Determining Entitlement to Graded National Courses in Diet 2020

Research and Evidence Report

CONFIDENTIAL

May 2020

Version: 03 May 2020 v1

1 Setting the Scene

SQA has consider a range of options around determining candidate entitlement to graded National Courses in Diet 2020 as a result of the impact of the COVD-19 public health emergency. An *Alternative Certification Model (ACM)* is being implemented to allow SQA to result graded National Courses in Diet 2020 with reliability, equity and fairness – over time, across subjects and levels – in the absence of any candidate performance assessment information. The core element of the ACM is teacher estimates, moderated by SQA, and centres have been asked to provide estimates based on a refined band scale and a rank order of candidates within each refined band.

SQA decided to request a more granular estimate scale and rankings to support more nuanced decision-making – in the absence of candidate marks – through the stages of the *Alternative Certification Model.* Requests for both an extended estimate scale and rankings were made to address two important aspects of accuracy of teacher estimation, i.e. absolute accuracy where the actual grade a candidate achieves is estimated against a national standard and the rank order, which is relative accuracy judgement in the sense that, while it is not possible to determine the actual grade, it is possible to rank candidates in a class relative to each other.

. SQA asked for both teacher estimates and rank order to ensure that we had the maximum amount of information available to inform decisions on grades for individual candidates and the quality assurance process surrounding that. This decision was made by the Board of Management prior to final agreement on the Alternate Certification Model, at a stage when a number of options where under consideration. Other regulators in the rest of the UK where in a similar situation of having to announce what information was required from centres whilst still developing and testing out a range of models.

CONFIDENTIAL

This *Research and Evidence Report* outlines the work undertaken since early March 2020 to determine candidate entitlement to graded National Courses in 2020 as a result of the decision not to run the business-as-usual Diet 2020. The work outlined in this report highlights the challenges of relatively low-uptake qualifications at national and centre-levels, an absence and lack of use of standardised prior attainment information and finally, no assessment data. This means that SQA is limited in adopting a purely statistical approach to moderation. . The data we will have in 2020 includes estimates, rank orders, prior attainment data for a significant proportion of Higher and Advanced Highers. In addition, we have historical data for the previous three years [just 3 years?], including estimates, actual results by subject and centre.

The *Code of Practice Governance Group* has been the main internal decision-making vehicle throughout this period. Further advice has been commissioned from Alpha Plus, an education service business, who specialise in standards, assessment and statistics, around technical validation.

While centres are encouraged to submit Estimates each year, for Diet 2020 this is a prerequisite. This is required for all candidate entries across the 128 graded National Courses and must be submitted to SQA by Friday 29th May 2020.

The timeline outlined below provides an indication of how SQA has proactively responded to this evolving issue including the refinement of the *Alternative Certification Model*:

12th March 2020	SQA's priority remained the delivery of a safe and secure diet of examinations;
19th March 2020	Cabinet Secretary announces the closure of schools in Scotland and the cancellation of Diet 2020 examinations;
20th March 2020	Centres requested to continue to prepare for completion and submission of Coursework assessment components;

24th March 2020	Centres advised that, due to public health advice, Higher, Advanced Higher and late submission National 5 Coursework should not be submitted. Ongoing

	consideration of how to safely and securely mark National 5 Coursework already submitted;
02nd April 2020	Centres advised that National 5 Coursework will not be marked due to latest public health advice and that an enhanced estimate scale and rankings would be required;
20th April 2020	Centres provided with information on estimating and rank ordering candidates for Diet 2020.

2 Guiding Principles

Three guiding principles have underpinned our approach to developing and implementing an alternative certification model for 2020:

CONFIDENTIAL

- Fairness to all learners, across all centres;
- To protect, as far as possible, identifiable groups of candidates through the process so that they are not systematically advantaged or disadvantaged
- Safe and secure certification of our qualifications, while following the latest public health advice, ; Maintaining the integrity and credibility of our qualifications system, ensuring that standards are maintained over time, within the constraints of 2020, in the interests of learners and the credibility of the qualifications that they achieve in 2020.

3 Business-as-Usual Environment

3.1 Setting Grade Boundaries

SQA has a responsibility to individual learners and to the wider community to ensure that the standard of our qualifications is set appropriately and maintained over time and across courses. This means that we have to make sure that the grade a candidate receives recognises achievement against the established grade descriptions of the course. It also means that we have to make sure that it is not easier or harder to achieve the same result across different courses.[1]

This is achieved through the development of course assessments based on an assessment 'blueprint' and consistent application of detailed 'fit-for-purpose' marking schemes through quality-assured marking processes. Finally, during Awarding Meetings each year grade boundaries are set following a consideration of a range of qualitative and quantitative information, for the current year and the three previous years. The boundaries set are – Upper A (Band 1), Lower A (Band 2) and Lower C

[1] A Guide to Setting Grade Boundaries

(Band 6). All other grades and boundaries are automatically calculated based on these boundaries. There is no other mechanism currently used for setting grade boundaries. Our approach to Awarding is discussed and approved each year at Qualifications Committee.

The combination of the above activities provides SQA with the confidence to award graded National Courses.

SQA does not operate an explicit norm-referenced system where a fixed-proportion of grades is awarded each year. Awarding Meetings are held individually and there is no process to shape national level performance. However, the approach does result in a relatively stable national system – as outlined in the tables below. Subject-by-subject variability is acknowledged, for example, larger uptake qualifications are more stable

Table One: National 5 A-D Distribution (2016-2019)

National 5					
	A	**B**	**C**	**D**	**A - D**
2019	35.2%	23.8%	19.3%	12.5%	90.8%
2018	35.2%	23.0%	19.3%	12.3%	89.8%
2017	37.2%	23.8%	18.6%	6.5%	86.1%
2016	36.6%	23.7%	19.2%	6.6%	86.1%
Extended Grade D Scale from Diet 2018					

CONFIDENTIAL

Table Two: Higher A-D Distribution (2016-2019)

Higher					
	A	**B**	**C**	**D**	**A-D**
2019	28.5%	24.0%	22.4%	15.0%	89.9%
2018	28.7%	25.3%	23.1%	8.7%	85.8%
2017	29.0%	25.5%	22.8%	8.3%	85.6%
2016	29.5%	25.2%	22.8%	8.3%	85.8%
Extended Grade D Scale from Diet 2019					

Table Three: Advanced Higher A-D Distribution (2016-2019)

	A	**B**	**C**	**D**	**A-D**
2019	32.0%	24.9%	22.7%	8.3%	87.9%
2018	32.6%	25.9%	22.3%	7.9%	88.7%
2017	32.0%	25.6%	22.8%	7.9%	88.3%
2016	33.8%	25.9%	22.4%	7.4%	89.4%
Extended Grade D Scale from Diet 2020					

Follow-on Point

The year-on-year stability outlined in the above tables suggest that the setting of national distributions in 2020 based on a consistent historical-based approach is possible. However, subject-by- subject variation may need to be considered. The types of situations where this may not be appropriate, include the following:

1. Where we made a significant or material change to Course assessment in that period.
2. Where there was a significant adjustment in the period due to a non-functioning assessment component.
3. Where there was a "reset" of standard.
4. Where the actions of one centre with significant presentations in a small cohort served to skew the decision-making or distribution.
5. Low uptake Courses

However, making national-level adjustments to grades may disadvantage some centres that have not overestimated grades. Therefore, the first step in the awarding process should be an exercise to moderate centre estimates to address severity and leniency, as far as possible.
The last step in the awarding process should be a national awarding review of the resulting national distribution.

