

RESEARCH AND ANALYSIS

Teacher Assessed Grades in summer 2021: interviews

Analysis of interviews with teaching staff and students

ofqual

Authors

- Steve Holmes, Asteria Brylka, Nathan Case, Latoya Clarke, Emma Howard, Ellie Keys, Diana Tonin and Stuart Cadwallader of the Standards, Research and Analysis Directorate.

With thanks to

- The teaching staff and the students who gave their time to speak to us in depth and share their experiences and views.

Contents

Authors	2
With thanks to.....	2
Executive summary	7
The teacher assessed grades process	8
Design	8
Evidence	9
Deciding and quality assuring TAGs	9
Concerns and issues	11
Introduction	13
Previous research	15
The current study	18
Differences between TAGs, CAGs and normal assessment	18
Change in the role of teaching staff	18
Experience	19
Anonymity	19
Methods	20
Overview.....	20
Interviews	20
Participants	21
Teaching staff	21
Students	24
Analysis approach.....	26
Structure of the interview analysis	26
The TAG process	27
Other considerations in the TAG process	27
Overall confidence in the TAG process	27

Beyond the TAG process	27
A note on terminology.....	27
The TAG process	29
Opinions on changes to the assessment arrangements in January.....	29
Design of the process used by centres to award TAGs	30
Guidance from government and awarding organisations	30
Pre-TAG activities and considerations	33
Evidence used to make judgements.....	37
Pre-guidance evidence	37
Post-guidance evidence	39
Student experience of the evidence-collection process	52
Evaluation of evidence and Internal Quality Assurance.....	57
Marking of individual pieces of evidence	57
Grading individual pieces of evidence	65
Calculation and/or judgement of TAGs within departments	70
Quality assurance of TAGs	77
The vocational and technical qualifications TAG process	80
Timelines and detail of guidance for TAGs in VTQs	81
Combining evidence and arriving at final TAGs for VTQs	82
Contrasts between Awarding Organisations	84
Other considerations in the TAG process	86
Fairness and minimising bias.....	86
Information, training and awareness	86
Reasonable adjustments, special consideration, and access arrangements.....	87
Access arrangements	88
Special consideration	90
Issues relating to specific types of student	91
Private candidates	92
Pressure experienced by teaching staff	93
The burden of determining TAGs	94

External quality assurance.....	96
Personal responsibility	98
Pressure from SLT.....	99
External pressures from parents, students and the media.....	100
Student well-being.....	101
The period of uncertainty.....	101
The assessment period.....	103
Fears for the future while awaiting results	106
Overall confidence in the TAG process.....	108
Perceived validity of the teacher assessments	108
Assessment conditions	108
The evidence used	110
Manageability of the assessments.....	111
Content coverage	112
Fairness across students with different patterns of learning or effort.....	113
The role of teacher judgement	114
Perceived consistency of the TAG process	116
Within-centre consistency.....	116
Between-centre consistency.....	118
Overall view of the TAGs.....	121
Looking beyond the TAG process.....	124
Readiness of 2020-2021 cohort	124
General preparedness for next steps	124
Loss and gain of skills.....	125
Expertise gained from the process	126
Discussion	128
Confidence in TAGs, but a heavy workload.....	128
Consistency across centres	129
Amount of testing.....	130
Upskilling following the TAG process.....	131

Teacher Assessed Grades in summer 2021: interviews

Sample strengths and limitations..... 131

Comparison of the survey and the interviews..... 132

Overall perceptions of TAGs..... 133

References.....134

Executive summary

In January 2021 the government confirmed that summer 2021 assessments could not go ahead as planned due to the coronavirus (COVID-19) pandemic. The decision was taken that students were to be awarded grades for general qualifications (GQs: mainly GCSEs, AS, and A levels) and many vocational and technical qualifications (VTQs: for example BTECs, applied generals) using teacher judgements. The intention was that these teacher assessed grades (TAGs) were to be based on evidence produced by the students that could be externally quality assured. Only content that a centre had been able to teach was to be assessed, and a variety of types of evidence could be used to support the holistic judgements centres were asked to make.

To support evaluation of the effectiveness and impact of the assessment arrangements in 2021 and to inform contingency planning for 2022, we carried out a project consisting of surveys and interviews of teaching staff involved in judging TAGs, and students receiving TAGs. This report details the interviews with teaching staff and students carried out between the submission of TAGs to awarding bodies (on 18th June for GQs and some VTQs) and results days (on 10th August for AS/A level and many level 3 VTQs, and 12th August for GCSE and many level 1/2 VTQs).

We interviewed a sample of 39 teaching staff who were selected to represent a range of job roles, qualifications and centre types. We also spoke to all 14 students that agreed to participate. Although only a limited number of students responded to the invites, those that did represented a range of year groups and centre types. All interviews used the same set of questions, but interviewees were free to expand on any of the topics if they wanted to. The interviews took place in the weeks prior to results being issued and therefore reflect the views of students and teachers at that time, before individual results were known.

It is important to note that this report reflects the views and experiences of a self-selecting sample of respondents - those who indicated a willingness to speak to us after completing the online survey and then agreed to be interviewed when contacted. This is normal in any voluntary qualitative study. However, because of this, we cannot assume that our samples are fully representative of the national population of teaching staff or students in all respects. Despite this limitation, the research still represents a wide cross-section of views and experiences. It allows us to consider the ways in which TAG processes may have differed in practice, in the context of specific schools.

The teacher assessed grades process

Design

Because of the limited time available to carry out the process and the different circumstances each centre faced, the guidance issued ensured that grading decisions would be based on robust evidence but allowed centres reasonable flexibility in how they designed the process and the evidence they used to determine TAGs. According to the teaching staff we interviewed, it was usual for senior management to take the lead in determining the design of the teacher assessment process, in terms of the evidence that should be used and the controls that should be in place. Qualification-specific decisions around the design of assessments and the way the evidence should be weighted were usually delegated to departments, where the head of department took the lead. A few centres allowed departments much more leeway in terms of the kinds of evidence and controls they used, so long as they were consistent with published guidance. Centres also sometimes worked with each other, sharing and discussing their ideas and approaches.

The centre policy (a document produced by the centre to record their process for determining and quality assuring TAGs) was used and perceived in a variety of ways by those we spoke to. Many used it to help them with the planning and design of the process, others used it to document their decisions, while some viewed it largely as a 'box-ticking' exercise.

Teaching staff we spoke to suggested that the timescales for release of guidance, exemplification and assessment materials sometimes caused them problems, as they had to make a decision about whether to go ahead with their planning or delay until the required documents were published. Senior managers reported that they felt unable to share their plans with teachers, parents and students, who were anxious to know what the process for determining TAGs would be, until final guidance had been published. This was reflected in the student interviews, where a great deal of uncertainty and worry was reported during the period between exams being cancelled and the plans for TAGs being announced.

Overall, the teaching staff we interviewed were generally happy with the teacher assessment process they designed and confident in the grades they had determined. They spoke extensively about the lengths they went to in order to collect the best evidence they could and to produce outcomes that were reflective of the student's abilities. This effort resulted in a high workload.

There were also some specific concerns staff spoke about. Assessment materials provided by awarding organisations were usually perceived to be insufficient because they did not include sufficient new assessment tasks. There was also concern that, because it was decided that materials should be available for all to access, motivated students could learn mark schemes and practice responses.

Because of this, many staff we spoke to described how departments spent a considerable amount of time selecting questions they thought students were less likely to have practised. At the same time, they were also trying to design complete assessments that were both representative of taught content and of appropriate demand. Students also commented in interviews that the availability of the materials meant that, in their experience, some students had been learning mark schemes and pre-prepared answers.

Evidence

The guidance released by Ofqual for GQs was designed to be flexible, to allow centres to choose how best to assess their students. Tests under exam conditions were by far the most common type of evidence and carried the most weight. While some centres used a small number of exam-length tests, it was more common for departments to run a series of shorter, often topic-based, mini-tests. Teaching staff that preferred this approach reported that it allowed them to account for any topic that had not been taught and to fit the assessments around the teaching schedule. Other sources of evidence such as mock exams, homework, classwork and coursework also contributed to the grades. Evidence collected under less controlled conditions (or before students knew such evidence might count towards their final grade) was often used to check that the results of later tests were not significantly out of line with the normal performance level of the student.

The marking process that interviewees described, including standardising, anonymisation, multiple marking, and moderating, was carried out slightly differently in every centre. Individual student work was often marked by multiple people. Staff spoke extensively about how they did their best to implement some of the procedures and processes that AOs would use for external exam marking. This was a major effort within departments, and in many the expertise of staff who were experienced examiners or moderators was important. Some departments did not have this expertise, and there was some suggestion that more help in terms of training could have been provided by exam boards.

Deciding and quality assuring TAGs

The approaches that centres took to determining their TAGs for GQs, based on the evidence that they collected, differed widely. Those teaching staff we interviewed often described using quite a numerical or analytic approach for determining their initial TAGs. Individual pieces of evidence were usually graded, sometimes using grade descriptors, sometimes through marking against fixed grade boundaries. Such boundaries were derived in a variety of ways, such as on the basis of whole past papers, by taking an average of several past-papers, or by analysis of the individual questions used in the tests. This last approach was particularly true for the shorter, more bespoke tests that centres used.

The resulting set of grades for the different assessments was then used to decide the TAG. This could involve taking a weighted average across the grades, identifying the highest grade a student consistently demonstrated, or concentrating on the most recent grades obtained in the more controlled tests (those that took place under exam conditions).

A minority of those interviewed worked with marks rather than grades, and weighted the marks to generate an overall score that was then mapped to a grade (often based on the grade descriptors). Almost all of the interview sample reported that they used their professional judgement to take a holistic view and that they would adjust initial TAGs, or re-weight different types of evidence, if the initial process resulted in TAGs that did not look right based on their knowledge of how they would expect students to perform.

Grade descriptors were provided by the JCQ to support teachers with their TAG judgements for GQs. It was expected that they would be used to support grading decisions by detailing the expected level of performance at specific grades in a subject. Despite some limited examples of their use, these grade descriptors were often not considered to be useful for grading individual pieces of evidence, as they were generally perceived to be too vague, or not appropriate for most subjects. However, they were thought to be more useful for deciding or checking overall TAGs. Some centres rank ordered their students based on the evidence collected and then worked down this rank order, considering the overall student performance against the grade descriptors to allocate groups of students to grades. At other times the grade descriptors were useful as part of a holistic check of TAGs that had been determined through more analytic approaches, or for supporting judgements for students who were close to grade boundaries.

All final TAGs were the work of multiple staff. They were quality assured by departments carrying out moderation and discussing and debating the TAGs. In addition, senior managers of the centre were always involved in scrutinising and querying both individual grades and profiles of grades within qualifications. The latter often involved a comparison of individual student TAGs across subjects. The interviewees commonly reported the use of results profiles from past years and described various pressures from senior management to adjust TAGs, sometimes up, sometimes down. The majority of class teachers we interviewed did not experience any direct pressure to adjust their TAGs based on the performance of previous cohorts, but were often asked by management to defend or explain their TAGs. A minority did report pressure, or described how their grades were revised by management during the centre's internal quality assurance process.

Finally, for those interviewees involved in judging TAGs for VTQs, the determination of TAGs largely seemed to be straightforward because the modular nature of most VTQs meant that fewer additional assessments were required to support TAG decisions. Many unit assessments had already been completed, and so it was often only a sub-set of units that required specific evidence collection for TAGs. The staff

were also well versed in carrying out internal assessment in many VTQs and felt that determining TAGs was not entirely different from normal practice. The design of the process was more often delegated to departments than it was for GQs due to the diversity of qualification designs. Tight timescales, and a lack of consistency of process and messaging across awarding organisations were often raised as issues in this context.

Concerns and issues

The main concern that came up in the interviews was around comparability between centres. While teaching staff viewed their own TAGs as valid and reliable, both staff and students often spoke about less well controlled, or even inadequate, systems that they believed other centres had used. For example, there were concerns from both teachers and students about the potential for other centres to provide too much help to their students and therefore give them over-generous grades. A few students in the survey also referred to variation across subjects or classes in their own centre, with some classes being given advanced notice of test content while others weren't.

We cannot be certain how common such practices were, as they were not usually experienced first-hand by those who contributed to our study. As noted above, our interviewees represent a self-selecting sample, and may be drawn predominantly from centres that took a careful, thorough approach to TAGs. Regardless, teaching staff in interviews were not confident that the external quality assurance process would ensure consistency across centres. Some thought that the most extreme cases might be detected and corrected, or that the threat of the quality assurance process would have encouraged most centres not to abuse the system.

The biggest negative reported was the amount of time and effort required and the resulting stress for staff. Teaching staff spoke at length about how running their own exam series, as they saw it, from start to finish took a major effort, particularly in the context of their other responsibilities. There was a general feeling that support, through guidance, information and assessment materials could have been better, and could have arrived sooner, from all official bodies and awarding organisations, particularly the exam boards. Stress levels were reported to be high for most teaching staff, as they were for students, because of uncertainty about what was going to happen.

Students we spoke to also talked about the pressure that came from the intensity of the assessment process for TAGs, particularly in GQs. Many experienced more testing than they would have if external exams had gone ahead, a result of centres wishing to collect a large pool of evidence to support their process for determining TAGs. The causes of stress may have differed across the various stages of the process, but students reported sustained levels of stress and pressure, sometimes affecting their well-being.

There appeared to be little concern around bias against groups of students based on protected characteristics or socio-economic status. This may have been because awareness of unconscious bias was high, a result of the experience of undertaking processes to minimise bias while judging CAGs in 2020, alongside more formal training this year than was possible in 2020. The fact that TAGs were based on evidence produced by the students themselves was also felt by teachers to reduce the risk of bias. However, some student interviewees were a little concerned about the possibility of bias against students other than themselves who might have had a difficult relationship with their teachers (though none had experienced bias personally).

Teaching staff did, however, sometimes report having difficulty deciding TAGs for students who demonstrated inconsistent performance. There were also concerns that those students who were good at high-stakes, final external exams would be disadvantaged, though it seemed a less significant issue this year, since most students were still assessed through exam-like methods.

Overall, these findings support our evaluation of the impact of the assessment arrangements that were used in 2021 on both teachers and students, helping us to understand their experiences and what they did in practice. They have also been used to help inform contingency planning for 2022.

Introduction

In January 2021, the government announced that GCSE, AS and A level exams would not go ahead in the summer as planned because of the disruption to students' education caused by the pandemic. Likewise, it was the government's policy position that it was not viable for timetabled exams and assessments for many vocational, technical and other general qualifications to take place. Following this announcement, schools and colleges began planning for the process of determining teacher assessed grades (TAGs). [Guidance was published by Ofqual](#) on 24 March ([and subsequently by JCQ](#) and awarding organisations) and while much planning and discussion took place before Easter, the main collection of evidence for TAGs began in schools following the return to face-to-face teaching after the Easter holidays.

TAGs were to reflect the grade level at which students were working, based only upon content that had been taught by the centre in each of their qualifications. This process covered all general qualifications (GQs) and many vocational and technical qualifications (VTQs), with the intention being to allow students to progress to their next stage of learning or training despite the disruption caused by the pandemic.

Separate guidance was issued by Ofqual and JCQ for GQs while awarding organisations (AOs) issued their own guidance for their VTQs. The guidance issued was designed to allow flexibility for centres, recognising the tight timescales and different circumstances they all faced. A more tightly constrained approach might have been difficult for some centres to follow.

All TAGs needed to be based on a range of evidence completed as part of the course which demonstrated the student's performance on the subject content they had been taught. A significant difference was that in some modular VTQs the TAGs were determined for individual units within a qualification. This included some unit-level TAGs for first year students on some longer courses (2 year or more) VTQs. For other VTQs, and all GQs, a single qualification-level TAG was required. In all cases TAGs were to be determined using the same grading scale that each qualification would normally use.

The evidence used to support TAGs could be of a variety of types, including coursework, non-examined assessment, class work and classroom tests, mocks, and tests created and administered under exam-like conditions specifically to support TAG judgements. While schools and colleges had a certain amount of evidence available from before the announcement that summer assessments would be cancelled, for most centres collection of further evidence was a significant task to be completed, particularly once they re-opened after Easter.

Centres also knew that the TAGs they determined could be externally quality assured by the exam boards (for GQs) and AOs (for VTQs) through the review of selected student evidence. This was to confirm that the TAGs represented

reasonable academic judgement. As part of the TAG submission process for GQs in June, a sample of student evidence was uploaded from each centre. For VTQs the quality assurance process either followed the same model as for GQs or AOs created bespoke processes, sometimes adapting their normal moderation or verification processes to review TAG evidence.

As regulator of qualifications, including those that were awarded using TAGs, Ofqual needs to understand how the arrangements worked in terms of both the processes used, and the views of those involved, to learn for the future. While normal assessments are taking place this summer with some adaptations, TAGs formed a part of the contingency arrangements that would have been used if examinations had not been viable in 2022.

It is worth considering the arrangements used in 2020 when evaluating the arrangements for TAGs in summer 2021. The process for determining TAGs was different to that used for determining centre assessed grades (CAGs) in summer 2020. TAGs represent the grades at which students had demonstrated evidence of achievement, based on assessments that they had completed on content they had been taught. There was no element of prediction as to how a student would have performed if final assessments had gone ahead, as was the case for CAGs. TAGs were therefore determined through teachers' evidence-based judgements of completed work and assessments rather than prediction.

