Degree classifications are extremely crude - and pretty useless.' Guardian 16/07/2012
https://www.theguardian.com/education/2012/jul/16/degree-classification-useless-marks-transcript
Accessed 15/01/208

Woolf, H. and Turner, D. (1997) Honours classifications: the need for transparency. *The New Academic*, (Autumn), pp. 10-12.
https://www.seda.ac.uk/new-academic/NA6.3
Accessed 26/01/2018

Yorke, M., Woolf, H., Stowell, M., Allen, R., Haines, C., Redding, M., Scurry, D., Taylor-Russell, G., Turnbull, W., & Walker, L. (2008) Enigmatic variations: honours degree assessment regulations in the UK. Higher Education Quarterly 62 (3), pp.157-180.
http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2273.2008.00389.x/epdf
Accessed 26/01/2018

Yorke, M., Barnett,G., Evanson, P., Haines, C., Jenkins,D., Knight, P., Scurry,D., Stowell, M. and Woolf, H. (2004) Some effects of the award algorithm on honours degree classifications in UK higher education. *Assessment and Evaluation in Higher Education*, 29 (4), pp. 401-413.
http://www.tandfonline.com/doi/abs/10.1080/02602930310001689000?journalCode=caeh20
Accessed 30/01/2018

Recent UWE Economics Papers
See http://www1.uwe.ac.uk/bl/research/bcef/publications.aspx for a full list.

2018

1803 **The use of differential weighting and discounting in degree algorithms and their impact on classification inflation and equity: A further analysis**
David O. Allen

1802 **Unambiguous inference in sign-restricted VAR models**
Robert Calvert Jump

1801 **Degree algorithms, grade inflation and equity: the UK higher education sector**
David O. Allen

2017

1706 **Internal rationality, heterogeneity and complexity in the new Keynesian model**
Cars Hommes, Robert Calvert Jump and Paul Levine

1705 **The regionalization of national input–output tables: a study of South Korean regions**
Anthony T. Flegg and Timo Tohmo

1704 **The impact of quantitative easing on aggregate mutual fund flows in the UK**
Iris Biefang-Frisancho Mariscal

1703 **Where are the female CFOs?**
Gail Webber, Don J Webber, Dominic Page and Tim Hinks

1702 **Mental health and employment transitions: a slippery slope**
Don J Webber, Dominic Page and Michail Veliziotis

1701 **SMEs access to formal finance in post-communist economies: do institutional structure and political connectedness matter?**
Kobil Ruziev and Don J Webber

2016

1611 **Curriculum reform in UK economics: a critique**
Andrew Mearman, Sebastian Berger and Danielle Guizzo

1610 **Can indeterminacy and self-fulfilling expectations help explain international business cycles?**
Stephen McKnight and Laura Povoledo

1609 **Pricing behaviour and the role of trade openness in the transmission of monetary shocks**
Laura Povoledo

1608 **Measuring compliance with minimum wages**
Felix Ritchie, Michail Veliziotis, Hilary Drew and Damian Whittard

1607 **Can a change in attitudes improve effective access to administrative data for research?**
Felix Ritchie

1606 **Application of ethical concerns for the natural environment into business design: a novel business model framework**
Peter Bradley, Glenn Parry and Nicholas O'Regan