CONFIDENTIAL

3.2 National Courses

Graded National Courses (National 5, Higher and Advanced Higher) are assessed through a combination of Coursework and/or Question Papers as outlined in the relevant Course Specification. However, in prescribed circumstances other existing processes, e.g. Exceptional Circumstances, may overwrite these assessment requirements. There are 128 graded National Courses as outlined in Table One.

Table Four: Graded National Courses (Diet 2019)

Level	Number of Subjects
National 5	48
Higher	46
Advanced Higher	34

Presentation Patterns Across Centres

Entry patterns for the SQA National Course portfolio are characterised by high and low-uptake qualifications; and new, returning or very low-uptake centres. For example, in Diet 2019 national level entry figures varied across National 5 from 74 (Urdu) to 45,593 (English), across Higher from 61 (Gaelic) to 35,461 (English), and across Advanced Higher from 9 (Music Portfolio) to 3,706 (Mathematics). Advanced Higher has very low national entry numbers in some subject areas.

Additionally, even in large uptake subjects there can be significant variation. In Diet 2019, the largest number of entries at centre level for National 5 was for Mathematics, but entry levels ranged across centres from 1 to 319 (average 98). For Higher, the largest number of entries at centre level was for English, ranging across centres from 1 to 314 (average 87). For Advanced Higher, the largest number of entries at centre level was for Mathematics, with a range across centres from 1 to 61 (average 12).

CONFIDENTIAL

In Diet 2019, there were 21,488 distinct centre/subject/levels entered for National Courses (National 5, Higher and Advanced Higher). On the whole, 15,588 (73%) of these distinct centre/subject/level combinations had 25 or less candidates entered and resulted. Table Two breaks this down by level.

Table Five: Centre/Subject/Level Combinations (Diet 2019) By Level

	Overall	25 or less candidates
National 5	9,126	5,539 (61%)
Higher	8,077	6,176 (73%)
Advanced Higher	3,924	3,873 (99%)

For National 5, at least half of class entry sizes were made up of 19 or fewer candidates; for Higher, at least half of class entry sizes were made up of 14 or fewer candidates; and for Advanced Higher, as least half of class entry sizes were made up of 4 or fewer candidates. Table Three outlines the full cumulative distribution of entries for Diet 2019. Relatively small numbers of candidates distributed across many centres is likely to make it challenging to be able to make statistically significant decisions across centres and nationally in some low-uptake subjects. This has significant implications for the use of a purely statistical approach for centre moderation purposes.

4 Diet 2020 Decision-making Data

4.1 Estimates

Use of Estimates in Business-As-Usual Awarding

Other than the skills of the estimator, the 'accuracy' of Estimates is dependent on three main variables – valid evidence of performance in centre to inform the estimate; the application of candidates in their assessment; and the reliability of SQA marking, quality assurance and awarding. Estimates are currently used in Awarding Meetings as an indication of cohort ability. The weight placed on this will vary from subject to subject but, to date, it has not been a significant aspect of business-as-usual decision making.

Accuracy of Centre Estimates

Centre Estimates are not always accurate. Table Six outlines overall estimating accuracy at a national level for Diet 2019 based on the business-as-usual nine-point Band scale. Estimating performance does, however, vary across centres, subjects and levels. The table shows how close centres were to accurately estimating individual candidate attainment against what *each* candidate actually attainment in the SQA assessment. While the majority of candidates achieve within -/+ one band of their estimate, around a third are outwith this. As a result, this suggests that some form of further moderation of centre estimates is required to address under and over-estimation.

Table Six: Overall Estimating Accuracy (Diet 2019)

N5	Above Estimate			Same Band	Below Estimate		
	> 2 bands	2 bands	1 band		1 band	2 bands	>2 bands
	25412	32864	58380	88142	48178	19841	11105
	9.0%	11.6%	20.6%	31.0%	17.0%	7.0%	3.9%
	20.5%			68.6%		10.9%	

H	*Above Estimate*			*Same Band*	*Below Estimate*		
	> 2 bands	2 bands	1 band		1 band	2 bands	>2 bands
	17013	20608	35111	50596	32119	15738	10665
	9.4%	11.3%	19.3%	27.8%	17.7%	8.7%	5.9%
	20.7%			64.8%		14.5%	

AH	Above Estimate			Same Band	Below Estimate		
	> 2 bands	2 bands	1 band		1 band	2 bands	>2 bands
	2091	2282	3971	6249	3930	2301	2237
	9.1%	9.9%	17.2%	27.1%	17.0%	10.0%	9.7%
	19.0%			61.4%		19.7%	

CONFIDENTIAL

Refined Band Scale

In our business-as-usual approach, a nine-point band scale is used for centres to base their Estimates on and for purposes of SQA certification. As outlined above, a refined nineteen-point band scale has been introduced for Diet 2020 in order to more closely reflect mark distributions. Each of the business-as-usual bands has been split into two, except for lower A and D, which have been split into three to give more granularity at the decision points for national awarding purposes. This is mapped to the nine-

point band scale and is outlined in Table Seven.

Table Seven: Refined Band Scale

Grade	Band (BAU)	Refined band		Notional % range
A	1	1 Upper	1	93–100
A	1	1 Lower	2	85 – 92
A	2	2 Upper	3	80–84
A	2	2 Middle	4	75–79
A	2	2 Lower	5	70–74
B	3	3 Upper	6	67–69
B	3	3 Lower	7	65–66
B	4	4 Upper	8	62–64
B	4	4 Lower	9	60–61
C	5	5 Upper	10	57–59
C	5	5 Lower	11	55–56
C	6	6 Upper	12	52–54
C	6	6 Lower	13	50–51
D	7	7 Upper	14	47–49
D	7	7 Middle	15	44–46
D	7	7 Lower	16	40–43
No Award	8	8 Upper	17	35–39
No Award	8	8 Lower	18	30–34
No Award	9	9	19	0–29

Previous Studies

Studies from across the UK, for GCSEs, AS level and A level, about the accuracy of teacher estimates in comparison to actual results achieved, show similar trends to SQA's data.[2]

The findings about individual variables are broadly similar: subject has a small but unsystematic effect; gender and age have a small effects which are inconsistent across subjects; centre type has a small effect which can be speculated to be attributable to correlation between centre type and attainment. There are likely some effects on prediction accuracy of ethnicity (that is more over prediction for some ethnic minority groups) and disadvantage (that is more over prediction for the more disadvantaged in general and less over-prediction for the more high attainers)but those effects have not been properly estimated.[3]

4.2 Rank Ordering of Candidates

CONFIDENTIAL

In Diet 2020, centres are required to provide a rank order for each of their candidates within each refined band.