The impact of the pandemic on learning in 2021 was different to that in 2020. In 2021, the disruption to teaching and learning meant that, in many instances, it had not been possible for teachers to deliver the whole curriculum, whereas in 2020 almost all content had been taught and the disruption in the form of school closures arose at a point shortly before summer assessments would start to take place. The extent to which content had been delivered in 2021 varied widely across centres, qualifications, and different parts of the country. Therefore, the judgement as to the grade the student was performing at was restricted to tasks covering content that had been taught. This was to account, as far as possible, for the different levels of missed learning that had been experienced by students.

Following on from our [studies of how teaching staff had made their judgements](#) of CAGs in 2020, we carried out similar work to understand how the TAG process in 2021 had been managed, and how decisions had been made. This year we also included feedback from students, since they had generally completed additional assessments knowing that they would support their grades. For the CAGs in 2020 students were not actively involved, since those judgements were based on work and assessments they had completed prior to the announcement that normal assessment arrangements were cancelled.

To strengthen our understanding of how the 2021 assessment arrangements were perceived, and to inform the [development of contingency arrangements](#) should normal assessments have been cancelled in 2022, we carried out an online survey and follow-up interviews with both teaching staff and students. This report details the

teaching staff and student interviews. A separate report [describes the surveys we carried out](#).

Finally, it is important to remember that these interviews were carried out before the qualification results based on TAGs were given to students. Students should not have been aware of their TAGs. Therefore, the views of students particularly on final TAGs may involve some assumptions. Teaching staff were almost all aware of whether their TAGs had been queried through the external quality assurance process of the awarding organisations or had been accepted. Therefore, teacher views on final TAGs are much more definitive.

Previous research

We conducted a rapid review of research on teacher judgement that has been carried out worldwide (but published in the English language) since the literature review we undertook for the CAG report in 2020 (Holmes et al, 2021). Teacher judgement has become a topic of interest in some countries due to the disruption caused by COVID-19 and the need to develop new ways of awarding qualifications. However, many studies and reviews are not directly relevant to the situation in England. Our literature search revealed just a few recent studies of relevance to the TAG process. We do not repeat the findings from Ofqual analyses of the 2020 CAG outcomes. Interested readers are referred to Holmes et al (2021), Noden et al (2021) and Stratton, Zanini and Noden (2021).

We saw some parallels to TAGs in a study by Jönsson, Balan, and Hartell (2021). They ran a study comparing what they described as holistic and analytic approaches to grading in Sweden. In the holistic approach, a large set of student work was compared as a full body of work to a set of grading criteria to decide each student grade. This was a similar method to that used at the end of Swedish lower secondary education to decide grades. The between-teacher consistency of this approach was compared to that of an analytic approach. Here, each piece of student work, focussed on different skills, was assessed individually through comparison to the grading criteria, to build up a set of grades that could then be combined through a more numerical approach. This study found slightly higher agreement between teacher grades when they graded analytically, than holistically.

Other recent research highlighted a wide variety of factors, other than marks or grades for completed work, that can influence teacher judgements of final school grades. As part of their study of grading decisions of secondary teachers in Canada and China, Cheng, DeLuca, Braund, Yan, and Rasooli, (2020) picked out a range of these factors in analysing their focus group discussions. Contextual factors such as the students' effort, attitude and participation in class were all important for teacher judgements. Putting more weight on students' actual reasoning, thinking, and practical skills (a more holistic judgement) than on component grades was also mentioned as a common practice.

Other considerations that affected teacher judgements related to the perceived internal pressure from senior leadership teams and external pressure from parents and students, and awareness of the implications that final grades have for students' education and future life opportunities. The overriding concern of teachers when grading students was fairness, although this was understood and executed slightly differently in the two national contexts. However, teachers reported finding it difficult to grade fairly at all times, especially in the case of hardworking but low-performing students.

Similar factors affecting teacher judgements also emerge from Arrafii (2020) who found that in Indonesian secondary schools, in addition to student performance and cognitive abilities, teachers took account of factors such as in-class engagement, participation, and collaboration, as well as extracurricular achievements in externally organised, recognised competitions or programmes. The grades for completed work which students achieved during the year were weighted differently by different teachers. For some teachers, the core element for making judgements on overall student attainment were these partial grades. However, another commonly used approach was to pay less attention to the grades obtained by students from in-class assessments and home assignments, because of possible cheating. Instead, teachers supported their judgements by referring to the other factors mentioned above.

This recent literature, as well as the long-standing literature on teacher judgements, all show that these judgements are not carried out in a uniform manner and the guidelines supporting these judgements differ in their rigidity. One of the major challenges of teacher judgement that the reviewed studies point at (see Arrafii, 2020; Jönsson et al., 2021) is grade inflation. Generosity in grade judgements is believed to occur because teachers consider not only the cognitive and performance indicators, but also the range of contextual factors that are not directly related to student attainment. The difficulty of being entirely objective and impartial when teachers are fully aware of the consequences of their decisions for students also contributes to teacher-allocated grades being more generous than those resulting from regulated, external exams (see e.g., Cheng et al., 2020).

Similarly, issues of bias in teacher judgements regarding protected characteristics and socio-economic status have been reviewed recently by Lee and Walter (2020) and Lee and Newton (2021). We do not revisit their findings here, but note some other factors highlighted by recent studies that may bias judgements.

Student behaviours such as non-engagement and non-participation, poor interactions with classmates, excessive talking in class, and non-completion of required tasks were identified by Ferman and Fontes (2021) as being influential on teacher assessment in Brazilian middle and high schools. Controlling for the result of standardised tests and demographics, students who were perceived by teachers to exhibit these behaviours in class were graded lower than those that did not manifest difficult behaviours.

Carbonneau (2020) focussed on the effect that conflict between students and teachers, often manifested in difficult student behaviour, had on judgements of student ability on schools in the United States. Perhaps counter-intuitively, they showed that as teachers perceived the conflict between them and a student to escalate, the accuracy of their judgement of a student's ability relative to standardised test scores increased. This was explained by the generally inflated evaluations by teachers for students with whom levels of conflict were low. In a conflict situation the student evaluation would be relatively lower, bringing it closer to reality. However, the other students remained generously graded and so bias against the poorly-behaved students was present in the case of this study.

Teacher judgement can also be biased by the personality traits of students. As shown by Westphal, Lazarides, and Vock (2021), more conscientious students received better teacher-assigned grades in mathematics. With rising student conscientiousness, teacher-allocated grades also became more closely related to standardised tests scores. These results were somewhat echoed by Brandt, Becker, Tetzner, Brunner, and Kuhl (2021) in Germany, who showed that teachers' perceptions of student conscientiousness and openness were more strongly associated with teacher-assigned grades than standardised test results in both mathematics and native language. There is, of course, a relationship between personality traits and behaviour so these findings are consistent with those described above.

Other student characteristics that may affect judgements include less favourable teacher grading of students that are overweight and obese - particularly boys – in comparison to average or underweight students (Dian and Triventi, 2021). This was more pronounced at higher levels of attainment, with overweight and obese students, again in a German context, being less likely to achieve high or medium-high grades, regardless of having the same level of objective ability as non-overweight students. The same effect of weight and BMI was also found by Black and de New (2020), who also observed that taller students (at age 10-11 in Australia) were graded better by teachers than students of an average height.

The reviewed literature shows considerable evidence of teacher judgement being susceptible to biases around individual characteristics of students. As shown by Lee and Newton (2021), the findings are not generally so clear cut for many protected characteristics. Teachers do not make their judgements about student attainment in a social vacuum. While protected characteristics may be focussed on in bias training and analysis of data, it may also be difficult to minimise perceptions that lead to bias against individual students, particularly where the judgement process is more holistic and less based on marks or grades. The TAG process and the guidance associated with it, particularly that around [making objective judgements](#), was designed to minimise the kind of potential affects discussed here.

The current study

Having considered other recent research that has looked at teacher judgement generally, the current study explores how TAGs were judged in England in 2021. The interviews described in this report follow-up on the surveys we carried out, to further explore some of the experiences of teaching staff and students. The focus of the interviews was on two research questions:

1. In determining the TAGs, what sources of evidence were considered and how were they combined?
2. What steps were taken to ensure fairness (i.e. to minimise bias) in judgements and to what extent were these perceived to be effective?

While these questions were sometimes best answered by the teaching staff, students also gave insight from their perspective. Their views on fairness, and their experience of the whole process, were also useful in helping to design contingency arrangements for 2022.

Differences between TAGs, CAGs and normal assessment

Last year we considered the differences between CAGs and normal assessment arrangements. This year TAGs again meant a change in the role of teachers, although the shift in role from CAGs to TAGs was smaller than the shift from normal assessments to CAGs.

Change in the role of teaching staff

In both 2020 and 2021 teaching staff across the country had to move from their usual position of trying to maximise the performance of their students under normal assessment arrangements, to being the assessor. In other words, they had to switch from being a formative assessor to a summative assessor. This was particularly the case for GQ teaching staff.

There was perhaps less tension between the teacher and assessor roles in 2020, since there was a clear separation between actual teaching prior to March 2020, and the judgement of CAGs following the announcement that summer assessments were cancelled. In 2021, teaching and learning could continue alongside evidence collection for TAGs, meaning that teaching staff had to balance these two roles. This was potentially challenging, with the potential to affect the teacher-student relationship.

A significant difference between 2020 and 2021 was that rather than making a prediction of likely future performance in (cancelled) assessments, as required for

CAGs in 2020, the TAG judgements in 2021 contained no element of forward prediction. They were an evaluation of current performance based entirely on evidence produced by the student.

Therefore, there was a degree of uncertainty inherent to CAGs that was not present for TAGs. For CAGs, teachers were required to judge whether students would have continued on a steady performance trajectory from the level they were working at when schools were closed, as evidence by marks and grades on completed assessments, or whether this trajectory would have changed. For TAGs, performance in completed assessments or work had to be judged in comparison to the performance level expected for each grade. Existing evidence could be used, but because this may not have been designed and implemented specifically to support this task, there was an option to collect new evidence specifically designed to support TAGs. Many centres chose to do this.

Experience

The need for centres to collect robust evidence of the performance of their students often required them to be actively involved in designing and running assessments. This task was carried out by a wide variety of staff within departments, with varied teaching experience, but also with varied knowledge of formal assessment practice.

There is therefore a contrast with external assessments, where awarding organisations deliver formal training to all of their examiners and assessors including yearly standardisation. This involves forensic reading of candidate responses and discussion of why features or characteristics of the answers capture certain mark-worthy qualities. Where departments had teachers who were examiners, their experience would have been useful in supporting the TAG process, but this would not have been the case in all centres. VTQ staff would have been more experienced in designing and running assessments due to the internal assessment common in these qualifications.

Anonymity

TAGs were determined by individuals who knew their students well. To support them in making fair judgements, Ofqual released guidance about making objective judgements, which included raising awareness of unconscious bias. Some centres did introduce an element of anonymisation to certain stages of the TAG process to help with this. Anonymisation was more difficult for CAGs, since the predictive element throughout the process generally required some knowledge of the student. By contrast, external examinations are marked anonymously, with no knowledge of the student producing the response. Indeed, exam boards and awarding organisations have safeguards in place to prevent examiners from encountering examination scripts from known individuals.

Methods

Overview

Through a series of interviews, this report explores the intricacies of the grading process and the context in which judgements were made, in addition to themes including fairness and reflections for the future. The details provided in the sections that follow complement the findings from the quantitative survey by offering detailed insight into the whole TAG process in individual centres, including the variation seen between centres.

Interviews

A total of 53 interviews were conducted with 39 teaching staff and 14 students between 13 July 2021 and 5 August 2021. All interviews were carried out via video conferencing software, such as Microsoft Teams or Zoom, according to the preference of the interviewee, and they were free to turn their cameras on or off. Although interviews were predominantly one-to-one, all students were provided with the option of inviting a parent/guardian to attend for support.

Each interview was led by one of seven Ofqual researchers, each of whom conducted at least 4 teacher and/or student interviews. Almost all student interviews and just under half of teaching staff interviews were attended by more than one researcher, with one conducting the interview and the other(s) acting (primarily) as an observer. This was for the purpose of (i) standardising researchers in their use of the interview schedule and (ii) making sure that interviews with students under 18 did not cause them any distress.

Two separate semi-structured interview schedules for teachers and students were designed to guide the discussions. The teacher interviews aimed to elicit information and views about the process by which the TAG judgements were made, whilst the student interviews focused on experiences during the process and perceptions of fairness. Different teacher roles had a different involvement in the TAG process and consequently some topics and questions were more relevant than others. Interviewers made slight adaptations to questions, depending on the specific role of the interviewee.

Although the interview schedule provided a structure and focus for the interviews, there was flexibility to engage in discussion specific to the interviewee's experiences. The broad nature of the interviews meant that we could not discuss every aspect of the TAG process consistently. The interviews lasted between approximately 40 minutes and 2 hours (the mean duration was approximately 1 hour). All interviews were audio recorded (with participant consent) and transcribed verbatim by an external transcription organisation.

Participants

Participants were recruited using Ofqual's surveys of teaching staff and students. At the end of both surveys, respondents were invited to take part in a follow-up interview. This offer was removed from the teaching staff survey around a week before it closed due to the number of positive responses received relative to the number of interviews we could carry out. A total of 519 teaching staff and 104 students agreed to be contacted, of which 119 teaching staff and all students were emailed and invited to a follow-up interview. Response rates to invitations to interview were lower for students compared to teaching staff, which could be explained by the timing of interviews (i.e., invitations sent to participants during the summer holidays), how confident students felt in sharing their views and how engaged students were in the process.

Prior to interviews taking place participants were provided with information sheets, informed consent forms and had the opportunity to ask questions. They were assured of complete confidentiality and told that any information gathered during the interviews would be used for the purposes of the research only. Participants were able to withdraw from the interview at any time.

Teaching staff

We interviewed 39 teaching staff in total. Those invited to take part reflected a range of educational contexts based on role, centre type, and subject and qualification specialty. Often, the participants were involved with making judgements for more than one type of qualification and subject. See the following tables for the number of interviewees across role and centre type (Table 1), GQ subject (Table 2) and qualification type (Table 3). 'Senior oversight' in Table 3 indicates senior managers (head or deputy head of centre or other senior leadership team members). In Tables 2 and 3 counts for subject and qualification do not sum to the number of teaching staff since multiple qualifications were sometimes taught by each member of staff.

Table 1. Teaching staff interviewees by role and centre type

Role	Sixth form college	Academy	Comprehensive	FE establishment	Independent	Selective	Training Provider	UTC	Tertiary College	Other	Total
Head of centre			1	1							2
Deputy head of centre	1	1			1	1					4
Head of department	1	2	2		4		1	1			11
Deputy head of department					1				1		2
Senior leadership team member		2			1						3
Teacher and/or tutor	3	2	3	2	2	3			1	1	17
Total	5	7	5	3	9	4	1	1	2	1	39

Table 2. Teaching staff interviewees by GQ subjects taught

Subject	Count
Ancient History	1
Art design	1
Biology	2
Business	1
Chemistry	2
Citizenships Studies	1
Classical Civilisation	2
Combined Sciences	3
Economics	3
English language	4
English literature	3
Food and Nutrition	1
French	1
Further Maths	2
Geography	3
German	1
History	4
Latin	1
Maths	4
Music	1
Physics	3
Religious Studies	1
Sociology	2

Table 3. Teaching staff interviewees by qualification types taught

Qualification type	Total
AS/A level	19
BTEC	4
EPQ	4
Functional skills	2
GCSE	24
Senior oversight	9
Technical and applied generals (exc BTEC)	1

Students

We interviewed 14 students in total. Those invited to take part reflected a range of centre type and subjects, however most students were enrolled on to A level or GCSE qualifications. See the following tables for the number of interviewees across centre type (Table 4), subject (Table 5) and qualification type (Table 6). As for teaching staff, counts for the subjects in Tables 5 do not sum to the number of students since multiple qualifications were taken by each student.

Table 4. Student interviewees by centre type

Centre Type	Total
College	4
Secondary	4
Independent	4
Selective	2
Total	14

Table 5. Student interviewees by GQ subjects taken

Subject	Total
Art design	3
Biology	7
Chemistry	7
Combined Sciences	2
Computer Science	1
Drama	3
Economics	2
English language	6
English literature	7
Food and Nutrition	1
French	5
Further Maths	3
Geography	3
German	1
History	5
Maths	8
Music	1
Physics	5
Physical Education	1
Politics	1
Religious Studies	7
Sociology	1
Spanish	1

Table 6. Student interviewees by qualification type taken

Qualification type	Count
GCSE	6
AS and/or A level	7
BTEC	1

Analysis approach

A thematic approach was adopted to analyse the interview transcripts. The teaching staff and student interviews were analysed separately by seven Ofqual researchers using the qualitative analysis software package Nvivo. All analysis was conducted in accordance with guidance published by Braun and Clarke (2006) and Nowell, Norris, White and Moules (2017) in three distinct stages:

1. An initial coding scheme was developed separately for the teaching staff and student interviews by the research team to reflect the broad topics found within the interviews. The utility and application of these frameworks were initially tested by all researchers on the same transcripts.
2. Once the coding scheme had been trialled, there was detailed discussion among researchers to further refine codes and assess agreement to ensure consistency during the analysis process. Conventions for coding interviews were agreed on, such as coding passages of text rather than single sentences (to ensure the context was captured) and referring to both audio and visual recordings alongside transcripts where there were ambiguities in the transcript (allowing the researcher to account for body language/ tone of voice when interpreting what was said).
3. As a final stage, all interviews were coded, in no particular order, using the finalised coding schemes.

Structure of the interview analysis

Although the teaching staff and student transcripts were analysed separately, the findings from the analysis have been merged where appropriate to provide a rich picture of the TAGs process from two different perspectives.

It is important to emphasise the exploratory nature of this research, as it seeks to deepen our understanding of the quantitative findings gathered from the teaching staff and student surveys. It is not possible for research of this kind to represent the views of all individuals awarding and receiving TAGs in 2021 due to the size of the sample and the level of detail accrued in each interview. Consequently, this report does not numerically quantify the number of participants endorsing particular experiences or points of view, as this could be misleading. Instead, we report

whether the views were commonly expressed or whether they were specific to one or two individuals.

The key themes that emerge throughout this qualitative study are supported by quotes extracted from interview transcripts. These are embedded within the report to incorporate the participants' voices and experiences, which are central to research of this nature. To maintain participant anonymity the names of centres and teaching staff/students have been replaced with teaching roles and centre type labels, giving some context for the quotes included.

The analysis that follows is divided across the following main sections:

The TAG process.

Describes the design and implementation of the process used to determine TAGs. This covers everything from the initial announcement of the cancellation of summer assessments, through to the internal quality assurance of grades ready to be submitted to awarding organisations.

Other considerations in the TAG process.

Describes issues of determining TAGs for specific types of students and the considerations given to ensuring that TAGs were as fair for all as they could be. Teacher and student stress are also discussed.

Overall confidence in the TAG process.

Describes views of how valid the TAGs were, considering various aspects of the process, together with perceptions of the reliability of the TAGs.

Beyond the TAG process.

Describes views on the readiness of the 2020-2021 cohort to progress to their next year and reflections on expertise that staff gained through carrying out the TAG process.

A note on terminology

We use the word 'centre' throughout to indicate all the different types of institutions involved in producing teacher assessment grades, be they schools, colleges, training providers or other types of educational establishment. Job role descriptions have also been simplified (as in Table 1) and do not necessarily reflect the variety of titles used by different centres for similar roles.

We also use 'SLT' throughout the report (other than in interviewee quotes) to refer to the senior leadership team of the centre. Teaching staff and students refer to senior management or senior leaders within their centre in different ways.

The TAG process

This chapter sets out the processes by which TAGs were determined. The discussions highlighted several similarities and differences in approach, both within and between centres. These were related to three main stages of the process: the design of the process, the evidence used by teachers to make judgements, and how the TAGs were quality assured within centres. This section focuses mainly on the process for GQs, as the majority of our interviewees taught these qualifications or worked on the TAGs for them. While some of these considerations were also somewhat relevant to VTQs, additional details specific to VTQs are explored towards the end of this section. Finally, we consider the contrasts teaching staff made between the different approaches used by different exam boards.

Opinions on changes to the assessment arrangements in January

A number of teaching staff we spoke to reported that they felt a sense of relief following the announcement in January that exams could not go ahead. Many expected that summer exams would not go ahead due to the disruptions caused by the COVID-19 pandemic and therefore initially felt positive about the decision.

I think our reaction initially was we kind of felt like it was coming in some ways in that internally we'd been saying they can't sit exams like normal for quite some time. So it felt like we're finally, at least we've got the answer. Even some of the students were starting to say are they really going to happen, are you sure? So initially it was a little bit of a relief to say now we know that that's not happening. And then who knows what's going to happen instead. **teacher and/or tutor, academy**

However, for some, the announcement came as a surprise and resulted in feelings of disappointment. In these cases, the teaching staff felt that normal assessments could have gone ahead in a socially distanced format.

I was really disappointed actually. I'd spent all of the autumn term telling my kids that the government would move hell and high water to make sure that the GCSEs went on . [...] We could have used village halls. We could have. Just the same as we've used for COVID vaccinations. We could have used all sorts of places to host these exams and still had kids socially distanced and safe. And I didn't feel that the kids having exams in exam halls would have been detrimental to them. **teacher and/or tutor, comprehensive**

Many of those we interviewed had started collecting evidence after the announcement of the cancellation of the exams. However, there was a general

feeling of frustration with the government because they felt the decision to cancel exams had been made late in the academic year and that a contingency plan should have been put in place sooner, and this meant the whole planning stage became quite rushed.

And I think that was the overriding feeling of frustration that the chaos of the last few months could have been avoided [...]. So contingency, schools would always rather have two plans one of which we don't use rather than this kind of constant reacting. **senior leadership team member, academy**

Design of the process used by centres to award TAGs

This section looks at the design of the teacher assessment process within centres. This includes the use of guidance and information from government and AOs to support the centres in designing their process, the activities that took place to support this and considerations that centres made.

Guidance from government and awarding organisations

Following the announcement that students would receive grades by teacher assessment, guidance and advice about how to carry out the task was provided by Ofqual, DfE, the AOs and the JCQ. These documents were used as the primary information for centres to consider when designing the arrangements for teacher assessments.

Most of the teaching staff we spoke to expected the guidance indicating how to undertake teacher assessments to be sent out soon after the announcement was made that exams could not go ahead. The delay in receiving the guidance, whilst consultation on the approach was undertaken, was felt to have had a negative impact on both teaching staff and students.

I think mostly it was just probably trying to work out what could we do in terms of assessment, and I think in the end we came to the conclusion we just had to wait for the guidance. There's no point in trying to plan anything until we saw the guidance, and knew what we had to do. [...] We said look, until we know we're just going to assume that whatever we do in the summer will be similar to real exams, and we just kept pressing ahead. **Head of department, comprehensive**

The rest of this section details views on the information and guidance issued by Ofqual and then by AOs.

Ofqual guidance for GQs

The GQ guidance from Ofqual was issued to ensure that grading decisions would be based on robust evidence, while allowing centres reasonable flexibility in how they designed the process to make it manageable for them. There were mixed views around the usefulness of this guidance. While some centres felt it was useful and clear, others felt that it was too generic - although they appreciated that it covered a wide range of subjects, it was not considered detailed enough to be helpful to design the process.

It tended to be very vague, it was like, 'oh, this is the sort of vague process that you need to use', which I can sort of understand because it has to cover literally every subject in every school, but on an individual basis it's just not helpful, it doesn't tell us what we need to do. We basically have to say, 'OK, we have to make up our own plan and do what we're going to do'. **Teacher and/or tutor, selective**

All those we interviewed felt frustrated by the timing of Ofqual's guidance. They explained the guidance was published at a time that made it difficult for centres to promptly react and respond to it, particularly where the Easter break followed immediately after the guidance publication.

There was, I'm going to be honest, a disappointment [...] The guidance as to what was going to happen came out in the afternoon of the Friday we finished for Easter, which was not great timing. **Head of centre, further education establishment**

For some, follow-up communications from Ofqual that were intended to be helpful in clarifying the initial guidance, resulted in their understanding changing, and so initial plans had to be adapted in response. This caused stress within some centres and in some instances delayed the centres' agreeing their final process.

It felt like constant evolution or revolution actually, it was like as soon as you thought you knew what was taking place then everything changed. And the school was saying it was because there was a blog from Ofqual that was daily making changes and it wasn't their fault that they were having to change the centre policy regularly because it was happening in line with national changes. So, it didn't feel very pleasant, that's for sure, a very stressful time. **Teacher and/or tutor, selective**

Awarding organisation guidance

Following the outcome of the consultation and the guidance issued by Ofqual on 24 March 2021, the Joint Council for Qualifications (JCQ) for GQs and VTQ awarding organisations produced and shared their guidance with centres. Within the area of general qualifications, the view was that the guidance was made available too late,

not at an optimal time just before the Easter holiday and it was too long. One head of department explained that they had condensed the information provided by Ofqual and the AOs before sharing it with the department.

Did we have enough information? It was things like some of the information was too large, like I was told that I had to train my department using the materials produced by the exam board and Ofqual and the documentation was too large, that I actually, the documentation that I sent to the department that I said to them they had to read was 800 pages. Now, most of that was exemplar from the exam boards of student responses and things like that, but if you're sending teachers 800 pages worth of documentation, they're not going to read any of it, so I distilled that stuff down and things like that. **Head of department, independent**

Some teaching staff felt that it was unreasonable that schools were requested to design a process and submit their grades in a short window of time, whilst at the same time, the guidance from AOs as how to do this was not issued promptly.

It doesn't seem reasonable to me that, from 5th January, where you know that exams have been cancelled [...] to I think it was 26th March, that the results of the consultation came out. [...] We then had to work through the Easter holidays, we still didn't have the stuff from the exam boards until 19th April quite often. That lead time from the national bodies who are providing the guidance is just too long, when then schools are supposed to do everything in a much shorter period of time. It's not reasonable. **Deputy head of centre, independent**

There were mixed views around the materials provided on the standardisation and the moderation of the TAGs. One person, for example, said that the guide from JCQ was helpful because it had examples that could be used to guide the process.

The worked examples there from JCQ really did help, because there they were showing that it is not an average of grades, it's where they're working now. So, if you've got evidence to say they've improved recently then that's OK, so, as an example, if you've got D, D, D, C, C, C, C, B, B, they're now at a B, because you get that, you get that beautiful blossoming of knowledge and understanding and everything comes together, so that was really helpful. **Deputy head of centre, sixth form college**

On the other hand, others felt that more guidance around the mark schemes and the standardisation process was needed to ensure that centres across the country were following a similar process. We consider issues of consistency throughout the section entitled ['Perceived consistency of the TAG process'](#).

Pre-TAG activities and considerations

Before centres undertook their teacher assessments, a number of pre-assessment matters were addressed. This largely related to designing the TAG process, writing the Centre Policy document, and issuing guidance and training to teaching staff.

Who was involved in designing the teacher assessments?

It was clear that SLT took a leading role in the design of teacher assessments in almost all our interviews. However, generally, individual departments and teaching staff were consulted early in the process. SLT tended to collate the information shared by Ofqual, JCQ, the exam boards and the AOs. They then designed a high-level approach to teacher assessments (with input from departments) as a starting point for departments to then devise their own assessment plans in more detail.

We had a 'heads of department' meeting in the spring term in which I outlined a timeline for what we would be doing when. [...] I explained where the practice exams were going to sit. I set up dates that we would be running the assessments in the summer term. [...] Departments had quite a bit of flexibility [...] We had set out the principle that we would run this final assessment period, then departments needed to find a way of assessing enough of the course to be able to give a grade on those final May assessments. **Deputy head of centre, independent**

Some teaching staff discussed how their centre collaborated with or discussed the process with other centres, either formally, or just through personal contacts. Where this was the case, interviewees felt this was a useful means by which they could refine ideas and check whether they were taking a suitable approach.

The first sort of layer I suppose was from a leadership team, working together, lots of reading, lots of talking to other senior leaders in other schools and sort of looking outward. A huge amount of time researching, reading and understanding, talking to people. [...] We work with an excellence group of schools, very similar context to us. [...] There was lots of, sort of, Key Stage 3 schools that converted to 11-16 the same sort of time as us. So, we've got quite a tight network with them, all be it not part of the same trust. **Deputy head of centre, academy**

A minority of the teaching staff we interviewed told us that they were solely responsible for their own assessment plans. In these cases, SLT did not specify an approach to be taken by each department, instead they were given the freedom to decide how to assess the students. SLT did not provide any specific guidance beyond the centre policy document and the guidance provided by Ofqual or the AOs.

Although teachers felt confident overall about their own judgments, one reflected on the issues around centres following a different process nationally.

We were given no guidance. We were told do what you think is best. And so because that came from the very top, SLT, our senior leadership team, said the same. They said look, this is going to look different for every single faculty. Your assessments are different, your subjects are different. We trust you to make the right decisions basically, which was great and I was very confident in what we were doing but I do know that that means nationally there was no consistency with how any of this had been done. **Head of department, academy**

Centre Policy document

Following the publication of the final guidance from Ofqual, JCQ and the awarding organisations, centres finalised their approach to determining grades. Though it was not necessary for all VTQs, Centres were usually required to create a centre policy that “reflected the centre’s approach to assessment and quality assuring the centre determined grades they awarded to students, based on the evidence they have produced” ([JCQ Guidance on the determination of grades for A/AS Levels and GCSEs for Summer 2021](#)). Centres submitted this document to the exam boards or AOs for review.

In most interviews we conducted it was clear that SLT had taken a lead role in writing the centre policy document and disseminating the document to staff, students and parents. The role of the centre policy document was different across centres: some of those we interviewed felt that it led the process, whilst others felt it was more to document the entire process after it had been designed. There was a minority that felt that it was an administrative exercise and did not find it very helpful.

Those who felt positively about the centre policy document explained that it was helpful to have all the information shared by the government, Ofqual and the awarding organisations summarised in one place.

[The centre policy] was the main document that the senior leadership team synthesised everything from Ofqual and the government into this one place so that we could refer to it and it was very clear exactly what we could and couldn’t do and what the procedures were for things. So that appeared very soon after the last document was produced by Ofqual and then that was something that was referred to frequently during the training. **Head of department, independent**

As one interviewee explained, the centre policy document was also used as a check at the end of the process to ensure that the centre carried out and submitted everything they were expected to.

So, that policy document was instrumental in helping us to provide all the evidence to make sure we were doing everything that we were supposed to be submitting, that we hadn’t left anything out of the process and also to do with

additional arrangements, accessibility, etc. **Teacher and/or tutor, further education establishment**

One deputy head of centre found it particularly helpful that JCQ provided a template to support them to write the centre policy. This also allowed them to check and compare their process to other centres.

It focused us right at that start because it was, we were going down a tunnel that we didn't quite know where it was going to turn and what the outcome was going to be. [...] The template was absolutely essential. I couldn't have written it without the template. But it enabled us to compare in quite a concise way with other centres. So, we shared our draft policy with other centres and looked at whether, what we'd done, and we added extra bits in on the basis of the feedback we got from other people. **Deputy head of centre, academy**

On the other hand, some teaching staff explained how it was a time-consuming administrative task and was not particularly helpful. This was largely because the document was too vague as it had to consider the processes used by different departments and qualifications.

The centre policy document then took all the processes from the different departments and amalgamated them together into one very woolly document, which sort of explained everything but none of it in any detail. So, as long as departments were [...] sort of fitting in with that, they could kind of do what they wanted or what they were going to do anyway. I think they wrote the policy document to fit what we'd already decided to do. **Teacher and/or tutor, selective**

Some felt that the centre policy was needed to capture the process but did not feel that it was a document for internal use. They felt that the document served to assure parents and students that a rigorous process had been used, but also aided their thinking about later stages of the process, such as potential appeals.

So, the process of having to write one I think was sensible. I never referred to it again, I never looked at it again once it had been written. It felt like it was a document for external agencies, it was for the parents, it was for us to submit. It was not a document for me to use, it was a document for my boss to have ready for there to be an appeals procedure in there, for that to have been thought about in advance, how are we going to handle that, what information are we going to give out at what stage. It was just, I suppose in a way it's like it forced us into having these thoughts and deciding on them and writing them down. **Head of department, independent**

Some interviewees perceived that the centre policy document was approved by exam boards or AOs too late in the process to be helpful. This was thought to be problematic as centres continued with teacher assessments without knowing if their

process had been approved. In these cases, they felt unable to communicate with students and parents about what their process would be.

Weeks later, and I mean almost at the end of the process, we phoned up again and said, look, we've heard nothing back about our centre policy, is it approved? And then eventually they said 'yes, it is, it's absolutely fine'. So, throughout the whole of the actual process it was not on our website, because we were not allowed to put it onto the website or share it with anybody until it had been approved. And that for me was slightly farcical, because how is that helpful to a student, or their parents or carers, or the staff indeed, if it's not actually been approved, so we were working all the way through that process with this centre policy, but somebody might have said, 'well, that's not right', [...] so that lateness has been an issue all the way through. **Deputy head of centre, sixth form college**

Internal training and discussion

Teaching staff were often provided with documents or summaries to guide their assessments. As we saw above, SLT often condensed the guidance from Ofqual and AOs into a more concise format, and these were provided as internal training documents.

One of the interviewees in a VTQ context explained that the VTQ team in the centre offered help, training and support to the colleagues assessing GQs. They felt this was helpful because the vocational staff were well-practiced in assessing and moderating their students' work.

In terms of coming up with training and everything, because we've run these vocational courses for quite a while, all staff members are already trained on assessing and internal moderation and external moderation, so the staff delivering on the vocationalals, we felt were in a lot stronger position because this is what we do normally. We do the coursework, we have to assess it, we have to internally moderate it. [...] And so we ended up being trainers to staff who just delivered GCSE or A-levels because this was just a complete new situation that they were having to face. Whereas we were used to it. So, we did become trainers and aiding all these different, other subjects and other staff members, which did help them. **Senior leadership team member, academy**

A few of the teaching staff reported a relative absence of formal training on how to assess evidence. Instead, they were simply referred to documents to read.