Unique rankings with no ties are expected within each refined band for most National Courses. This will help for large multi-class cohorts, we recognise that it may be extremely difficult to distinguish between every candidate and centres have been strongly advised to minimise the number of ties and the number of candidates who are tied, within any refined band[4].

Baird (1997) concluded, most of the evidence suggested that centres were good at rank ordering students.[5] SQA has not used candidate rankings in assessment decision-making for many, but it waspreviously used to form part of an alternative evidence appeals process. Ranking is still based on professional teacher judgement but removes the need for teachers to make specific grading decisions. In this approach teachers are being asked to rank candidates on their ability relative to one another. However, there are challenges of comparable decision-making within and across centres and if used as the only source of data this could result in candidates of equal ability intra/inter centres gaining different grades. Centre rank orders must therefore be linked to an Estimate in any approach. For example, a centre may have 100 candidates ranked 1 to 100 but the highest ranked candidate may be estimated at Band 6 / Grade C (current 1-9 scale) therefore this would have to be factored into how any historical distribution is laid over the centre rank order. The inclusion of 'ties' in a rank potentially hinders the efficacy of rank orders and we have advised centres to use ties sparingly, only for large cohort groups.

Follow-on Points

- Rank orders must always be referenced to a centre Estimate;
- Rank orders are only meaningful within individual centres;
- Rank orders should be used in the centre moderation process to assist in addressing differing patterns of leniency or severity at different points on the grade distribution
- ~~Rank ordering may also be~~ useful in Post-certification Review process

[2] British Educational Research Association, Volume31, No 1 Feb 2005, Teacher Estimates of candidates grades: Curriculum 2000 Advanced Level Qualifications, Debra Dhillon, AQA, UK
Scottish Examination Board (SEB) (1996) Teacher Estimates and SCE Examinations (Dalkeith, SEB)
[3] Ming WEI Lee, Merlin Walter, 2020 Ofqual Research And Analysis: Equality Impact Assessment Literature Review
[4] Information for Centres – Producing Estimates Session 2019-20
[5] Baird J A (1997) Teachers Estimates of A level Performance , Guilford, Internal Report RAC/763, Associated Examining Board

Ofqual are alsoexploring how rank orderings could be used in the awarding process for 2020. They are relying on the cohort level value-added relationship between candidate prior attainment and subject level grade. In this model, the overall relationship (typically used at cohort level) would be used to predict the grade outcomes for individual centres based on prior attainment profile for the current year. This would essentially force each centre's outcomes to match a single national value-added relationship. For each centre the only estimation data used would be the rank order. The centre distribution would be overlaid upon the rank order to determine candidates' grades. This is one of a number of approaches under consideration.

4.3 Prior Attainment

SQA Data on Prior Attainment

SQA only holds meaningful prior attainment data at Higher and Advanced Higher. Prior attainment is not available at National 5 or for all candidates in Higher and Advanced Higher due to bypassing and 'crash' courses. However, it is a useful predictor of performance where present. It is currently used as an indication of cohort ability in Awarding Meetings. The extent of prior attainment (at the subject/level immediately below) in Diet 2019 is summarised in Tables Eight and Nine.

Table Eight: National 5 Prior Attainment in Higher Courses (Diet 2019)

Year	Level	Percentage of Candidates	
		No Prior	Prior
2017	H	23.27	76.73
2018	H	22.74	77.26
2019	H	21.94	78.06

CONFIDENTIAL

Table Nine: Higher Prior Attainment in Advanced Higher Courses (Diet 2019)

Year	Level	Percentage of Candidates	
		No Prior	Prior
2017	AH	2.18	97.82
2018	AH	2.07	97.93
2019	AH	2.11	97.89

There are variations in prior attainment by subject area. Prior attainment volumes are higher for higher-uptake subjects. In some National Courses at Higher, it can be as low as 5% in Classical Studies. However, prior attainment – where available – will continue to be used as a general indication of cohort ability within a National Course at a national level and as an additional measure in any centre moderation activity.

Non-SQA Data on Prior Attainment

The approaches to awarding that are likely to be used in other parts of the UK in the summer of 2020 are heavily based and dependent on data on the prior attainment of candidates who would have sat examinations this year. A major source of this data is the regime of Key Stage 2 tests which has, as one of its main purposes, to produce a measure of cohort ability that can be used to inform setting of grade boundaries for general qualifications.

Whilst SQA has prior attainment data from those candidates at Higher and Advanced Higher who have previously sat qualifications at lower levels we do not have a source of prior attainment data comparable to the Key Stage 2 tests that could be used to inform awarding decisions at National 5.

The closest comparable assessment in Scotland to the Key Stage 2 tests is the Scottish National Standardised Assessments (SNSA), introduced in 2017. These assessments, in Literacy and Numeracy, are completed by learners in P1, P4, P7 and S3. Their primary purpose is '…to help identify children's progress, providing diagnostic information to support teachers' professional judgement' https://standardisedassessment.gov.scot/ This purpose means that SNSA assessments are not delivered at a fixed point in the school year.

The primary purpose of SNSA and the way it is administered means that it does not provide a useable source of prior attainment data for SQA as part of its Alternative Certification Model for Diet 2020. Teacher professional judgement (ACEL) will be used to inform the journey of learners from BGE into the senior phase and will take into account a wide range of other evidence including SNSA.

5 Consideration of Technical Options

Several technical options for the Alternative Certification Model have been considered to date and are outlined below.

5.1 Directly Awarding Centre Estimates

Description

In this case, SQA would directly award the Estimates submitted by Centre. All centres would simply be resulted with the current band 1-9 Estimate. No moderation of the Estimates would be undertaken.

Technical Work to Date

Preliminary work looked at estimated grades and resulted grades in Diet 2019.For estimates with a completion status of "C", 48% of grades estimated at National 5 matched grades resulted; at Higher, 44%; and at Advanced Higher, 43%. Estimate accuracy varied by grade with greater grade accuracy seen at "A" compared to other grade levels. It is important to understand the difference between candidate matched estimate/result data and looking simply at the overall picture at the end, as the overall figures may look close but accuracy at candidate level was moderate at best.

If Diet 2019 estimates had been awarded, *overall* the national level A-C rates would have been very similar to the national level A-C rates resulted for National 5 and Higher but the national level Grade A rates would have been much lower. This is because, while Grade A estimates were on the whole more accurate, generally lower numbers of Grade As were estimated than resulted in Diet 2019. However, we know that this averages out learners who achieved more than their Estimate and others who achieved less than their Estimate.

Table Ten: Estimates and Published Results

National 5:	A	A-B	A-C	A-D	NA
Estimates	27.8%	51.8%	78.6%	90.8%	9.2%
Published results	35.1%	58.9%	78.2%	90.7%	9.3%

CONFIDENTIAL

Higher:	**A**	**A-B**	**A-C**	**A-D**	**NA**
Estimates	23.8%	48.6%	77.0%	89.6%	10.4%
Published results	28.3%	52.3%	74.8%	89.9%	10.1%
Advanced Higher:	**A**	**A-B**	**A-C**	**A-D**	**NA**
Estimates	27.0%	56.1%	84.0%	92.7%	7.3%
Published results	31.8%	56.7%	79.4%	87.8%	12.2%

Note - Estimates percentages and published result percentages have been calculated on different groups of candidates, as some candidates do not have an estimate/result.

Subject level A-C rates and Grade A rates would have varied considerably – with some subjects seeing large percentage point increases or decreases overall.

For Higher, the top and bottom ten subjects which would have seen an increase/decrease in their overall A and A-C rate from resulting estimates in 2018/19 are presented in the table below

Table Eleven: Estimates and A; A-C Rate

Subject – Higher	**P.P diff A rate**
ESOL	13.9
Psychology	11.0
Sociology	6.6
Environmental Science	4.4
Fashion and Textile Technology	3.9
Music Technology	3.2
Health and Food Technology	2.7
Graphic Communication	2.5
Classical Studies	1.9
Dance	1.4
Politics	-6.7
Mathematics	-6.8
Gaidhlig	-8.1
Economics	-8.8
Music	-9.6
Accounting	-10.4
Physical Education	-10.8
Administration and IT	-10.9
Business Management	-12.3
Latin	-15.2
Care	37.9
Psychology	33.6
Design and Manufacture	24.2
Sociology	20.8
Health and Food Technology	20.6
Fashion and Textile Technology	18.5
Childcare and Development	16.9
Music Technology	15.3
Media	14.8
Engineering Science	14.3
Economics	-2.6
Biology	-3.0
Latin	-3.1
Gaelic (Learners)	-3.8
Politics	-4.8
Administration and IT	-4.8
Physical Education	-4.9
Mathematics	-5.9
Business Management	-6.5
Human Biology	-7.1

CONFIDENTIAL

Within the above, centre results varied considerably with some benefitting from over estimation and others self-penalised from under-estimation.

This perhaps reflects a business-as-usual environment where Estimates are arguably not as important as they will be in Diet 2020. It is likely that the higher stakes, increased efficacy of a more granular band scale and; the enhanced support information, training and additional requirements for centre approval of Estimates should all contribute to improved Estimating practice. This is a central element of the *Alternative Certification Model.*

Opportunities

- places trust in professional teacher judgement

Points for Consideration

- known variable accuracy in estimation by centres
- some candidates would get appropriate grades, some would get a lower grade then evidence merits and others would get a more generous grade than the evidence merits
- unlikely that A-D distribution will reflect any recent national distributions.

4.2 Multiple Linear Regression

A multiple linear regression model was developed and tested. This is a statistical technique that uses several explanatory variables to predict the outcome of another variable of interest. The model was developed using three variables: Estimates, Coursework marks and prior learner attainment to predict an overall course 'mark' for each candidate. The three variables are currently used in Awarding Meeting decision-making to varying degrees.

CONFIDENTIAL

The model was developed using Diet 2019 data as this is the only full component level data available. Five National 5 and Higher subjects were assessed: English, Mathematics, Chemistry, Physical Education and History – selected to cover a small number of different subject/assessment types. The preliminary statistical modelling indicated that, should the variables be available, they could not reliably predict candidate performance in the previous Diet for subjects/levels with high entry numbers and in the (unrealistic, optimal) 'complete' data case where all learners had estimates, prior attainment and coursework marks.

In all cases, *at least* 40% of predicted grades did not match actual grade outcomes in Diet 2019. Now that Coursework marks will not be available in Diet 2020 and, with prior attainment information varying from subject/level to subject/level, the use of a predictive modelling approach has not been progressed further. More explorative models based on those variables were not explored due to the data no longer being available with the removal of Coursework (e.g. similar models with additional centre level factors).

4.3 Awarding – National Moderation Only

Description
For each National Course, a national target grade and/or band distribution will be set based on agreed criteria. Centre estimates would be fit to this distribution, with all estimates of the same category receiving the same result (e.g. the Higher Mathematics distribution is set as Grade A: 30%, Grade B: 20%, etc. Starting with the first band, entire bands are added to the A category until the national subject/level distribution A percentage set is achieved or exceeded). This would be an entirely automated process with no moderation at *centre* level.

Technical Work Undertaken

This option was assessed on the basis of the original estimate scale (1-9) using Diet 2019 data and with target distributions set to actual Diet 2019 result distributions awarded as per August certification. The process assumed a 'target' distribution (here, set to be the Diet 2019 observed attainment distribution – the sensible target when testing the approach on previous data) and allocated each band in turn starting from the best band (band 1) to the A result until the target A rate was exceeded. At this point, allocations of a number of bands have been made to result 'A'. The next available band is then allocated to result 'B', and this process continues until the A-B cumulative 'target' distribution is exceeded. Similarly, the next available band is then added to the result 'C', and this process continued until the cumulative A-C 'target' distribution is exceeded, and so on. The allocations here could be automated (coded), and in this case always exceeded the target distributions.

The outcomes indicated that the original estimate scale was not sufficiently granular, which was not surprising given that SQA uses nine estimate bands, one of which contains the perceived strongest learners estimated to be in the 85%+ category, and two bands {8,9} that indicate 'No Award' levels of performance. This means that SQA really only has bands 2,3,4,5,6,7 to allocate to result outcomes {A,B,C,D}. Due to every estimated band being treated the same and the cumulative distribution rate required to be the same or exceeded with the approach used, the result was national subject/level distributions that were far from the starting targets in some cases. Below demonstrates the end result of a national moderation approach applied to Diet 2019 data, which allows the operation to be understood and the associated outcomes, albeit in a year where we only had 1-9 estimates and in the case where the target distribution is known (as we know what we awarded based on certificated outcomes, the 'target' here is the actual Diet 2019 august certification cumulative A, A-B, A-C, A-D distributions).