There wasn't a lot of training. There was the initial, I think, heads of department meeting where they went through what evidence was going to be in the basket, and some training on how to come to a final grade when you've got your list of different grades, a little bit on bias. So, it tended to be more: you were referred to

these documents, which you need to read, rather than really active training. It was more emailing round the documents. **Head of department, academy**

Although active training in how to apply a teacher assessment approach was lacking early on, much of this occurred as assessments were completed, and marking and grading started. This is described within the section entitled '[Marking of individual pieces of evidence](#)'. Nearly all those we interviewed also explained that initial training did include consideration of fairness and bias. This is further discussed in the section entitled '[Fairness and minimising bias](#)'.

Evidence used to make judgements

Following the generally high-level plan from SLT on how the TAG process should proceed, each department had to i) select the evidence to be used for each subject, ii) devise the assessment plan, iii) decide the number of and the content of the assessments, iv) mark the assessments, and v) assign grades to each student. This section focuses on the first three steps of the process. The latter two steps are covered in the section entitled '[Evaluation of evidence and Internal Quality](#)'.

This section specifically addresses the types of evidence that supported the TAGs. This includes details of evidence completed before the final guidance was published (particularly mock exams) and post-guidance evidence. Within our sample of interviewees, most centres based their decisions heavily on post-guidance assessments taken under exam conditions, with mocks contributing in some centres, often depending on the conditions under which they were taken. We look at these aspects firstly from the perspective of the teaching staff, and then consider the students' experience.

Pre-guidance evidence

Most of the teaching staff we interviewed felt that final, end of course, assessments were the most reliable way to assess the students, and in some centres these were the only pieces of evidence used. However other evidence completed before the publication of the guidance, such as mock exams, non-exam assessment (NEA), class work and class tests did often contribute to the TAG judgements.

It is worth noting that differences between subjects are likely to have arisen, however, we did not have a sufficient sample to conclude any subject-level differences, and neither do we have space in this report to reflect all the varied considerations specific to particular subjects. As such, this section highlights cross-cutting considerations.

We first take a closer look at the role of mock exams and other pre-guidance evidence in the TAGs.

Mock exams

There was variation across centres regarding the timing of their mock exams. Some teaching staff reported completing the mock exams before the Christmas break, applying their normal invigilation and exam-conditions.

I did one in November and it was a full mock paper on the very first unit that we cover in the first year, which is Greek art. They knew it was coming, it was in proper mock, or as best we could, mock conditions, with all the usual extra time, use of computers and stuff. But it was a proper mock paper. And it was, and that's an hour and 45 minutes. So that was the closest that they were going to get to a full actual paper. **Teacher and/or tutor, sixth form college**

Others who were planning on carrying out their mock exams in January felt unable to do so because of the national lockdown and did not go ahead with them, whereas some continued with mock exams as planned, but undertook them remotely. In these instances, invigilators used online video conferencing software to watch over students as they took their assessments. These remotely-completed mocks often didn't carry much weight in final decisions when centres had concerns that their remote invigilation processes could not guarantee fully-controlled exam conditions.

Other pre-guidance evidence

In addition to mock exams, most (but not all) centres collected a wide range of other work that students had completed prior to the announcement of exam-cancellation and the release of guidance. This included NEA routinely undertaken by students on particular courses, class work and class tests.

For GQ subjects with NEA content, the teaching staff we spoke to commented that this was a useful source of evidence. One teacher said that the department was able to use the NEA as they had a record of the parts completed at home and those completed in class.

I was very acutely aware of subjects that had NEAs still in progress, once we got the guidance obviously, we knew that they didn't have to be complete, that we could use them, they were a really good source of evidence. Didn't matter whether bits had been done at home and we got the curriculum leaders to look at those NEAs and risk-rate them really in terms of how much was done at school, how much was done at home, you know, how reliable was what they'd got from their NEAs. **Deputy head of centre, academy**

A minority of centres did not include any pre-guidance evidence. In these cases, they reflected on how formative work provides opportunities for feedback, and that this was an important part of learning. Ultimately, these teachers felt it would have been unfair to use evidence from when students had not completed the course, as this would be a poor reflection of their ability and potential.

We see the learning process as one where you should be free to make mistakes all the way through. That you learn from your feedback, that feedback is therefore honest and open, and we don't focus on grades as we go through the process. So we didn't have a bank of evidence already accumulating towards the final grade. **Deputy head of centre, independent**

For VTQs, coursework represented a trusted source of evidence as much was completed (and often marked and internally moderated) before January. The Extended Project Qualification (EPQ) fell into the same category, being project based and largely completed (and in some instances also marked) by the time the announcement of the cancellation of the exams was made. One teacher reflected on this.

Some candidates had already handed in and completed their EPQs, so had actually produced full reports, 10,000 words long in some cases, and were just gearing up to do their presentations, which of course we then did early in January with Microsoft Teams. [...] What I've got in front of me is marked, second marked and what will be soon to be internally standardised work and all the evidence is present. [...] I think we'll do what we traditionally planned because it is a real grade rather than some kind of arbitrary grade arrived at by different means, this is evidenced work. **Teacher and/or tutor, tertiary college**

Overall, centres varied in how much weight they gave to pre-guidance evidence relative to post-guidance assessment. The relative weighting of evidence was decided by individual centres, departments or staff. This is further discussed in the section on ['Weighting evidence when determining TAGs'](#).

Other more formative types of evidence such as class work, class tests, homework and home-based tasks or tests were typically considered more useful for checking that TAGs arrived at using more recent evidence were reasonably consistent with the student's longer-term performance.

Post-guidance evidence

The bulk of the evidence that contributed towards the TAGs came from post-guidance assessments that were undertaken after Easter. In this section we describe how centres designed and delivered these assessments.

The design of the post-guidance assessments

This section explores the types of assessments that took place post-guidance, who was responsible for creating them, the use of the assessment materials provided by the AOs, the format and number of assessments and how content coverage was managed. Finally, we consider some subject-level differences in the design of the post-guidance assessments.

We identified two main approaches to these assessments. Some centres ran a series of in-class mini assessments, which usually required less than 1 hour to complete and were on specific topics. Other centres ran fewer but longer assessments, which each covered a range of topics.

Who created the assessments?

The most common approach mentioned was for departments to design the assessments considering the centre policy document and the guidance from Ofqual and AOs. The assessments were usually designed within the department using a continuum of approaches. At one end of this continuum, individual teachers took responsibility for designing the assessments, while at the other heads of departments took full responsibility. Usually though, there was a degree of collaboration across these roles.

There was a lot of work on all of us going away, thinking about [the assessments] as a department, coming back and then sharing those things and those questions that we had, and the dilemmas that we had. **Head of department, sixth form college**

In this centre, just a few people were involved in the design of the assessments, to ensure their integrity.

I have two Deputies, and it was a centre decision that we would minimise the number of staff that had prepared these exam papers for security. So basically, only the three of us worked on making these exam papers so that most teachers hadn't seen them, so that in that period when they were preparing their students for these internal exams they didn't know what was on them so that they weren't able to offer any guidance. **Head of department, independent**

Assessment material produced by the AOs

To design the assessment, teaching staff used a combination of past-papers, materials provided by the AOs, and materials produced by the department. All our teaching staff interviewees were expecting AOs to provide unseen materials that they could use and adapt for the post-guidance assessments. When the materials were released, however, they were disappointed that much (or as below, apparently all) of the content was taken from past papers, which they already had access to.

We were supposed to get some material, more material to be able to assess from, which I didn't think it was really good enough what was given because it was all just previous past paper stuff, it wasn't anything new. So in my opinion the stuff that we were given at the time wasn't really up to the standard that it should have been. **Teacher and/or tutor, sixth form college**

Teachers explained that the material was shared too late, and it was not as helpful as they expected. A teacher explained that the material published was grouped by

topic and so for them not usable as published, requiring them to construct their own assessments.

It was very topic based. So there was a huge volume of it, which was quite topic specific, so having taught the whole course we wanted to just give them an exam paper that covered everything. If you were trying to use the exam boards' that was topic based, to try and cover I think a broad enough spectrum of topics you would have had to have given them far too many tests. I think it would have been too difficult. **Teacher and/or tutor, independent**

Overall, the feeling was that the material provided by the exam boards to support assessment design was not helpful because so much may have already been seen by students.

With also the hope that by then the exam boards would have produced some more materials that we could use for exams, because we're ploughing through secure content of exam materials in getting them to sit exams and getting them to sit questions that they hadn't seen before and the more times we have to do it the less there is available. **Teacher and/or tutor, selective**

The teachers' disappointment with the resources provided by exam boards this year was also reflected on by some students. Students felt that the publishing of the previous assessment materials and mark schemes on exam board websites was unfortunate. There was a sense that as these materials were accessible to all students, their usefulness for teachers was limited. Moreover, students perceived that releasing such materials could give an unfair advantage to those students who located the papers online and were therefore able to practice the questions.

Yeah, the exam boards have given us good exams, but they put them online for everybody to use, including the students. So, a lot of people went on those exam board pages, got the test, searched up all the answers for them, and the next day took the test. [...] If they kept the exams off of the internet, and maybe released them after the exam fine, I do not mind. But the fact that they were the whole time for anybody to use—and I understand that they wanted to make it fair and everything, because I respect that—however, it wasn't very well thought through, because the whole exam was on there. **year 11 student, independent**

Format of the assessments

All those we interviewed looked at past papers in some way to design the post-guidance assessments. The assessments took a range of forms, including full and unedited/unchanged past papers, combinations of different past papers, past papers that teachers had made some changes to and papers that teachers had written from scratch. The decision was often driven by an intention to ensure that students were assessed on questions they were unlikely to have seen before.

We used the 2020 paper as the one that they hadn't seen already. So, we used topic questions from that. So, assessment week one was a skills-based paper, which we essentially wrote ourselves using questions from lots of different places.

Teacher and/or tutor, academy

We'd already used quite a few past paper questions we couldn't use a full past paper so we kind of used a selection of questions from different past papers to make up a full paper. **Teacher and/or tutor, independent**

Some centres had teaching staff who were experienced examiners, and they felt this experience enabled them to create good quality assessments of their own.

At the end of the day most of our teachers are very experienced professionals, a lot of them are examiners so we knew what good robust assessments needed to look like. **Senior leadership team member, independent**

For others though, they felt they did not have the experience, time or resources to undertake this task. One teacher felt empathy for those centres that did not have much assessment experience.

Pitching a listening task or a reading task is incredibly difficult. That's why the exams are so expensive because that's what they do, that's their job and that is not a job I am trained or qualified or feel in any way ready to do. And it is incredibly time consuming and it is time I did not have, so we relied on the textbook publishers and exam board experience to provide us with resources that we could use. **Head of department, independent**

I work with [two exam boards] as an examiner and as a senior examiner. [...] This was my nineteenth year of teaching and it was like, OK, right [I can do this]. I feel sorry for teachers who don't have that much experience and who don't work for the exam boards and who had less of a picture as to maybe what they could do.

Head of department, sixth form college

Creating assessments from scratch or as a combination of questions from past papers was often a challenge. In particular, the teaching staff found it difficult to determine how the marks should be mapped onto final grades.

I think that was the most difficult aspect, because we had designed our own exams, and [...] we weren't allowed to use old past papers, because it was felt that there could be cheating there, students could get hold of the material; [...] So we had to put all that together using various past papers and course books and assessment materials and come up with grade boundaries. **Head of department, academy**

Logistical limitations also influenced the format of the assessments, in particular the time available to sit the assessments. A few centres opted for assessments that were similar in length to formal, end of course, exams. In many centres though, this was not feasible as they did not have the time and facilities to run these. There was

also an awareness that students had been informed that exams were cancelled, and they felt it was unfair to students to expect them to sit full-length assessments.

It wouldn't have been right or fair to say 'hey, electronics is normally a two-and-a-half-hour paper or a three hour paper, we're going to set a two-and-a-half or three-hour paper but not call it an A-level'. The spirit of TAGs was about more data points, slightly shorter assessments. **Senior leadership team member, independent**

Overall, most of those we interviewed, including the head of a science department talking about GCSEs below, indicated that the assessments were shorter than typical exams.

We did nine or 10 short exams. So for the triple students they were short 45-minute exams, and for the combined students they were 30-minute exams, each one with the two hour preceding, us revising the topics and stuff like that with them first. **Head of department, comprehensive**

Where there were constraints on time and facilities to run full-length assessments, some of the teaching staff we spoke to explained that they adapted existing past paper to fit the time-window that was available, typically by modifying and/or removing questions.

We didn't have the time available for the exam. I think we only had like a two-hour slot and it was a two hour 15 exam, so we had a two hour slot in our exam timetable, so we had to cut down those exams a little bit. **Teacher and/or tutor, selective**

The number of pieces of evidence

There were a range of approaches to how many pieces of evidence were used to support the TAGs. In some centres, departments had the flexibility to choose the number of pieces of evidence used.

So I'd say typically it was three data points for every subject. [...] I think Greek or Latin had seven or nine, but they were tiny classes with much more class-based assessments [...] I think the TAGs process gave you a huge amount of flexibility, you could do whatever you wanted. But we were multi-point, slightly shorter, but exam-based assessment. **Senior leadership team member, independent**

In other cases though, the requirement was determined by SLT. Sometimes SLT were particularly prescriptive about what constituted a valid piece of evidence and how many pieces of evidence were required.

So basically again the senior leadership team suggested about five to six pieces per student and they strongly, well there were some things that were said, essentially it was the mock exam from the year before which had been done online, then it was a couple of block tests from the autumn term, then it was the

mock that they did in the spring term once we returned from lockdown and then in the summer term we did two lots of assessments and so we used those as well and that made up the six. **Head of department, independent**

Sometimes SLT allowed some flexibility for departments to choose the pieces of evidence that best suited the subject.

The college was talking about using a basket of evidence made up of four buckets. [...] They imposed bucket one as a blanket requirement. It was going to be the average of the five best assessments across the two years for every student in every subject, and that would be worth 20% of the final grade. And then that leaves three other buckets up for debate. If you had a subject that did NEA that would go into bucket four. **Teacher and/or tutor, sixth form college**

Content coverage

The guidance was clear that students should only be assessed on the content that they had been taught. Although there was no minimum requirement of taught content set by the AOs, centres were asked to confirm that sufficient content had been taught to award a valid TAG.

From the interviews, it emerged that there was some variation in the proportion of the specification covered across centres and subjects. The majority of teaching staff we interviewed commented that over 75% of the course was taught, with some having covered all the course content, across remote and face-to-face teaching.

The vast majority were in the middle, and what we went for was assessing them on about 75% of what they would have done in the real exams, which basically mirrored what we taught them. We'd covered about 75% of the course, so it all fitted quite nicely. **Head of department, comprehensive**

One teacher explained that, although they had largely covered the content on the specification, it was difficult to be sure of each student's actual level of engagement, particularly when teaching remotely. This is likely to have resulted in a mismatch between the content that teachers had taught, and the content that students believed they had learnt.

We largely covered it, but it's very difficult to see how much engagement you're getting from the other end and I think for the people who were going to work hard, it was quite good because effectively they almost get like a one-to-one tutor, for the people who weren't going to work very hard, it's very easy to not bother doing anything. **Teacher and/or tutor, selective**

Some centres assessed students only on content that had been taught in person, so as not to disadvantage the students who had engaged less successfully with online learning.

We were told we couldn't assess them on material that they hadn't done in class if it had just been during lockdown. [...] So each paper was out of 80 and I think we maybe knocked 20 to 25% off to say 'this is the material we weren't able to teach or we taught superficially'. [...] We just had a look at the exam papers and just said 'OK, we've taught [this], we haven't taught this, we've taught this superficially'. **Teacher and/or tutor, comprehensive**

In some instances, students also reported that their centres made decisions on inclusion of content based on the quality of remote learning.

For maths, I think we finished the curriculum quite early—like in January—so we'd been doing revision since then. [...] That carried on for a long time for us, doing a lot of online lessons, so it was very difficult for a good chunk of the course. But my college with the assessments, there's not that much stuff that we've not been taught at all. So, they've obviously taken that stuff off, but that still left the majority of the course on. So, my college made the decision the stuff we were taught—well, [that] we were sent out by email at the very beginning to teach ourselves—they've taken that stuff off as well. Because we've not really been taught that, we've just learned it. **year 13 student, college**

Centres ensured they covered the breadth of skills that would typically be assessed in a particular subject when designing the assessments, as this music teacher reflected.

So, we couldn't just go 'right, this person has done brilliantly in a question six on these three papers' [and give that grade]. But there's no question five in there. So, we would say we need to have a question five because it's covering different skills. And in fact, none of our assessments would have been purely on the [one skill]. They all were 'listening, plus an essay' or 'something, plus essay'. Because that's where you cover the AO4 [assessment objective 4] is in those essays. The listening tests on the whole, tend to cover AO3 [assessment objective 3]. **Head of department, independent**

Subject-level considerations

Whilst our sample was not sufficient to draw strong conclusions about systematic subject-level differences, there were some observations of differences in the design of assessments for specific subjects.