CONFIDENTIAL

Table Eleven: Change in Higher Grade Distribution

Subject			Percentage Point Change in Grade Distribution					
	Entries	Estimates	A	B	C	A-C	D	No Award
Chemistry	10,047	10,064	6.69	-1.26	5.84	11.27	-9.16	-2.11
English	35,461	35,408	5.85	1.39	-5.66	1.58	2.55	-4.12
History	9,987	9,989	8.13	0.52	-3.71	4.95	-1.01	-3.94
Mathematics	18,626	18,742	1.74	9.87	-4.07	7.54	-4.85	-2.69
Modern Studies	8,653	8,688	5.68	0.8	-3.38	3.08	-1.69	-1.39
Physical Education	9,896	9,931	2.86	0.59	3.79	7.24	-6.75	-0.49

The above demonstrates the use and limitations of the current estimate scale. Note that the Grade A and A-C rates are always higher here due to using the rule 'add more estimate bands until the cumulative distribution is exceeded'. It would also be possible to choose the allocation that led to allocations that were closest to the target thresholds (using absolute values). Exploring all the finer details of options was not explored as each approach required custom coding and the method is held to be adequately demonstrated above.

The refined estimate scale for Diet 2020 is more likely to result in national level distributions that are closer to the starting target distributions set, since each band of the refined estimate scale represents fewer candidates.

The main challenge with national moderation is that it cannot account for systematic over or under estimation/bias at a centre level, although checks could be built in after initial decisions are made to identify outliers. The outliers could then have their refined estimates adjusted up or down as necessary, and the national moderation process repeated using the adjusted refined estimates. Alternatively it could be more appropriate to carry out centre moderation first to address over and under-estimation and then to conduct a national awarding check on the resulting distributions.

Although the focus of this paper is not concerned with the use of statistical archives/data, the statistical archives are snapshots taken at specific time points and form datasets that were never intended to be used with respect to measures or decisions at individual data subject level. In summary, a combination of the centres based on the 1-9 band scale and current estimating practice would be unlikely to result in an acceptable outcome. The move to a greater number of estimation bands and the increased support provide by SQA to centres on their estimating practice should, all things remaining equal, mitigate this.

Opportunities

- places full trust in professional teacher judgement
- more managed SQA national distribution
- learners with same band nationally will be treated the same way
- automated process
- national level outcomes may look somewhat more like those of previous years
- allow an equivalent of the grade boundaries to be published as for each subject and level
- practical solution

CONFIDENTIAL

Points for Consideration

- variable accuracy in centre estimation
- limitations of the current nine-point scale
- ignores centre or local authority variation in estimation accuracy
- good cohort centres treated equally to overestimating centres
- some candidates would get appropriate grades, some would get a lower grade then evidence merits and others would get a more generous grade than the evidence merits

4.4 Centre-level Moderation – Using Previous Centre Outcomes

Description
For each National Course, a national target grade and/or band distribution will be set based on agreed criteria. Centre estimates would be fitted to this distribution, with all estimates of the same category receiving the same result (e.g. the Higher Maths distribution is set as Grade A: 30%, Grade B: 20%, etc. Starting with the first band, entire bands are added to the A category until the national subject/level distribution Grade A percentage set is achieved or exceeded). At the end of this phase, all estimates of the same value would have the same result in a subject/level. A second phase would then attempt to address centre outcomes that appear very different based on some historic centre level performance measure(s). Descriptive summary measures of outcomes will be produced for all centre subject/levels and will include year on year percentage point changes in resulted Grade A rate and A-C rate; and year on year differences in average Insight tariff scores and any other identified/agreed measures for consideration. It may be possible to produce inferential statistics where sample sizes are large enough, and assumptions are met. Given the entries within centres at subject level the use of these may be restricted.

Technical Work Undertaken
In the absence of Diet 2020 data, Diet 2019 data has been assessed. Estimates on the current band scale were fitted to the starting target distribution of the actual national level result distribution for Diet 2019 as above in national level awarding. At a centre/subject/level, the data that SQA holds historically allows the calculation of a number of summary values and approximate measures which may indicate the relative strength of performance and qualification attainment in previous sessions. These include:

CONFIDENTIAL

- The number of valid estimates received per year
- The number of resulted candidates entered per year
- The number of years of estimate/result information
- Average prior attainment information (Higher and Advanced Higher) at centre level by year, where available, which will contain the numbers & proportions obtaining each result and measures derived from that (such as average band).
- The no (and %) of candidates estimated by band/grade per year [estimate distribution]
- The no (and %) of candidates resulted by band/grade per year [result distribution]
- Centre estimation band accuracy per year (*for candidates estimated *and* resulted)
- The total (and average) calculated Insight tariff score based on estimates per year, using the insight tariff scores for course awards.
- The total (and average) calculated Insight tariff score based on results per year, using the insight tariff scores for course awards.
- Historic total assessment scores for candidates (Marks from previous years, although obviously no equivalent of marks will exist for Diet 2019/20 unless estimates are treated as such)

A large amount of development work has taken place to form suitable datasets containing such information and measures. Three main datasets were required to be formed:
- relevant record/candidate level data;
- candidate records across time to an aggregated centre/subject/level by year
- creation of centre level measures of the type described above for each subject/level.

The main comparison for this year in terms of ‘checks’ at the centre level are likely to be between the results obtained after the *Awarding – National Moderation* stage and those awarded historically to that centre, noting that the national moderation process is dictated by the submitted estimates. The options for flagging results for centre level moderation are presented in this context. Options include:

- ranking centres by difference from a three-year[6] historic average (e.g. difference in resulted A rate or A-C rate).
- ranking centres by difference in their average resulted Insight tariff score from a previous year/ three year historic average
- flagging centres with resulted A-C or A rates which are out with historic minimum or maximum ranges
- multiple criteria within the above options.
- alternative approaches.

The same subset of centres from previous [centres a-e] have been selected to demonstrate what the output of the national level awarding would have been in Diet 2019, and options for centre flagging for centre level moderation.

However, in principle, if SQA could determine measures to calculate in advance and how to prioritise these, centres could be flagged (eg this centre may or not be the most extreme change in points observed after national moderation and this sort of information could warrant further investigation), but does not indicate how to proceed or overrule any estimates.

For some of the centres with high estimates, it may be possible to calculate a chi-square statistic (or some goodness-of-fit measure) to assess whether the outcome distribution is significantly different from the three-year historic average distribution above for each centre. However, such approaches would require non-trivial entries, and non-trivial expected frequencies in each of the result categories to be considered. Even in these uncommon cases, lack of fit does not then give any information on acceptable ways to correct. Centres could be ranked on such differences and flagged for centre level moderation.

CONFIDENTIAL

There are challenges due to the number of combinations of centre/qual/level involved, problematic low sample sizes in a number of centres, missing historic data. This is likely to be challenging to automate and may require some form of involvement of qualifications staff. A fully automated process would require decision rules to be fully coded in a software solution to cover all scenarios.

Opportunities

- may address centre or local authority variation from the national distribution
- opportunities for automation

Points for Consideration

- variable accuracy in estimation
- limitations of the current nine-point scale
- may reward or penalise Diet 2020 cohorts based on previous centre/local authority performance
- no guarantee that learners with the same Estimated band will receive same Grade
- unlikely that A-D distribution will reflect any recent national distributions
- complex coding / decision rules required.

[6] SQA is using a three-year historical dataset of three million entries to inform Diet 2020 activity.

Table Twelve: Example of National Level Moderation – Higher Mathematics (Diet 2019)

Outcome of National Moderation								
Centre	**Estimates**	**A**	**B**	**C**	**A-C**	**D**	**NA**	**Ave. Insight Tariff Points**
A	73	12.30%	39.70%	27.40%	79.50%	1.40%	19.20%	143
B	175	57.10%	15.40%	8.60%	81.10%	0.00%	18.90%	158
C	77	19.50%	45.50%	16.90%	81.80%	18.20%	0.00%	177
D	91	48.40%	22.00%	17.60%	87.90%	11.00%	1.10%	183
E	213	72.80%	21.10%	3.30%	97.20%	2.30%	0.50%	196
Moderation Using Previous Centre Outcomes (2016 - 2018)								
Centre		**A**	**B**	**C**	**AC**	**D**	**NA**	**Ave. Insight Tariff Points**
a		No historical values for centre						
b		60.70%	21.40%	10.40%	92.50%	3.90%	3.60%	185
c		19.90%	21.00%	24.20%	65.10%	18.10%	16.70%	145
d		22.20%	21.00%	25.00%	68.30%	16.30%	15.50%	148
e		55.30%	23.30%	13.00%	91.50%	5.70%	2.70%	184
Percentage Point Difference between National Moderation and Previous Centre Outcomes								
Centre		**A**	**B**	**C**	**A-C**	**D**	**NA**	**Insight Tariff Points**
a		No historical values for centre						
b		-3.6	-6	-1.8	-11.4	-3.9	15.3	-26.9
c		-0.4	24.5	-7.3	16.7	0.1	-16.7	31.9
d		26.1	1	-7.4	19.7	-5.3	-14.4	35.3
e		17.5	-2.2	-9.7	5.6	-3.4	-2.2	11.2

CONFIDENTIAL

4.5 Awarding – Using Centre Supplied Rank Ordering

Descriptions
For each National Course/centre, a centre distribution is applied e.g. the Higher Mathematics distribution is set as Grade A: 30%, Grade B: 20%, etc. Candidates are allocated to grades from rank 1 to the final ranked candidate until quotas are filled.

Technical Work to Date
Diet 2020 is the first time for a number of years that ranking information has been collected. As a result, we do not have real centre rank order data to work on at this point. Work is continuing here to model this. Findings to date have indicated that rank orderings may be useful in terms of allowing a lower level of refinement in the allocation of results to learners, but doing this in a systematic way *that is able to reflect genuine changes to cohort strengths within a centre* leads to the same problem that it depends on knowledge of what the expected centre outcomes for that qualification should be i.e. this shares all the same challenges of the centre level moderation up approaches.

Essentially, if a centre level expected distribution of attainment could be reliably arrived at, ranking would be potentially useful in the allocation to that distribution. However, as outlined in previous sections, low-uptake qualifications and/or centres complicates this approach. Nevertheless rank orders will be useful in centre moderation approaches.

Opportunities
- provides more granular decision-making;
- facilitates differential adjustments across the grade distribution at centre level

Points for Consideration
- rankings cannot be compared across centres

CONFIDENTIAL

5 Approaches in England, Wales and Northern Ireland

SQA's approach to awarding this year has been informed from its earliest stages by an awareness and understanding of the approaches being adopted elsewhere across the UK and by regular engagement and information sharing with the other UK regulators. Our approach from the outset has been to be consistent with other UK regulators where this is appropriate and to develop approaches that reflect the particular circumstances of the Scottish educational and qualifications system where necessary. All four countries of the UK face challenges in awarding this year in the absence of performance data. However in Scotland, SQA faces some additional challenges to those faced by Ofqual in England. Centre cohorts in Scotland are often small, which precludes using ranking data for many subjects/centres; there is no reliable prior performance data at all for National 5, and not always at Higher and Advanced Higher either.

Whilst the specific details of the approaches being developed by the other regulators are still under development, the broad principles of the approaches being adopted have been made public. The following sections contain a high-level overview of these approaches.

5.1 England

In its consultation '[Exceptional arrangements for exam grading and assessment in 2020]' Ofqual has stated the aims of its standardisation approach to the awarding of qualifications in 2020 as follows:

- to provide candidates with the grades that they would most likely have achieved had they been able to complete their assessments in summer 2020
- to apply a common standardisation approach, within and across subjects, for as many students as possible
- to use a method which is transparent and easy to explain, wherever possible, to encourage engagement and build confidence
- to protect, so far as is possible, all students from being systematically advantaged or disadvantaged, notwithstanding their socio-economic background or whether they have a protected characteristic
- to be deliverable by exam boards in a consistent and timely way that they can quality assure and can be overseen effectively by Ofqual.

CONFIDENTIAL

To support achievement of these aims Ofqual has published guidance indicating that for every candidate to be certificated centres will be required to submit to exam boards:

- A 'centre assessment' – reflecting the centre's judgement of 'the grade that each student is most likely to have achieved if they had sat their exams.'
- the rank order of students within each grade

In its consultation Ofqual states that its statistical standardisation will draw on the following sources of evidence:

- historical outcomes for each centre
- the prior attainment (Key Stage 2 or GCSE) of this year's students and those in previous years within each centre
- the expected national grade distribution for the subject given the prior attainment of the national entry

The consultation explores the implications of placing greater weight on each of these sources of evidence and the likely impact on the resulting grades awarded. Indicating its likely preferred model, the consultation asks that respondents give feedback on the question that '….using an approach to statistical standardisation which emphasises historical evidence of centre performance given the prior attainment of students is likely to be fairest for all students.'

The consultation also includes questions relating to the inclusion of information on the centres' 'trajectory' i.e. the trend in centre performance over recent years and to the most effective approach to mitigating the risk of bias in the standardisation model.

From our ongoing discussions with Ofqual we are aware that they have not yet settled on a final statistical model or models for the awarding process and are continuing to work with awarding bodies

to explore several options. We expect to receive further information on their evaluation of these options and final choices of model(s) as their thinking develops.

5.2 Northern Ireland

Council for the Curriculum, Examinations & Assessment (CCEA) has advised its centres that it will also use a combination of historical performance data together with centre-generated grades and student rank order information within those grades to award qualifications in 2020.

CCEA has also stated that it will subject all centre-generated grades through a process of standardisation using a model being developed with technical and statistical experts.

As with Ofqual no further details of this model have yet been made public.

CCEA has specified that the approach to each qualification type will be as follows:

- A Level – a calculated grade, based on statistical predictions informed by previous student performance combined with centre assessed grades and centre rank orders
- AS Level - the grade awarded will be based on previous student performance combined with centre assessed grades and centre rank orders
- GCSE – grades generated through a combination of average centre performance, centre assessed grades and rank ordering

Published information states that 'awarding organisations are working towards the model on which grades will be awarded and in which centre assessed grades and centre rank orders will be used'.