Some teachers felt there were differences between subjects in the need to use 'unseen' questions in assessments, with answers to previously seen questions more likely to be remembered in some subjects than in others.

Now, in subjects like maths and computer science, and probably the natural sciences, where if you give the students the questions and they go away and look at the answers, and can remember the answers, it didn't make a lot of sense to

give them a seen assessment. So, lots of those subjects went for two unseen assessments. **Teacher and/or tutor, sixth form college**

For these types of questions, previously seen questions could be used if they changed certain elements of the question, such as the specific data to be used in a calculation. They felt this allowed them to assess the student's ability, rather than their memorisation of a question from a seen past paper.

A lot of physics questions, at least, you can change the question quite easily just by changing the numbers, so we did quite a bit of that. It's quite hard to do, because you have to then pick numbers that are still going to work [...] you have to check that it doesn't affect something else. **Teacher and/or tutor, selective**

Other teaching staff also noted that where normal assessments contained optional questions, often in the more essay-based subjects, they retained this optionality in new assessments they created.

Giving a choice to students is not unusual in economics. Essentially on the first two papers all questions are optional. [...] So when the guidance was to give as much choice as possible, we decided on three pairs of questions on the seen assessment. [...] And then for the unseen assessment there were three multiple choice questions that were compulsory, and then a choice of three different contexts. **Teacher and/or tutor, sixth form college**

The teaching staff we spoke to often identified that practical subjects were more difficult to deliver during the pandemic, especially, for example, those that required any technical materials, laboratory-based or group work. A geography teacher reflected on the difficulties around transferring teaching and learning activities for such content online.

We, obviously for geography we lost the fieldwork relatively early on. We hadn't done any fieldwork with this cohort. But other than that we covered the content. I think it's probably relatively easy to deliver online when you're not trying to do science practicals or art or... as compared to other subjects. And a lot of our students I find they quite like relatively traditional teaching in terms of teacher led stuff anyway so that kind of switched online comparatively easily whereas some of the more interactive things obviously don't transfer quite as easily and that's where some people have lost more. **Teacher and/or tutor, academy**

Delivery and logistics of the post-guidance assessments

In this section we focus on how post-guidance assessments were undertaken at centres. There were different approaches seen in terms of spacing out assessments

or having a dedicated assessment period, and in the location and conditions under which assessments took place.

Ongoing teaching and revision time after Easter

First, we consider the balance of teaching and revision time after the Easter break, before most assessment began. Interviews with students were the most revealing of how time was spent after Easter. Many centres completed the delivery of new content online before Easter and then focused on revision, but other centres were still teaching new content, and approaches to revision time varied. It was noted in some teaching staff interviews that in normal years the revision period usually started in February and so students had lost some revision time.

It was more common that teaching finished and time after Easter was mostly spent revising. This could be on the basis that all content had already been covered, or just to teach as much content as possible before Easter and then to concentrate on revision.

After Easter when we went back, a lot of us had finished, I think we finished everything in lockdown - where they wanted us to be. [...] Everything was pretty much revision lessons, going through old content, practice questions, that sort of thing. Basically Q&As effectively, if we had anything we wanted to go through they'd plan a lesson on it, do revision lessons. And it was basically the same in every subject, it was no new content. **year 11 student, selective**

From after Easter I think, we were doing revision for all of our subjects. For some subjects [...] - like for maths I think - we finished the curriculum quite early, like in January, so we'd been doing revision since then. But for all subjects they stopped teaching. Even if we hadn't finished the course, the teaching stopped by Easter. And then we did revision after Easter, and we got sent revision lists for the exams that were in April and May. **year 13 student, college**

A couple of the students we spoke to detailed how at least in some of their subjects they were still learning new content after the Easter break.

We actually did have teaching of new material, which was hard when you're trying to learn things, and then they were saying 'well, here's some new stuff to learn as well'. Yeah, because obviously the exams were in timetable. [...] But it meant, in some lessons that we'd finished all the content we'd be doing revision in the lesson. [...] And then in some other subjects that we hadn't finished we'd be learning new things as well. **year 11 student, secondary**

This often meant that some topics were delivered at the last minute and intertwined with ongoing revision or TAG evidence collection.

In the weeks coming up to the four-week assessment period, I think there was maybe three or four weeks before that where maybe I think it might have been

the senior leadership team basically said to teachers right, stop now, just go over past material. And instead what lessons became was essentially revision periods before the exams. [...] But it was just learning new content up until about three or four weeks before the assessment period I should say. **year 13 student, selective**

In some instances, students suggested that specific topics that were to be tested were taught at the last minute.

There were a lot of topics which we still didn't know, which were for a fact going to be on that test. [...] I think teachers kind of panicked a little bit, and they were like: "Oh, we need to focus on the exams, because we don't have any evidence." And then they were like: "But they still don't know this," and so, the day before we were learning the stuff for the exams the next day. **year 11 student, independent**

There were mixed perceptions about the amount of time that was dedicated to teaching new material, revision, and time spent focussing on assessments for TAGs. As shown by the following quotes, some students were satisfied and found the approach adopted by their schools to be balanced:

But the way that it turned out in the end for the actual assessments, I don't think there was too much emphasis on it. I think it was a good amount of teaching time, and they dedicated the right amount of time to preparing people for what they were going to be assessed on. And the exam technique and things like that, there was a good amount of time for. **year 13 student, selective**

I don't know how else they could have done it if there wasn't emphasis on the assessments, but there wasn't too much, there wasn't too little – I was kind of right in the right amount at my college so, I was quite lucky. **year 12 student, college**

Other students, however, did suggest that the activities were unbalanced:

I worry that too much time went to assessments that maybe should have gone to preparing us to learn the content to the subjects we're doing next year that we have, we're up to par of what we should be to learn the content. **year 11 student, independent**

The scheduling of post-guidance assessments

A variety of approaches to scheduling assessments were described, depending on what the centre or department felt was best for their students. Some centres explained that they preferred to spread out the assessments, allowing a mix of revision and assessments after Easter.

We stopped teaching content at Easter and then we did a mixture of revision and assessment and sometimes mixed them. Because we have double lessons, so

some of these assessments were 20 minutes, 30 minutes. **Head of department, independent**

A student described a similar timetable that cycled between revision/exam preparation and then being assessed on the content they had revised.

We finished at May half term [...] the six weeks before that we did a week of exams, then a week of revision, and just kept doing a week on and a week off. So I guess three or four actual exam weeks. **year 13 student, independent**

Some teaching staff noted that with a more extended set of assessments, students had multiple opportunities to demonstrate their best performance but also to account for the on-going disruption.

I made the decision that having multiple effectively low stakes assessment opportunities spread out over a range of weeks and days during the week and everything, that would serve the best chance of giving students the opportunity to do these assessments and not be penalised because of the situation. **Head of department, academy**

Many other centres completed the assessments during a scheduled period after students had completed the course, mirroring an exam series. One deputy head of centre explained that scheduling a discrete assessment period was less stressful for students, suggesting that they would have been unable to concentrate on their learning were assessments to have been taking place alongside regular classes.

We did formal assessments [...] given the centre that we are, and different centres will have done this in different ways. But if we had tried to do in-lesson assessments, if girls had known that they were going into the second lesson of the day to do an assessment that was going to count in chemistry, there would have been no point in trying to teach them history in the first lesson of the day. [...] So we set up basically an exam period from 4th May to about 25th May, and we ran, for A-levels and for GCSEs we ran formal assessments during that time in all subjects to get final papers out of them. **Deputy head of centre, independent**

The same head of centre also explained that the students at the centre generally improve their performance towards the end of the course, meaning that the end of the course was a better time to assess them.

As I've already said our [students] do improve hugely towards the end of the course, so we were very keen to take advantage of that sensible suggestion that we should assess them at the end of the course. **Deputy head of centre, independent**

Teachers noted how they set the assessments to ensure that students had enough time between assessments to take breaks.

And we tried to do like an exam board would do and stagger it so your English and your Maths weren't back to back, your science wasn't back to back with your PE because there's a lot of overlap there for them. [...] Just supporting them really, just to make sure they could come in and have breakfast and whatever.

Teacher and/or tutor, comprehensive

Conditions under which post-guidance assessments occurred

Centres generally ran the post-guidance assessments under invigilated exam-like conditions. The replication of normal examination conditions was by far the most common approach described by our interviewees, including the provision of normal access arrangements. One of the teachers described how this worked in their centre.

The college also constructed an invigilation timetable and allocated rooms for these model assessments to take place in. I became the invigilator for a number of other subjects, never my own or none that I course lead, so I invigilated things like economics and drama in rooms that were unfamiliar to me or the students, on the clock, with assessment access arrangements in place, so 25% extra time for those that are awarded it, college made arrangements for the scribes, amanuensis where that was required, all of those kind of things. Other than the fact that we didn't use the official sports gymnasium as a formal exam room, pretty much everything else was as a traditional time-constrained exam, in silence. **Teacher and/or tutor, tertiary college**

One interviewee explained that they used a mix of open-book and exam condition assessments in their centre.

They sat this other paper as an open book. So they could bring all their notes. We gave them a notebook as well, like a homework book and we said, make notes on the videos you know, write the topic at the top that you're revising and then come back with that book and it's open book for your second exam. [...] And that was considered medium control because it was open book. [...] And then the third paper they got the topic list and the video clips but it wasn't open book.

Teacher and/or tutor, comprehensive

Some centres explained that they had back-up assessment papers planned in case any of the students missed any of the assessment sessions. In some instances, this meant having to prepare more than one assessment paper.

We did have back-up assessments. [...] So if a candidate managed to miss assessment one through whatever reason and also managed to miss assessment two or either/or, they were offered a third assessment which was different to assessment one or assessment two, so freshly created, because I'm afraid some of my colleagues were almost at their wits' end, because they created assessment one very carefully and meticulously and then they'd done the same thing for assessment two and then they unfortunately had a candidate that

hadn't been assessed by those methods so they had to generate a third and in some cases even a fourth assessment in order to offer that candidate an opportunity to take an assessment of some kind. **Teacher and/or tutor, tertiary college**

Advance information about assessments

The guidance specified that centres should only assess students on material that had been taught. Some teachers described how they informed the students of the general topics they were to be assessed on before the assessments took place.

I think we were also, we were instructed as well at some point to make it clear to students what the topics were going to include, but not so overt that they knew what the questions were. But to give them a very clear and public transparent overview of, what your assessments will contain. **Teacher and/or tutor, tertiary college**

An alternative approach in a few centres was to inform students of the content they would not be assessed on.

For the last exam, for the one in May we told them what, I think about three or four topics that weren't going to go on, but apart from that all the rest of the topics were on there. [...] The May exam we wanted them to achieve the best they possibly could without [...] telling them what was on there. **Teacher and/or tutor, sixth form college**

A few interviewees described how it was sometimes up to the students to work out what would be on forthcoming assessments.

If they'd been paying attention they could get quite a lot of hints even though we weren't explicitly telling them. Especially with the more able students sometimes it's half the game, it's not so much about their academic ability it's whether they're awake or not. Those more subtle things. **Teacher and/or tutor, academy**

Some centres used their revision time to emphasise topics that would then be assessed at the end of that revision session. This was a common approach when centres adopted a series of smaller and more topic-based assessments.

So, on the Monday morning they would come in and [revise], say, language and structure and then Monday afternoon they'd come and do [the] language and structure assessment. And that was the [arrangement for the] last six weeks. **Teacher and/or tutor, further education establishment**

One head of department explained that students were very anxious about the assessments and telling them what they would be assessed on was a way of making the process fairer.

Yeah, before we told them [the topics] there was quite a lot of anxiety around what were we going to cover and all that kind of stuff from some of the kids. So yes, I told them in advance just so that they knew what to focus on and they knew that it would be fair to them. Because, if they haven't covered some of the content then that wouldn't be fair. **Head of department, university technical college**

A teacher reflected on how students still had to apply the skills and knowledge in the assessments even if they knew the text they were to be assessed on.

They still had to apply the skills, so they knew which text we were doing, so we did, for example, a letter from a First World War soldier and an extract from Jack Munroe's book and the theme was food and they knew that, but they still had to go and do the analysis and some were amazing and some were really poor! [...] When I was looking at them and marking them, it reflected what those students' skills levels were at. **Teacher and/or tutor, further education establishment**

Not everyone felt that the way students were informed about the content of exams in advance was well managed. A minority of those we interviewed felt that some staff may have been giving overly -detailed information about, or teaching of, the content of the assessments.

I know in business studies for instance they tried to play the system. So they told the kids what was in the exams. They used a past paper, but I think they pretty much told them what topics were going to be in it. So when their marks came back they weren't spread, their marks came back very close together, which made it very difficult to draw lines and that was a very uncomfortable conversation with their head of business studies. **Deputy head of centre, selective**

And I think that some of the teachers, not explicitly but not far off it taught some of the test. So some of the questions, they would just have a look at the paper and maybe put them in as starters. **Teacher and/or tutor, comprehensive**

Student experience of the evidence-collection process

Preparing for assessments

Before considering the experience of the TAG evidence collection itself, students also spoke at length about how much of the course had been taught, and the effectiveness of remote learning. We do not cover this in detail as this was part of the teaching staff's consideration of content coverage in assessments used, described within the section entitled '[Content coverage](#)'.

As well as the in-class revision described in the section entitled '[Ongoing teaching and revision time after Easter](#)', in a few instances students we spoke to reported that they undertook additional independent learning activities, such as using free online learning resources and courses. Peer support as well as active parental support were also perceived as important:

Yeah, I mean we all were sitting there on our free periods helping each other, being like: "Oh, I think this question's going to come up, make sure you know that one." You could tell everyone just knew how important it was, and we were all trying to help each other out. Because if you know one of your friends is less good with memorisation of something, you'll sit there and help that friend with that. And if there's someone else who's good at memorisation but they might not understand it, you're helping them with that. So, we were all helping each other with our strong points, passing it on to each other, because everyone was just so stressed, and we didn't have time to do it all by ourselves. **year 13 student, college**

And then also obviously my mum is a single parent, and she has been so supportive. She's been just helping me get focused and just on top of everything. And so, she's been my main motivator along with my sister. And so, without them I would have probably failed all my exams, but they helped me get into revision and things, just so I could get those grades. **year 11 student, independent**

A couple of students also reported that they received additional support from their teachers who provided additional lessons on topics that were omitted during online classes.

But then my teacher said one thing that basically made them realise that learning 15 poems was a lot harder than just doing Christmas Carol. So, it ended up only me and another boy wanting to do conflict poetry. But because we were so adamant on it, our teacher ended up giving us "private" lessons in the last two weeks. I think it was before we went back in and that was quite intensive. **year 11 student, selective**

Experience of assessments

The students described a similar range of evidence types as described by the teaching staff we spoke to within the section entitled '[Evidence used to make judgements](#)'. We do not repeat this detail here.

A few students described the experience of remote assessments such as mocks before Easter. For some students, these pieces of evidence were collected under remote invigilation, while other were not formally monitored whilst undertaking these tasks, rather they were trusted to complete the work without conferring and checking against sources.

Online mocks were quite hard. Like doing revision online and then doing more revision and then having mocks in your house. We had to have our cameras on and facing our paper and stuff so they could check we weren't cheating. **year 13 student, college**

I remember that we did live lessons and our teacher adhered to the timetable. I think it was at the end of every week they'd give us a question that we did on our own and we'd submit it back to them. And you know, we weren't on a call or anything. And their justification for that is 'it will be evidence'. So, they gave us exam-style questions for us to type up and do without any outside influence. **year 13 student, secondary**

With regards to the degree to which students were informed what they would be assessed on in advance, the students we spoke to felt they had only a vague idea. This was typically where they knew the topic(s) the assessment would be covering, but nothing specific about the questions.

The most information we probably knew was things like, for example, English, we knew that the literature was going to be a Macbeth assessment and for language we knew that it was going to be either paper 1 section A or paper 2 section A. [...] And the other things basically we knew what topic it was on, like what big broad topic it was on. **year 11 student, selective**

For the college ones they told us some of the topics we should focus on, but they didn't tell us any of the questions that were specific. But for the ones I did externally, those ones I was just told to focus on most of the topics, I wasn't given any specific ones. **year 13 student, college, private candidate**

However, a couple of students pointed out that they or other students would often recognise questions and be familiar with the mark schemes and correct answers because the assessments predominantly consisted of questions from past papers that students had been provided with as formative learning materials, or were otherwise accessible via exam board's online resources.

Depending on the exam boards, we got to see a lot of the questions beforehand, and we had access to a lot of the mark schemes, it was very different, the preparation, compared to a normal exam, because you had an idea of the answers, so you could work for the exact questions more. **year 13 student, college**

Because you could access the mark schemes [...] it would kind of become more of who can memorise the mark scheme the best, rather than who knows the content, and who can do it as well. So they [the teachers] did kind of say that they weren't going to do that [use the past papers], but then actually they included a lot, most of the questions from the website. So then further down the line when people realised that they'd been the same questions they'd done from that

practise on the website, they were kind of like on it, 'let's try and do all these questions, revise the mark scheme'. **year 11 student, secondary**

Overall, most of the students we spoke to felt that they were assessed more under the teacher assessment arrangements than if exams had not been cancelled. For many this had implications for their well-being (which we discuss further in the section entitled '[The assessment period](#)').