5.3 Wales

CONFIDENTIAL

In its document 'Information for Centres on the Submission of Centre Assessment Grades' Qualifications Wales states that it too requires centres to submit a centre assessment grade based on professional judgement the rank order of candidates within each grade as the basis for its 2020 awarding decisions.

No further information has been made publicly available on the models or approaches Qualifications Wales plans to use to derive final grades from this information provided by centres. We are aware from our ongoing dialogue that they are considering similar models to those being evaluated by the other regulators.

6 Next Steps: Centre-level Moderation and Awarding

Our proposed approach is to carry out an initial assessment of centre estimation accuracy, then to carry out centre moderation and to finish with a national awarding check at the end of the process.

.

6.1 Initial Assessment of centre estimation 'accuracy'

When centres submit their estimate and rank order information on 29 May we will do an initial piece of work to assess at a national and centre level the 'accuracy' of centre estimations. This will be done by comparing this years estimates against previous years estimates and actual attainment at national and school level.

Work to date with the education system has focused on enhancing the accuracy of teacher estimates for 2020. SQA has shared the past three years data of estimates vs actual results attained with centres. This will enable centres to assess which departments have historically had a tendency to over or under-estimates and if so at which parts of the grade distribution. SQA has also provided information for centres to assist with this process. In this information we have asked centres to quality assure their estimates at both departmental and whole centre level. For state schools, there is also a key role for each Local Authority to play in quality assuring estimates. However, we will not know what the impact of these measures is until we receive the data and do the initial national and centre level checks.

6.2 Centre-level Moderation Approaches

Centre-level Moderation Assumptions
Centre level moderation of Estimates will consider a combination of two factors – *Consistency of Estimating* and *Outcomes-based.*

Consistency of Estimating
This involves a consideration of how a centre has been historically estimating candidate performance in a qualification over the past three years and whether the Estimates submitted for Diet 2020 are in line with this. Centres will be provided with this information in early May 2020.

Outcomes-Based
This involves a consideration of centre attainment (including grade distributions) in a subject/level over the past three years. This is largely a measure of historical cohort ability.

As outlined in Section One, the challenges of relatively low-uptake qualifications at national and centre-levels, an absence and lack of use of standardised prior attainment information and finally, no assessment data means that SQA is limited in the predictive statistical approach SQA can deploy in Diet 2020.

This does not mean that statistical information and analysis is not central to our approach in Diet 2020. However, it does mean that it will need to be deployed based on consistent decision rules (or

heuristics), the use of SQA professional expertise and, liaison with schools, colleges and/or local authorities, where their attainment patterns may not reflect previous patterns.

To support the above, SQA statisticians have created dataset of around three million entries has been developed to inform Awarding and Centre-level in Diet 2020. This is comprised of graded National Course information from the past three years. This will be developed further during May 2020 involving both SQA statisticians and external data analytics expertise and more advanced tools. It will build on the work undertaken to date and outlined in previous sections of this report. The development of a set of *Centre-level Moderation Models* is central to our approach. The work will address:

- Exploration of a wider range of statistical techniques to compare centre performance over time;
- Approaches to bringing together low-uptake centres for centre performance comparisons over time, for example, at a whole school and local authority level;
- Consideration of inclusion of prior attainment;
- Consideration of use of centre rank orders;
- Consideration of centre level prior attainment;
- Consideration Of centre-estimating accuracy;
- Development of consistent decision-rules (heuristics) to inform centre-level moderation practice.

Additionally, we will consider how the following may be used, for example, as complementary evidence in our approach:

- Identification of relevant qualitative information;
- Contribution of Senior Appointees.

Our proposed approach is outlined below.

CONFIDENTIAL

Centre-level Moderation

To address the above, centre-level moderation will take place to address under and over-estimation – essentially, to identify centres who are under or over-estimating in terms of previous centre attainment and/or previous consistency of estimating. Genuine changes in performance due to different cohorts is a reasonable possibility and year to year centre variability is not unreasonable particularly when entry numbers are low/moderate. The analytical work in the development of the Centre-level Moderation Tools will inform what valid criteria can be used to determine under or over-estimators – based on the information that SQA has available.

The real challenge is identifying where centres have overestimated (or underestimated). With such small cohorts, this will be very hard to do. If a statistical check was carried out of this year's performance against an average of prior years' performance, then centres with fewer than ten candidates would probably not show as statistically significantly different to expectations, unless the swings in grades were extreme. Given 29 per cent of N5, 38 per cent of H and 85 per cent of AH are in centres with ten or fewer candidates, then it is likely to be challenging to identify overestimation (or underestimation) on a subject by subject basis. However, it might be possible to explore level of estimation accuracy by aggregating subjects up for a whole centre, or it might be possible to do for larger entry subjects in a centre and then make an assumption for other subjects that the same pattern would apply and look to see if there is additional evidence which supports this conclusion (such as prior attainment information, or information on standardised differences from average performance). It may also be possible to aggregate the data up to Local Authority level.

Using a centre level moderation method where all candidates are adjusted by the same amount in the same direction is also potentially unfair. For some centres, candidates would achieve no level 9s or no level 1s (depending on direction of adjustment). Clearly there will be candidates in some of these centres for whom this is deeply unfair. Whilst changes in these levels 9 and 1 s make no difference to the final A to D grades that the candidates receive, it could make a difference at more critical (for the candidate) level and grade boundaries such as 7 to 8, as candidates would be receiving no award. This may well be hard to justify for some candidates. Also, if a centre estimates leniently at critical grades and severely elsewhere, where it matters less to the candidates, then the overall outcome may well be no adjustment. In this instance it could be argued that the grades would not be a fair reflection

of candidate performance. This is why we believe that making use of the rank order, where possible, could give fairer results to candidates. It breaks the rule that all candidates with the same fine level would receive the same grade, but it does allow a more nuanced adjustment approach. It is hard to see that it is fairer to adjust everyone by the same amount in a centre, compared to just adjusting those students that the centre deems to be least secure in the level. The data actually shows that historically certain grades are underestimated and others overestimated (e.g. AH underestimate at A and overestimate at C), so applying a flat adjustment to all candidates in a centre is unlikely to correct this issue. This seems to us to be a justification for not applying the same adjustment to all candidates in a centre for a subject[1].

There is substantial complexity in establishing the "best" approach to detecting centre deviance. We expect that such an approach will incorporate multiple checks and complex rules to determine which should be used in which situation (in addition to incorporating prior attainment information where available). We have laid out some options in the main body of this report.
Finally, and in summary, it is difficult for SQA to say definitively what its approach should be to centre moderation before it has received estimation data back from schools. There is no single approach or model which will be appropriate across the board; a mixed economy of approaches should be considered with ranking data and prior attainment data used where possible, falling back to simpler models where this is not possible. SQA will flesh out further the range of options, and only determine how these will be used once the data has been returned and initial analysis carried out.