It was like constant assessments throughout from when we returned to school until May and then at the end of May we did full exam conditions, there wasn't just lesson tests it was full exam conditions and then we did that again in June. **year 11 student, independent**

Those summer assessments, it was the week before we finished. They finished on Monday 24th May. [...] for the first day I had English literature, an hour; chemistry, an hour; RS, 30 minutes. And then the next day I had maths, an hour; biology, an hour; history, an hour. And then the next day I had physics, an hour; and English language, an hour. **year 11 student, selective**

There was also a case whereby a student felt frustrated and disappointed when they were told that a number of assessments they had worked towards would not contribute towards their final grade.

But we did so many assessments in those classes, but the school then came out and said that that couldn't be used as evidence because not the whole year had done every single one of those, so those particular ones couldn't be used as evidence. **year 11 student, selective**

Not all students we spoke to felt they were being over-assessed. A couple felt that they had been assessed less under the teacher assessment arrangements than they would under normal assessment arrangements. These students had experienced fewer, longer assessments, rather than the shorter topic-based tests other centres used.

No, I think we probably had less assessment because we didn't have the same number of papers for each subject as we would have had normally. And [...] it wasn't like we had extra assessments throughout the year, they [the centre] just do a lot of assessments in a normal year. And then at the end we only had one, like two papers for maths instead of three in total and whatever. **year 13 student, college**

Other students mentioned how much they would have been assessed anyway as part of the centre's normal learning and monitoring process. They did not feel as though the arrangements in their centre this year deviated much from that.

Awareness of the evidence used and a sense of agency

The students we spoke to varied in their knowledge of the evidence that contributed towards their TAGs. Some students indicated that centres had not been completely transparent so that they would not be able to determine the grade they would be awarded.

They've not told us properly, because they didn't want to make it too obvious what grades we're going to get to us. So, they told us that it's going to be a mixture of them, some from first year, some from second year, but I don't know which ones. **year 13 student, college**

In other cases, students were well-informed about the pieces of evidence and the ultimate weighting of the evidence for the final TAG.

So when we actually got back into college on March 8th, about that week and the week after, all of my subjects gave me a complete list of everything that they were going to assess me on - and it had all the information on it, so dates, each topic, how much it would count for and if we could like resit some of them. **year 12 student, college**

Indeed, a small number of teaching staff spoke about how they involved students in the choice of evidence that were used. A deputy head of centre explained that they wanted to be upfront with students around the evidence used to award TAGs and give them opportunity to comment.

The next stage for us was to communicate as per our policy to tell the students what pieces of assessment they would be assessed on. So that they then had an opportunity to come back to us, via a Google form, come back to us and say 'actually I don't think that's fair. I couldn't do my NEA' or whatever they wanted to tell us about, they had a full opportunity to do that. [...] It was quite focused on what the evidence was going to be. [...] A report to parents and students saying these are the four pieces of evidence that we're going to use to decide upon your final Teacher Assessed Grade. **Deputy head of centre, academy**

The degree to which students were aware of which pieces of evidence contributed towards their TAGs had a large impact on the agency they felt they had in the assessment process. Those who were not aware of which pieces of evidence would be counted experienced a diminished sense of agency. This, in some cases, led to an increased sense of unfairness and the perception of increased pressure on every piece of work, in case it was going to be used as evidence.

[I had] no involvement whatsoever [in the assessment process]. It was a lot of teacher meetings. [...] And it was so annoying, because they weren't talking to

us, they were talking at us. And they weren't involving us in the conversations that we should have been in, because these are our grades! **year 11 student, independent**

But we didn't have any input over it [the assessments]. [...] Since we came back in September it was sort of like the whole country was under the impression that every exam we did mattered. So, like every assessment week and class test was quite stressful because you thought it might be used as evidence and then actually none of it was used. **year 13 student, college**

A few students, on the other hand, did feel a strong sense of agency over their assessments and resultant grades. Some were even involved in discussion of which pieces of evidence would count towards their TAG.

[The teachers] gave us like a questionnaire with all the options that they were considering of how many exams and things like that, and they let us choose when we would do them. [...] [In the end] the teachers picked what they thought would be the best. They did ask if we agreed, but it all seemed to be the right sort of thing, so we just let them get on with that. **year 13 student, independent**

Evaluation of evidence and Internal Quality

Assurance

This section looks at how centres dealt with the evidence collected. In the analysis that follows we broadly separate the process into two stages. The first stage includes the marking and moderation of the evidence collected, including, in many centres, turning marks for individual pieces of evidence into grades. The second stage is the agreement process within departments to decide the TAGs followed by the internal quality assurance process (IQA) of the TAGs within centres, usually involving SLT.

Marking of individual pieces of evidence

This section describes the process put in place by centres to manage the marking of the individual assessments forming the evidence base on which to decide TAGs. This includes ensuring the same marking standard across multiple teachers and checking the accuracy and consistency of the marking. The next section describes converting marks to grades for each individual piece of evidence, where this occurred.

This process mainly applies to the evidence collected after the cancellation of the exams and mainly to the post-guidance assessments. Some centres noted that they had marked and standardised mocks and other pre-guidance evidence as standard practice at the centre, however this was usually not discussed in detail in the interviews.

Before describing the different marking processes, it should be noted that many schools and colleges had teaching staff who are examiners or moderators for an exam board. They were able to apply and share their expertise within their centres.

We started off by sharing across the Trust, everyone who's been an examiner in any subject, any year as long as it was new spec. And that meant we had a kind of pool of people who were that bit more trained that we could call upon. **Senior leadership team member, academy**

There were two main points at which the marking standard could be agreed: through marker standardisation before marking began, or through checking and comparing marks after the initial marking had been completed, a process which sometimes involved multiple marking. Some centres focussed on one of these approaches, while many implemented aspects of both.

As this senior leader notes, individual departments were usually left to manage the marking and standardisation process themselves.

I wasn't involved in standardisation or moderation, they happen within the team, so they all standardised and they all moderated, one before the assessments and one after, of course, to make sure that they are (a) all marking consistently and (b) using the grade descriptors or criteria consistently and effectively, so that was all done by the teachers and the Heads of Department. **Deputy head of centre, sixth form college**

We now look at the approaches taken for standardisation, marking and moderation within departments, in turn.

Standardisation of marking

Standardisation of marking involves agreeing the marking standard across the team of teachers carrying out the marking. In many respects, this followed similar approaches to exam board standardisation of examiners, which is not surprising given the number of teachers with examining experience. The mark scheme was discussed, together with marking and analysis of a selection of responses, so that everyone could gain a common view of how to interpret and apply the mark scheme.

The starting point for marking was typically the official mark schemes produced by the exam boards. We saw in the section entitled '[Post-guidance evidence](#)' on the post-guidance evidence used to make judgements that centres used a mixture of whole or part past papers or questions drawn from a variety of sources, including the materials that the exam boards compiled following the announcement of TAGs. In general, the mark schemes derived from past papers were felt to be useful in guiding marking. However, some found the mark schemes accompanying the assessment materials provided by exam boards to support the TAGs less helpful because they were perceived to be less detailed than those for past papers.

The mark scheme that [the exam board] provided for our shadow papers [the exam board materials] was a very basic mark scheme. [...] So I printed out the full mark scheme for the November paper, which it was based on, and used that to see whether or not I could give method marks or process marks to any of the bigger questions. [...] I felt that, in a true mark scheme you might get a process mark if it was a two mark question or a three mark question [...] for showing some working. [...] I think that [the exam board] should have provided a full mark scheme for those shadow papers to be honest for teachers. Not to be available to the wider public but teachers should have had a full mark scheme. **Teacher and/or tutor, comprehensive**

Some teaching staff reported that exam boards gave guidance and training on how to apply the mark schemes. However, usually it was felt that this came too late in the process to be helpful, as some centres were already well into their marking.

We used obviously the very late coming exam board training on how to apply the mark scheme. Half of it was arriving after the fact that we'd marked it but, we did the best we could. **Head of department, academy**

Teaching staff often commented on how significant efforts were made to create their own, in-house, training materials for standardisation before marking could commence. This head of department used exemplar scripts they had previously obtained as training materials.

We bought some papers from the exam board. [...] So, [this year] I sat at home with paint open, and I un-coloured all the ticks and crosses they'd put onto this work, found the papers where they came from, and basically gave them to the staff with the students' answers with the mark scheme, and then we all marked them. And then they gave them back to me, and I marked their marking if that makes sense. **Head of department, comprehensive**

Standardisation before marking was not always carried out though. In more objectively marked subjects where the students' responses are typically just 'right or wrong', this may have been less important. However, for subjects with more subjectively-marked responses, such as those with long answer responses and essay questions, this lack of standardisation was felt to be particularly problematic. It was perceived that different interpretations of how to apply the marking criteria would result in inconsistent marking without standardisation.

There was no standardisation, and what I understand by standardisation is you agree upon your interpretation of the mark scheme, which is very important with a subject that is essay-based, and that you use those grade descriptors from the exam board so that when you see something that is of a level four on the mark scheme you understand that's what you're looking at ... I raised this with the Head, that that was missing from their policy and it seemed to me they didn't

seem to understand the difference between standardisation and moderation.

Teacher and/or tutor, selective

Discussion between staff about the marking standard and how to apply the mark schemes also continued following the beginning of marking.

Marking

Fairness in marking was an important issue to the teaching staff we spoke to. There were two main approaches centres took to ensure the fair marking of assessments. Student work could either be anonymised, or student work could be distributed between markers so that staff did not mark just the students they taught. This might involve random allocations of students to markers, swapping whole classes between teachers, allocating specific questions (rather than students) to markers, or by having multiple markers for each piece.

Anonymising students' names was fairly common – typically, candidate numbers were used. Teaching staff felt this 'blind marking' approach helped reduce bias in marking.

We didn't have names on the front of the papers, so that the teachers didn't know who they were, so it felt like we were running our own set of external exams, so we did it as close as we could to that. **Head of department, independent**

I think the blind marking was a big development, there was a much bigger emphasis on unconscious bias and I think rightly so. [...] When [...] we did the grade analysis and [...] we were looking back at tracking marks and predictions and things it actually fitted, even after the unconscious bias blind mark, so that was reassuring. **Head of department, comprehensive**

The issue of recognising handwriting and undermining anonymisation was considered by some.

The students had used candidate numbers, which was good, and with lots of them typing work these days knowing handwriting, in terms of unconscious bias it is less of a problem, because we don't know their handwriting so much these days and so that was good. **Teacher and/or tutor, selective**

On the other hand, some centres decided not to use anonymised papers due to the risk of mistakes occurring.

We had a lot to discuss about whether we should have some sort of code on the papers and stuff like that. I think I was more worried that we'd get the codes wrong or something, and we'd end up putting the wrong marks onto the wrong kids or something. I wasn't really sure how that would work. **Head of department, comprehensive**

It was also common for centres to mix the allocation of students' work to markers. Several interviewees from larger centres reported that teachers were not permitted to carry out the initial marking for students they taught.

Each teacher would mark classes that they didn't teach and then they'd be moderated by another teacher that didn't teach [them] and when we're marking [...] we'd try not to be looking at the front of the paper to see whose paper it was and you'd just mark question by question. **Teacher and/or tutor, selective**

Other departments used a random allocation.

We'd got 80 students sitting sociology, I was teaching 40 of them, 20 were being taught by another person, 20 by a third person, and so the Head of Department took all 80 in and kind of shuffled them up, if you like, and divvied them out for marking amongst the three of us proportionately. **Teacher and/or tutor, selective**

Another approach was to allocate different questions to different markers, in a similar way to how live examination marking is often carried out by exam boards. As sometimes noted, there was less need for duplicate marking (such as in the moderation stage) when this approach was taken since individual harshness or leniency would cancel out.

The way we did it was we marked different parts of the assessments, so actually, although we had a lot of communication and a lot of chat about the different grades and how we were marking, we didn't feel we actually needed to countermark the other person's. **Head of department, independent**

Many interviewees described combining several approaches to marking: incorporating anonymisation and mixing allocations of students' work to markers.

We could spread out the marking. We made sure we were doing moderation as we went along. So we buddied up with one another, we changed the pairings, we did blind marking, we swapped classes with one another so we didn't know the students we were marking all the time. **Head of department, academy**

Multiple-person marking as part of the original marking process was mentioned several times in the interviews, outside of (and sometimes in addition to) a more formal post-marking moderation process.

In terms of marking, we anonymised all the papers. So, they had candidate number and not student name. We then marked students who were not our own, so the head of exams within the centre made sure that we were given batches of papers to mark. Every paper was at least double marked, most were triple marked. **Teacher and/or tutor, comprehensive**

It was clearly more difficult to implement such anonymisation and allocation approaches where there was only one teacher for a subject. The issue of marking

bias was also harder to control in these cases. However, later stages of checking, such as moderation - either by a direct line manager such as a head of department, or by SLT - could query unusual grades. As illustrated below, some checking across teachers in related subjects was often implemented.

So, for those of us that taught classes just by ourselves with no other teacher, we made sure to exchange papers, so somebody else had a look at them to see if it was roughly what they thought and then we discussed about why that was. And if there were any differences that were significant, that we felt like actually needed to be changed, then we did. **Head of department, independent**

One teacher from an exam centre for private candidates explained that they outsourced marking to an external organisation in those subjects where there was just one teacher.

We'd had an external marker do it, because the issue with history, although I'd done the training on marking it, I am the only history tutor and it is a subjective subject, so we did outsource some of the marking. So, we used [awarding organisation] for the mocks for the internal candidates and then when we marked in the summer we used an affiliate tutor and we both cross-marked some of the papers, she did some and that meant that across that 10 topics we were able to cover all of them with one or other of us feeling confident that we could do that topic. **Teacher and/or tutor, private candidate exam centre**

Moderation

We use the term moderation here in the same way as the teaching staff that we interviewed used it. During these interviews, the term was generally used to describe processes that occurred once marking had started to ensure that teachers were marking to the same standard (this was distinct from up-front training which tended to be described as standardisation). While much of this moderation occurred after evidence had been given a provisional mark, as can be seen in some of the quotes above, the marking and moderation processes were sometimes intertwined to a degree. Multiple-person marking and some checking was often employed from the start and throughout the process.

There are only two of us that were marking second year work anyway, so we largely sat and marked together and then would swap some things and moderate between ourselves to make sure that we were judging things at the right level, particularly with coursework. **Head of department, sixth form college**

In many of our interviews though, there was a more sequential process, with marking followed by moderation of those initial marks, where marks were checked for consistency and fairness. This usually did not involve blind marking, but was normally a review of the original marking.

For some, a moderation process was normal practice undertaken for regular scheduled assessments like mocks and coursework. Where this was the case, little additional work was required to implement this process.

We benefited from the fact that we were drawing on a practice that we'd done for years. We've moderated coursework. We've worked together for a long time. We've worked together when we were doing controlled assessments, and things like that. We moderate exams when we have exams, so we're used to that process. **Head of department, comprehensive**

Where moderation of assessments didn't routinely happen, more effort would have been required to set up such systems.

If there were any patterns observed across all of the interviews, it was that moderation involving multiple re-marking tended to be more extensive for subjects where marking is known to be harder, and perhaps more subjective.

Some departments did a lot more moderation than we did [...] I know languages, for instance, did a lot of it. [...] There's not that much in geography where the marking is [less] subjective, so, you're moderating for accuracy. Where[as] I think for some of the subjects, where perhaps it's a bit more subjective - the marking, they felt that they needed to mark a bit more, double mark a bit more, to make sure that that was as objective as possible. **Teacher and/or tutor, academy**

I think that by the time we got to the end of it, particularly with our science, maths and English colleagues, it was an incredible amount of hard work. English probably checked their marking I'd say a good five times. Because they wanted to make sure that they were absolutely fair and consistent for all children. **Head of centre, comprehensive**

We heard, in a few cases, that almost every piece of work that contributed as evidence towards the TAGs had the initial marking checked. But it was more common to use a sampling process during moderation, to pick out a selection of scripts to check that marking was consistent, accurate and fair. The script selection usually involved work distributed across the grades. Additional pieces of work for which the first marker had been less certain of the mark to award were also sometimes included.

We moderated within the department. So it would be top, bottom and anybody you felt really insecure about. A colleague would then read and go 'yes, I don't know this set work like you do, but, actually with this mark scheme and this information I would put them in this band'. So, we covered that. So, if there was someone who'd done outstandingly well on an essay that was unexpected, then it was checked by somebody else. **Head of department, independent**

Where inconsistencies in the marking were detected, it was usual for those involved to discuss their marks and come to an agreement.

I acted as second marker for physics and for biology. [...] If the first marker had maybe been a little bit uncharitable or had missed something in the marking scheme, for example, then I would go back to the first marker and we'd have a discussion and then come to some agreement [...] The job in hand was to go through every single paper to make sure that that error or that adjustment was fairly applied in every case. **Teacher and/or tutor, tertiary college**

As noted above, some form of additional checking was important where there was only a single teacher for a subject in the centre.