A possible set of statistical test/decision rules are set out in Appendix 1 to be used with decreasing cohort sizes. These will be explored further in the next round of development of models and testing using 2019 data sets.

Engagement
Based on the above centre-level moderation exercise, SQA will explore if it is feasible, within the time available, to engage with schools, colleges and/or local authorities to discuss any reasons for these differences.

6.3 Final Awarding

Any changes to the Estimates submitted by centres will be made based on agreed, consistent set of decision rules. The final band/grade distributions would be signed-off through a national awarding meeting as a final check following on from the application of the centre moderation models and the subsequent changes made.

[1] There may be other data which could be used to support an adjustment outside the straight, statistical data. The Ofqual consultation on vocational qualifications (VQs) (https://tinyurl.com/y9yem9x6) is arguably more nuanced and more like the challenge that SQA face than the GQ scenario. SQA may be advised to look at the approach for VQs in England.

6 Recommendation

The Code of Practice Governance Group is asked to discuss and agree the proposed way forward outlined in Section 6 and the next steps outlined below.

Next Steps

Outlined below are a suggested set of next steps for discussion:

- We are currently using two sets of contractors (Alpha Plus and a data analyst organisation) and we need to agree quickly how we best deploy both to help us to develop and test our centre moderation set of models;

- We need to have a cross-organisational discussion about the implications of suggested changes and impact upon delivery timelines and risks and how we mitigate those. Could we not develop, code and test the layers of centre moderation approaches but then delay decisions about exactly how deploy until we have the data? If we can't then need to understand the risks associates with developing, coding testing the layered model for centre moderation before we have the data and then consider how we mitigate those risks;

- We need to develop a communication plan around this – some key points in the Alpha Plus report for us to consider eg acknowledging the volume of low uptake qualifications and centre uptake that Scotland has;

- Advisory Council and Qualifications Committee are on Wednesday of this week and we need to agree what we will present to each of these groups.

Appendix 1

Options for statistical comparisons of centre-level performance in 2020 vs prior years
Below, we present a summary of statistical methods we consider plausible tests to detect unusual centre grade estimates with in 2020.

Table 1: Summary of possible statistical tests/decision rules, by decreasing cohort size required

Test	Cohort test conducted on	Required data	Required cohort size (per subject)	What the test tells us
Statistical test on 2020 vs prior performance	Centre*subject combination	Historic centre performance	High	If a centre's 2020 performance is significantly higher/lower than in prior years, for this subject
Statistical test on 2020 vs prior years' value add	Centre*subject combination	Current & historic centre value added (i.e. prior attainment vs achieved grades)	High	If a centre's 2020 value add is significantly higher/lower than in prior years, for this subject
Decision rule on standardised difference between 2020 and prior performance	Centre*subject combination	Historic centre performance	Moderate	If a centre's 2020 performance is a pre-determined number of SDs higher/lower within the population distribution than in prior years, for this subject
Decision rule on standardised difference between 2020 and prior years' value add	Centre*subject combination	Current & historic centre value added (i.e. prior attainment vs achieved grades)	Moderate	If a centre's 2020 performance is a pre-determined number of SDs higher/lower within the population distribution than in prior years, for this subject
Statistical test on 2020 vs prior performance	All subject outcomes aggregated within a centre	Historic centre performance	Low	If a centre's 2020 performance is significantly higher/lower than in prior years, across all subjects
Statistical test on 2020 vs prior years' value add	All subject outcomes aggregated within a centre	Current & historic centre value added (i.e. prior attainment vs achieved grades)	Low	If a centre's 2020 value add is significantly higher/lower than in prior years, across all subjects
Decision rule on standardised difference between 2020 and prior performance	All subject outcomes aggregated within a centre	Historic centre performance	Extremely low	If a centre's 2020 performance is a pre-determined number of SDs higher/lower within the population distribution than in prior years, across all subjects
Decision rule on standardised difference between 2020 and prior years' value add	All subject outcomes aggregated within a centre	Current & historic centre value added (i.e. prior attainment vs achieved grades)	Extremely low	If a centre's 2020 performance is a pre-determined number of SDs higher/lower within the population distribution than in prior years, across all subjects

Given that SQA's two main sources of data with which to verify 2020 grade predictions against are historic centre outcomes, and this combined with prior attainment information to produce historic value added (though this is much more limited in availability), these are the two variables on which it is plausible to statistically check whether a given centre's grades are unusually high or low in 2020.
As the "required cohort size" column shows, statistical testing requires a larger sample than using standardised difference (i.e. z-scores) to check whether a centre has substantially shifted rank within the population distribution, but is less robust as it only takes a point estimate into account rather than the entire distribution[2].
Similarly, in the numerous cases where there are <10 candidates per subject within a centre, it is possible to aggregate up subject-level statistics to perform a centre-level test. These tests are unable to detect cases where teachers of one subject have been lenient and other severe within a centre, but allow the detection of overall severity/leniency within a centre across all subject. This option requires further consideration by SQA, and its use may depend on the nature of entries within a centre. If, as may be the case for AH, a centre has almost exclusively very small entries for each subject it offers, then this approach seems at its most useful, as it allows a means of verifying results when other approaches would not be plausible. However, for more niche subjects at the lower levels, where certain subjects (Maths for example) see whole school cohort entry, an aggregated test will be biased by the larger cohort subjects. To counteract this it may be desirable to exclude larger cohort subjects from aggregate testing. Another option would be to implement both a statistical test on a subject-aggregate at the centre level, and a decision rule on standardised difference within each subject. I.e., if a centre appears to be lenient overall across all subjects, then only the subjects where the standardised difference exceeds a particular amount should be adjusted/moderated, so as to allow for the fact that only some subject teachers may have mis-estimated grades.
The final two lines of the table, on decision rules at the centre aggregate level, seem likely to be most useful in centres with extremely few candidates, where it is not possible to perform any of the other tests on the centre aggregate, let alone individual subjects.
A limitation of these statistical checks is that it is impossible to apply them to new centres' estimates[3]. There is unfortunately little that can be done about this, and as above, the same issue exists for small private centres in England. SQA will discuss this issue with Ofqual to see what, if any, action the English Awarding Organisations are taking in such instances.
As this discussion and the table above shows, there is substantial complexity in establishing the "best" approach to detecting centre deviance. We expect that such an approach will incorporate multiple checks and complex rules to determine which should be used in which situation (in addition to incorporating prior attainment information where available).

[2] Assuming a statistical test accounting for the distribution shape is utilised, such as Kolmogorov-Smirnov.
[3] A quite complex comparable outcomes-esque approach could be workable using just prior attainment information for candidates in a new school, but given the transience of this information we do not discuss this approach in depth.