[For] single person departments, the college appointed a critical friend, so you had usually your line manager who would come and you'd have to justify marks. And we went through must have been about 30% of the papers. [...] My line manager [...] was delighted with the process, because his marking of the papers was actually quite close to mine, and where he was a little bit out he was absolutely satisfied that I was able to explain why. **Teacher and/or tutor, sixth form college**

While the anonymisation of students was carried through the moderating process where possible, it was also the case that more senior individuals involved in moderating departmental marking did not really know the students and could therefore moderate without fear of unconscious bias.

I was present for them and the same thing with GCSE biology, [which I] had nothing to do with teaching, it was just names on a list. The only rare instances where I did know the student would be if they'd done an EPQ as well by coincidence, and that's pretty rare I think. **Teacher and/or tutor, tertiary college**

In some instances, there was dissatisfaction with the moderation process, or lack thereof. In these cases, it was usually a result of insufficient time available to undertake an effective moderation process.

The moderation was done too quickly, so it was like after school, the Head of Department [...] wanted] to get off at a certain time and, again, when you go to examining board standardisation days they last as long as they need to, because no-one's going anywhere until you all agree. But this moderation process, there was pressure to just agree, let's get this thing done sort of thing. So, no, I didn't feel that was done very well. **Teacher and/or tutor, selective**

In maths [the students] put their name on and we marked our own papers. We didn't do any moderation. If I wasn't too sure about a mark I would go and speak to another teacher and say 'what do you think? What do we do about this?'. And other than that, there was no moderation. We didn't mark each other's papers and I felt we should have done. I felt there should have been a degree of moderating but there wasn't because I don't know, maybe we ran out of time or maybe nobody else thought it was important. **Teacher and/or tutor, comprehensive**

The sampling of student work to be moderated was also sometimes a concern for those we interviewed. Particularly where the sample only selected those pieces of evidence for which teachers were most secure.

Every single time, the Head of Department chose the top, middle, bottom of your pile... that was the same and every time, of course, the top student is absolutely outstanding [...] and the bottom candidate as well [...] it was the same candidate every time [...] you've got 80 students there, most of whom were only looked at by one member of staff. [...] I felt the sampling was not good and should have been varied. **Teacher and/or tutor, selective**

In addition to within-centre moderation, some centres also worked together more widely to moderate their marking, share expertise and compare processes. Where possible, this involved centres collaborating with other sites within their trust or group, or with other local centres or professional networks.

We've got another site, [...] a sister college. A colleague down there, she sent work up to me, I sent work down to her and so we're a pretty close team. **Deputy head of department, tertiary college**

Where we didn't have any subjects in the trust, like I said earlier, we used local networks. [...] We met up as a group of four schools, and [that] included a local college. We all scanned in our three samples: top, middle and bottoms, blind marked them a couple of days before and then met online and discussed the grades. [...] We've got a drama teacher who went to uni with someone who teaches in [a different part of the country]. They met up online and did theirs [moderation]. **Senior leadership team member, academy**

Grading individual pieces of evidence

After marking the evidence, centres used different approaches to prepare for the final TAG judgements. Some centres worked with collections of marks, while others converted the mark for each piece of evidence into a grade. In both cases this information was collated, generally on a spreadsheet. Note that the official guidance issued stated that it was not necessary to grade each piece of evidence. However, in our sample, most interviewees did speak about grading evidence.

Where marks were turned into grades for each assessment, there were two main approaches, using existing or devised grade boundaries, or using grade descriptors provided by awarding organisations.

Use of grade boundaries

Grade boundaries were not provided by the AOs for any of the assessment materials released to support TAG judgements, since these materials were not in the form of

whole papers. Some of the centres we interviewed used previous years' grade boundaries to support grading.

Use of the most recent November 2020 paper grade boundaries was sometimes mentioned. There were varied views about the appropriateness of these grade boundaries as they were for a small, atypical cohort of students that had been disrupted due to COVID-19. Rather, it was more common to see grade boundaries from 2019 being used.

You just put in the three marks that you had and with the formula that the head of department had put in, it would convert that to a grade based on the November [2020] paper because obviously the November paper was actually sat and we did have grade boundaries from that. [...] I know a November cohort is completely different from a June cohort but it was still the best that we had. So I know it wasn't perfect but it was still a national grade boundary. **Teacher and/or tutor, comprehensive**

Where we could, so where exam board resources were used, we then converted the 2019 grade boundaries into percentages and used those to actually arrive at grades. Because we felt the 2020 ones were not terribly useful, because the cohort had been so small, because it had basically only been those that had been very unhappy with their grades had gone for it. So, we used 2019 and matched that up. **Head of department, independent**

Grade boundaries vary across years due to the awarding process which adjusts for differences in paper difficulty. Some interviewees therefore derived boundaries by taking averages across more than one paper, although this was not always straightforward.

We had to base them on an average of the 2019, '18 and '17 grade boundaries, which I don't think was fair because the grade boundaries are very different from one year to another. There's at least like a 5% difference between one year and another. And I think the actual spread was about 10% in some of the grade boundaries and the grade boundary fits the exam, and we thought the one that was sat in 2020 would have been the one to use. I think from our SLT's point of view, they wanted something that was defensible if it came to an argument as to why have you used these, [...] and they thought that the 2020 one was too different to be realistic. **Teacher and/or tutor, selective**

This head of department describes how they created an ad hoc grade boundary using previous papers.

For 80% of the students simply adding up their individual marks, turning that into what would have been a full exam series and using existing grade boundaries would chuck out a reliable grade. We used a previous grade boundary. We were told we weren't allowed to use November 2020 even though that was the only exam series that we had left because they didn't issue us with any new ones and

we'd used all of the others. So we were using November 2020 exam papers with somewhere between the November 2020 and 2019 grade boundaries because the 2020 boundaries were the only things that we had that had taken into consideration a COVID situation. **Head of department, academy**

Those centres that created bespoke tests using questions from several past papers also chose to devise their own ad hoc grade boundaries.

Because we'd already given them lots of previous questions we didn't give them the same ones [for the assessment, they] came from different exam papers. So, we kind of took an average grade boundary from each of the years that the questions came from. **Teacher and/or tutor, independent**

Often, interviewees spoke about circumstances in which the existing grade boundaries were no longer appropriate, particularly where students had seen the content of the assessment and mark scheme beforehand. In these cases, the grade boundaries were set higher to account for this.

The great difficulty we had was setting grade boundaries, because it was like we had grade boundaries for the papers that had been set externally so those were our guide, but our students had done way better than those grade boundaries. [...] Because our students, 'here's six exam papers to prepare, your questions are going to be taken from these'. So, they did. They literally learnt it - they were able to learn every answer. And then the marks were really high, so [we] had to set massively high-grade boundaries. **Head of department, independent**

My psychology teacher said I've had to not use the grade boundaries, because if I had 60% of my students would have got an A*. So he's moderated his down, because he said the grade boundaries were too [low]. **Head of centre, further education establishment**

The TAG guidance was intended to be flexible, so it is not surprising that a wide variety of approaches were taken in devising and applying grade boundaries to assessments. Sometimes pre-defined grade boundaries were not used, but instead grades were allocated based on an expected distribution of grades.

So if we did a test, one person would mark one section. We would then rank those students in terms of those marks and we would look historically at how many A*, As and Bs, that we get, then we would look at the average GCSE point score across the whole cohort to see what we thought the strength of the cohort was and then we would set grade boundaries on the basis of that. [...] So generally when we do effectively what exam boards do, although a much more simplistic version, is that we have in mind historically how many grades are likely to be got at the end. We rank the students in terms of their percentage and then we award say the top 30% an A* prediction and the next 40% an A prediction. **Teacher and/or tutor, selective**

Use of grade descriptors

Grade descriptors were provided by the exam boards and the JCQ as part of the materials to support teachers with their TAG judgements in GQs. It was expected that these would be used to help decide grades by detailing the expected level of performance at specific grades in a subject. Because of the kind of difficulties noted above with using fixed grade boundaries, especially where the departments had devised their own tests, these descriptors were intended to support decision-making around the grade-worthiness of student work. They would also create a level of standardisation across centres, as they should all be judging performance against the same performance standard.

From our limited sample of teachers, it appeared as though grade descriptors were considered most useful for essay-based subjects.

In terms of the materials, we didn't have a grade boundary. So, there was this whole, really difficult nut to crack as to what percentage have you got to get in this exam to get [each grade], and none of it's nationally standardised. [...] So we used the JCQ grade descriptions, particularly in subjects like English language, history. That really helped us to sort of root those [students] to the grade descriptors, stack up our students in rank order and then look at where they were hitting and where those grade boundaries came. So, we used 2019 grade boundaries just as a, 'right, OK we've got to start somewhere'. [...] Is that what you would expect? And then we dovetailed that with the grade descriptors as well. **Deputy head of centre, academy**

As the above quote illustrates, grade descriptors were often considered helpful when deciding grades for students whose performance was close to a boundary between two grades.

In several subjects, such as mathematics and the sciences, teachers noted that the grade descriptors were not very useful due to the types of assessments used, which were based more on accumulating marks for correct answers than for demonstrating skills at different levels of performance.

If the students were a bit borderline, then we would look more closely at their papers, we'd looked at the grade descriptors, and we were trying to look for evidence of the grade descriptors at certain grades. That was a quite difficult thing to do, because the grade descriptors as a mathematician were quite woolly. **Teacher and/or tutor, independent**

It was noted that question-specific descriptors may have been more helpful for these subjects than the broader skills-based descriptors.

Looking at things like the grade descriptors for science were just useless, just utterly useless. I couldn't work out how to apply any of those to science because if I don't know what a grade 8 is in terms of the exam question then how do I

apply a grade descriptor to that. How do I know what I'm looking for, it was meaningless. **Head of department, university technical college**

Quite a few teachers commented that the descriptors were just not good enough for effectively dividing all the students up into the correct grade for each piece of evidence. They suggested that this was because too many of the descriptors were based on relative comparisons to other grades.

There's been a constant focus on this word holistic [...] and there was the constant focus on grade descriptors, but in reality grade descriptors do not let you grade a piece of work. They are so unspecific, well I mean there were no grade descriptors for grade 9, and there were no grade descriptors for A*, as there never are. So to get a grade 9 you had to exceed the descriptors for grade 8. Well what does exceed mean? **Deputy head of centre, independent**

Inconsistency in the grade descriptors across subjects were also noted, here from a teacher of religious studies and classical civilisations.

The grades descriptors given by JCQ were wildly different from subject to subject. So my grade descriptors were so incredibly vague they could have been interpretative dance. That's how interpretative they were. But I know the science JCQ guidance were incredibly specific. So there's no parity with the grade descriptors. **Teacher and/or tutor, sixth form college**

Some interviewees suggested that particular grade descriptors were stated at quite a granular level, based on a sub-sample of questions within assessments, meaning that there was little evidence to compare to each descriptor.

We had to use the grade descriptors to come out with a grade for the test. Now in some ways they were quite hard, because some of those tests were then really short. So where you got a 30 mark test, we had five grade descriptors covering the five different aspects of science. One of them was linked to say mathematical skills, but they've only been two marks of maths in there. So it's then very hard when you've only got two marks to determine who's grade 9, grade 8, grade 7, grade 6. [...] Everything we read basically said you can't just use marks. We have to be judging them on grade descriptors. **Head of department, comprehensive**

Several teachers also commented that the grade descriptors were not consistent with grade boundaries from past papers.

Historically over the last five years since we've had number grades, [...] it's always been two thirds [of marks] roughly is a grade 9, one third [of marks] is a grade 4. If you then read the grade descriptors, for a grade 6 it says the student has mostly accurate understanding; for grade 8 it basically implies it's nearly perfect; and then grade 9 is beyond a grade 8. So it's what do you take as mostly accurate? Because I would say mostly accurate is probably about two thirds of it

is right. But that's a grade 9 on the questions we're using. **Head of department, comprehensive**

However, while the grade descriptors were considered of variable utility in grading individual evidence, it is worth noting that they were considered more useful when considering TAGs, as a more holistic indicator that the TAGs were aligned with the level of performance demonstrated across all the evidence collected.

Calculation and/or judgement of TAGs within departments

This section describes the different methods used in centres to interpret the evidence they had collated and determine TAGs for their students. This covers the initial work within departments. Later stages of the process, including internal quality assurance (IQA), are covered in the section entitled '[Quality assurance of TAGs](#)'.

It is important to recognise that these two stages often overlapped, as although IQA was led by SLT, it was also partly the responsibility of departments, and there were frequent discussion between SLT and departments regarding the TAG judgments during this process. The overriding impression from the interviews was that the process involved a significant amount of debate and discussion, from the marking of evidence up to the final agreement of the TAGs.

TAG judgements were usually supported by a grid of either grades or marks for each piece of evidence for each student on a spreadsheet.

So we got together as a department, and we obviously had all the data, all the results for all the pieces of evidence. So that was one picture that we had, and then what we did was we put together some spreadsheets, mathematicians as we are. So out of those three November papers that we've given them, one for classwork, one for homework and one for test, we picked out questions that hit the different assessment objectives, and we put those into a spreadsheet. And we put in how many marks they got for those questions, and we colour coded the spreadsheet red, amber and green, and then we kind of used that. **Teacher and/or tutor, independent**

The number of pieces of evidence combined to determine the TAG varied considerably across centres. Our companion survey suggested that most TAGs were based on 4 to 6 separate pieces of evidence.

The process of arriving at a TAG often involved combining the marks or grades from several assessments with teacher judgements and/or other evidence. Those who had converted each assessment into a grade could make a judgement about the level at which the student was working by looking across the grades, or by applying

an analytical calculation-based approach. In both cases, different evidence could carry different weight, based upon how well it was thought to represent student performance.

Those departments working with marks, often applied an analytical approach as well, differentially weighting marks for different evidence, to arrive at an overall mark across all selected evidence. This generated a rank order of students, and grades could be allocated either by a quota-like system based on how many students were expected to achieve each grade, or by considering the student evidence against the grade descriptors, and deciding where the division between grades fell on the rank order.

However, this was not an entirely abstract, mechanical process based on the spreadsheets. In most centres other evidence or considerations beyond assessments were considered and a holistic judgement by teaching staff was usually applied to ensure the TAGs were fair and reflective of overall student performance. To support this, folders containing the evidence for each student were maintained for easy access, so that they could be referred back to. In addition to the marks, the spreadsheets of evidence described above also often included additional pieces of information, such as tracking or predicted grades, attendance data or noteworthy circumstances for individuals.

Although we describe different approaches separately in the next sections for clarity, those we interviewed explained that a combination of data points, weighting of the evidence grade, grade boundaries and descriptors were used and balanced alongside the teacher's holistic professional judgement.

Weighting evidence when determining TAGs

When combining evidence to determine a TAG, the majority of teaching staff weighted evidence according to the conditions under which the evidence was collected, the content or type of evidence, or the date it was completed. Only a minority reported weighting all the evidence equally.

For instance, those assessments that preceded the announcement that exams would be cancelled were not weighted as heavily as post announcement assessments. Partly this was because students would have performed best right at the end of the year, while some teaching staff expressed that it would have been unfair for such older evidence to carry substantial weight as students did not know at the time that it would count towards their final grade.

We had a policy for exactly how we were going to apportion the grades and what weighting. I know I said ultimately, it's holistic, but what we've got to think of is the stuff that they did before March when they didn't know that these were going to count towards their grades, were only going to give a certain percentage of weight to, the subsequent ones we will give considerably more. **Head of centre, further education establishment**

A few teachers highlighted that assessments with the highest levels of control, such as assessments that were completed in class, were weighted more heavily compared to evidence such as homework or class work, which were completed under low levels of control.

So, a normal classroom test we weighted as one, if it was an online test we weighted it at half because they're not as trustworthy, if it was a big exam test which was multiple papers, then we effectively, if it was three papers in an exam hall, we weighted that as three, because we took the overall and we weighted it three times. **Teacher and/or tutor, selective**

While the post-guidance assessments usually carried the most weight, in many cases, because the TAG was based only on these assessments, all the evidence was weighted equally.

And so we ended up with these five assessment grades and then we had a meeting, they were all on a spreadsheet and then if there were three grades, if it was A, A, A, B, D, then the student got an A, if there were three grades the same that was what they got, or that was the TAG that was recommended to the board **Teacher and/or tutor, selective**

Although weighting the evidence was one of the approaches taken by centres, this was usually used in combination with other approaches discussed in this section. For example, this teacher explained how they used a combination of weighting and historic grade boundaries to inform the TAGs.

The senior leadership team decided what weighting each of the different pieces should have and so then I just added up all the different percentages, times, whatever the weighting was and divided it by the total and then that gave me an average percentage and then I used the grade boundaries from 2019 to have a look and see what grades they had. **Head of department, independent**

Use of historic grade data

As suggested in the guidance, centres often used data from previous years, when summer assessments had taken place, in the process of determining or checking TAGs. This was used both to guide judgements up front and as a check to ensure grade distributions were fair. It is worth noting that comparing TAGs to historic grade data was more usually done by SLT during their IQA process, however in this section we report some examples of the use of historic data within departments to make initial TAG judgements.

Looking at the centre's previous cohorts was often a starting point for the process, to lay out an expected distribution of grades.

In terms of the getting to the grades [...], we were looking back at historical grades within the school, we were looking at historical grades nationwide and

working out where we fell within that. [...] there was the expectation that our grades would be roughly in line with previous cohorts unless we could justify why they were not. **Head of department, independent**

When taking this approach, many centres also considered the ability of their current cohort relative to past ones to adjust their expectations, noting this was not unlike part of the normal awarding process for exams.

We can see historically that the GCSE point score is the best indicator of A-level outcome. [...] We would look historically at how many A*, As and Bs [...] that we get, then we would look at the average GCSE point score across the whole cohort to see what we thought the strength of the cohort was and then we would set grade boundaries on the basis of that. [...] We do effectively what exam boards do, although a much more simplistic version, is that we have in mind historically how many grades are likely to be got at the end. We rank the students in terms of their percentage and then we award say the top 30% an A* prediction and the next 40% an A prediction. **Teacher and/or tutor, selective**

Interviewees reported differences in which year's results were used as a reference point for TAGs. One teacher noted that checks involved benchmarking TAGs against grades achieved over the last three years. They explained that it was reassuring to see a slight reduction rather than inflation of grades awarded.

And we did grade comparisons over the last three years and it fitted with the grades we gave, they fitted very much closely with what we'd had [in the] last three examined years plus last year and they fitted very much in the distribution of the different grades. In fact I think the head of centre said, oh this year we were slightly lower I think with some of our history grades and we could have issued, in fact we went back and looked again and just checked our judgement, but we were still happy with what we'd given, so we used our professional judgement at that point. **Head of department, comprehensive**

However, there were concerns from others where comparisons to 2020 grades were made. This was seen as problematic due to the increased outcomes seen that year.

There was just an awful lot of pressure, a lot of pressure. And they [SLT] were comparing our grades [to] last year's grades which none of us believed had the integrity of the previous year's grades, so we wanted to base it on [2019] because last year we had a 10% increase in our profile. Which is phenomenal. **Teacher and/or tutor, comprehensive**

Analysing TAGs according to student groups and protected characteristics was also considered important for identifying any potential unfairness, and was an opportunity to adjust preliminary grades.

When we put our initial TAGs in the first time, they ran checks on the data versus things like disadvantaged students, EAL students, and came back and said

whether or not there was a significant disparity in your data, if it looked like some groups might have been disadvantaged. And I think if I remember rightly I was told that the change between the last predicted grades that we'd made for the students versus the final, well not the final TAGs, but the initial TAGs that we'd come up with, SEND students in my cohort had made significantly less progress. Do you think, is there anything in that, do you think that there's something you need to look at here? Have students been unfairly disadvantaged? **Teacher and/or tutor, sixth form college**

Professional judgement and/or holistic approach

Although the process for determining TAGs was driven by the evidence collected for each student, teaching staff we spoke to usually used their professional judgement and considered students on a case-by-case basis. In this section we analyse how teachers combined the evidence collected and the more analytic process described above with their professional judgments to achieve a holistic grading of each student.

While data points were usually the starting point for determining the TAG for each student, the grading process was intended to be holistic to account for the conditions under which evidence was produced. This head of department explained how they used data, but also tried to consider broader factors.

So we had the weightings and we were given templates for spreadsheets that we put it into and then from that you get the average percentage and the average grade and then the pupils had submitted any extenuating circumstances and we put those into the spreadsheet as well and from that you could then make a holistic judgement which is what I was trying to do, which is why it was so difficult. Because if it had just been numbers then whatever the average was you could have just given that and that would have been fine. But having to think about anything else that was going on, that was where it became slightly more complicated. **Head of department, independent**

Many departments used a more data-driven approach to produce initial TAGs and then used their professional judgement to check the grades.

The process we both used was, do the grades arrived at with the formal assessment give us a single grade? So, it was easy, in a sense, if a candidate had got B/B in that final assessment, you'd go, I think that's probably a B, and then to support that B we would look back at the previous 10 or 15 assessments and look at the pattern, so if they got a lot of Bs or they got a lot of As and Cs, that would probably support the judgement. **Teacher and/or tutor, tertiary college**

Some teachers suggested that basing TAG decisions primarily on the data provided a methodical approach, which was important for ensuring consistency and fairness

across students because it provided an unambiguous process for making judgements.

You've got a series of numbers and percentages and what the equivalent grade was and then it just became a very mechanical, methodical process that was a lot less stressful and it meant that you could be very consistent, because you'd see patterns emerging and you were able to take a consistent approach. **Senior leadership team member, independent**

A minority of centres implemented explicit formulas within their data spreadsheets to automate the weightings of the various evidence, which was then checked manually using professional judgment.

I went through manually and did lots of manual checking of pick a random student, manually check that not only has it picked the best five [pieces of evidence], has it calculated the weightings properly, has it averaged them right, has it done the right calculation, have we ended up with the right result. **Teacher and/or tutor, selective**

A similar process was mentioned in determining TAGs for vocational qualifications.

For the vocational, for level 2 it was quite a lot easier because the different boards already have a grading calculator that we use to guide us. And so if they had a merit on one unit and they would have got a distinction on another unit we were able to use that calculator to guide us. **Senior leadership team member, academy**

Teaching staff told us that a holistic approach was particularly important when formulating TAGs for students who achieved a variety of grades throughout the assessment process and when teachers/tutors felt some of the evidence was not representative of the student's skills or ability.

It's where you've got a student with C, A, D, B, D, C a real kind of up and down spikey profile. And yeah, so you look at it and you're like well this kid's got X number of As, X number of Bs, X number of Cs, we would say they were a B, say. And yeah, so we did plug it into a spreadsheet just to see what it came out with at the end. Add up the data effectively and there were some students it's like well yes, often we used it to basically support our professional judgement. **Deputy head of department, tertiary college**

I had a fight with one member of staff who got quite annoyed because he's had a, and I'm being horribly stereotypical, but he'd had a typically bright but lazy boy in the group who really hadn't worked particularly hard at all until March. And on his final two assignments he got A*s. And I said, well looking at everything I think you've got to give him an A, his previous work isn't as good and these final things don't cover everything. And he banged on the table, he said but he doesn't deserve it, he's not worked. And I said no, we've talked about this, objective

evidence and that's why I did it that way really, to try and make sure that we could guarantee the objectivity. **Head of centre, further education establishment**

Such students were also the focus of discussion amongst staff. They recognised the difficulty of the judgement and sought out additional expertise.

We did have some students who were a little bit all over the place. So there were some students who, for example one student he had Us, he had As, he had Cs, so then it was just a bit confusing as to what to do with them. So it was a bit, for the students like that, I made my judgement and then I went to my colleagues, we spoke about it as a department, then we spoke about it as a faculty and then the faculty lead then took it to the principal and then, well the assistant principal was leading it and then it was a decision made by, well looking at it as a whole rather than just me looking at it individually. **Teacher and/or tutor, sixth form college**

Professional judgement was often supported by discussion and the consideration of the wider experience of individual students, beyond just considering evidence in isolation.

We then got together as a group and the three teachers who taught them discussed each individual, looked at their grades. We did have, there was a kind of holistic view of things as well because there were some students who'd had, who'd really suffered with mental health issues and some who'd missed quite a lot from self-isolation and things like that. So again that [evidence] wasn't necessarily a fair reflection on their ability so we discussed and used other pieces of information, evidence. **Head of department, university technical college**

The published grade descriptors were used at this stage to check whether student work demonstrated the described performance level for the TAG. They were particularly useful for students whose performance was considered to be near a grade boundary.

For about 80% of the students, they were sitting in the middle of the grade boundary, it's nice and simple, that's the grade that they were given. And then sort of as like a sampling, internal sampling thing we would choose random folders and check them against the JCQ grade descriptors that they had issued. And just check that it sort of fitted with the written descriptor of what that grade might look like. [...] Then for the borderline students we had to essentially go through the folders themselves of the TAG assessments and we had to make a holistic judgment based on the grade descriptors. **Head of department, academy**

It is worth noting that the more qualitative type of judgement was not always applied. One teacher spoke about an analytic approach at their centre that excluded their qualitative judgements and that they were not able to challenge. They felt this could lead to unfairness for some students.

They took each piece of evidence and they gave it a weighting, but the weightings were different. And the final results that this thing generated we felt was not in line with what we would have predicted the students. So my own dissatisfaction with my centre primarily is focused on that really, I mean I think the intention was good, because I think the intention was to say actually teachers are not choosing these grades, this is how we're going to use the evidence. [...But] some of these kids [...] if their external assessment, the really important one, if they crashed and burnt on that, that [...] was a big part of the weighting. **Teacher and/or tutor, selective**

Quality assurance of TAGs

This section describes the internal quality assurance (IQA) process, in which SLT usually took a leading role. The majority of those we interviewed described a process where, once agreed within the department, TAGs were shared with SLT members for a check. In most instances the head of department was then involved in one or more meetings with SLT where the TAGs were discussed. For single-teacher qualifications the same process was followed, with the teacher meeting with SLT. The TAGs were scrutinised and the role of SLT was mainly to ensure that they were a fair reflection of the evidence combined with the teachers' professional judgment, and that the centre policy had been followed.

A member of SLT explained that they had multiple layers of checks in place, which was fairly typical across the centres of our interviewees. There were also checks for any administrative error.

Me and the head of department would have looked through that folder for a final, final time, not changed any marks because we were agreed that those papers had been accurately marked but potentially changed the grade. [...] So I guess the only other thing we did was that as a senior leadership team we took that whole tracker then when they'd all been entered and did a line by line. So we just went, we had a massive sheet of paper and went through each kid. [...] So after the line by line the head of centre and the exams officer then spent a significant amount of time then entering those and making sure it wasn't a glitch. You know, you hadn't accidentally put 2.3 when you meant 3.2, those kind of things. **Senior leadership team member, academy**

Several senior staff explained that the role of SLT was to scrutinise and challenge the decisions, but that the final decision was down to the professional judgment of the teachers and their departments.

So, when we were getting towards the stage that all the decisions had been made, we met with every curriculum leader, we scrutinised all the judgements that they'd made. We [were] sort of picking out students and looking at 'how did we arrive at that'. And then from that meeting there would be actions and then

that would arrive at the department head signing the declaration. [...] It's got to be reasonable and, where you questioned it, the curriculum leader went away, they reviewed their evidence and they brought it back to us. [...] It was only that they had to be happy because it was their results. They'd made those professional decisions. Our job was to challenge them really. **Deputy head of centre, academy**

They then did go up to be checked by senior management, our director of quality looked through all of ours. And they basically pinged a couple back and said you look at their nine pieces of data and you've put them down for an A, are you sure? Shouldn't they not be a B? And we just went no, professional judgement, it's an A. There were a couple they were pushing for us to lower their grades and we were adamant, we said no, because that is their level. **Deputy head of department, tertiary college**

Most of those we interviewed explained that if any changes to grades were suggested by SLT, head of departments or the class teachers then made decisions based on further consideration of the evidence, but in discussion with SLT.

It was fed back, [...] the Head of Department went back to the teacher, they discussed it. If they didn't agree then they came back to us, if they did agree changed it. **Deputy head of centre, sixth form college**

In the final SLT line by line check of 120 kids doing eight, nine subjects each [...] the action from that meeting was we sent them back to the middle leader and said could you just tell us, can you tell us a bit more about this grade. Are you really sure? And most of them stayed. We maybe changed a tiny amount of those. [...] So we got the second in department and the head of department and got them to talk us through the evidence. So it was always very collaborative. **Senior leadership team member, academy**

The guidance published by Ofqual and JCQ advised centres to consider previous years' results as part of the quality assurance check for TAGs. We found that this was frequently described as part of the final IQA check by senior management, although it was often also part of the checking process within departments, as noted earlier.

A number of interviewees reported that this sort of checking conducted by SLT resulted in requests to review and/or adjust grades. This senior leader described how departments where grades were not in line with previous years' data were asked to check their judgments against the previous results and justify any differences, but to ensure they reflected the evidence.

I did ask departments after they'd put in the grades [...] to have a look and see how it fitted [to 2019 grades]. And if they exceeded, or were under the 2019 grades, to provide some sort of rationale as to why that might be, so that we could then discuss that. [...] I mean I suppose that a teacher's holistic judgement

about what looks like a good distribution of grades is going to be shaped by the historical experience anyway. So we didn't do a massively formal process, and I certainly wanted to retain the principle that an individual would get the grade that their evidence suggested, because that has been the really important promise to youngsters this year. So it's very difficult to match that with in 2019 we only got 20 grade 9s, so we can only give 20 grade 9s this year. So we certainly didn't do that. **Deputy head of centre, independent**

The majority of class teachers we interviewed did not perceive any pressure to adjust their TAGs based on past performance.

They did question departments where you were out of line with prior data of several years. But the argument used, as long as you feel confident you can justify it if someone came in to look at the work, then that's fine. [...] I don't know of anybody that was told to put grades down, which would be very dodgy to be told that. That wasn't our experience of SLT, in the end they were pretty reasonable about it and realised that it's not a perfect system and probably will have some grade inflation and we'll just have to live with that. **Teacher and/or tutor, academy**

However, a minority of interviewees described greater pressure not to diverge too much from historical results.

When I had my meeting about the grades with my line manager, the deputy head, she said to me that some of my classes were fine because they were in line with the previous percent, like historic data, but some of them weren't and that was going to be an issue. And we discussed it and I said I thought this was the fairest grades for the students and she said she understood that, but that overall, she said 'think about the school as a whole'. **Head of department, independent**

The minority that experienced pressure to adjust TAGs to match previous years' performance thought this was unfair to their students.

The other thing they looked at was previous years' results. So if for example we gave too many A* this year we had to then decrease the amount of A* we gave just so that it was in line with the previous year, which is a bit of a difficult one isn't it, because you're thinking this student can get a grade A or an A* but I have to downgrade them because previously we didn't have this many As and A*. So it was a bit, yeah I think it was quite unfair that one in particular because it's not truly reflecting what they're going to achieve, it's reflecting what we as a college would have achieved. **Teacher and/or tutor, sixth form college**

In cases where class teachers were not directly involved in the IQA process, some reported that adjustments were made based on historic data without their agreement.

Then we submitted the grades, our head of department also had to check it. And after that our head of department had a meeting with a member of the SLT, and they had to double check it. [...] And then I think, yes, there was another process then, the trust looked at all the grades. And I know that there were a few students whose grade was pushed down, again because of the historical data. **Teacher and/or tutor, comprehensive**

Some of those we interviewed explained that the meeting with SLT worked as an additional quality assurance check. Members of SLT were often able to make a comparison of TAGs across all subjects for each student which could flag up issues.

We had to take the grades to the director of studies and we had to talk through with the director of studies exactly what we'd done and how we'd come to everything. So she was the top quality assurance [...] because she had a whole spreadsheet of data and we went through it and discussed it all. We agreed on it but she also looked across, she actually looked at the candidate across the whole. So if we looked at one candidate and we looked at what they might be getting somewhere else and we look at their previous assessments there was a look at yes we would expect them to be about there. So she was bringing the next level of assurance by checking across other subjects and what she knew of the candidates. **Deputy head of department, independent**

Others did not view the quality assurance by SLT positively, sometimes suggesting that the main quality assurance had in fact been carried out within the departments themselves.

So we did have somebody overseeing our department, but she had no idea about the work of the department. Luckily she was a maths teacher so we could actually get her to look at, check the data for us, but that was very much a tick box exercise. So I think in terms of the quality assurance and I think that was very much up to the discretion of the individual head of department. **Teacher and/or tutor, selective**

The vocational and technical qualifications TAG process

The findings in previous sections reflect a predominately GQ perspective. While many of the observations also apply somewhat to VTQs, as their normal assessment arrangements are often different to those for GQs there were some differences.

It is worth noting that this was a small sample of staff focussed mostly or entirely on VTQs. The main differences lay in the timeliness and detail of the guidance that was provided to support the determination of TAGs in VTQs, as well as how evidence was combined to arrive at a final grade.

Timelines and detail of guidance for TAGs in VTQs

In general, there was a feeling that VTQs had been overlooked by government, and that the focus of announcements and information was on GQs. One senior leader explained that the decision in January to cancel the exams was announced the day before the first BTEC exam in their centre started. They felt there was a lack of information about the arrangements for these assessments, and this caused confusion for students and centres.

[The cancellation of the exams] was announced the day before the first BTEC exams and it seemed to be that the vocational exams had been forgotten. And I was in conversations with my head teacher at half past five on the Wednesday morning, when the announcement was the night before,' to say, well what are we going to do?' [...] Because the announcement was just GCSE and A-level when the next day was the start of the level 3 exams and it just seemed like no one from Ofqual, [or] Department for Education, had even given a thought to all these thousands of students that were sitting exams that morning. **Senior leadership team member, academy**

The policy position from the Department for Education was that the exams in January should go ahead where it was safe to do so. It was up to centres to make that decision. One teacher explained that most of their students completed the exams, however, they had heard that at other centres exams did not go ahead.

The teaching staff we spoke to from a VTQ perspective felt that once guidance was issued by exams boards for GQs, the guidance relating to VTQs from some AOs was often missing or was lacking in sufficient detail. As a result, centres considered guidance from other AOs and for other qualifications to start to build an understanding of what arrangements they could put in place. This created additional administrative work for centres.

This is where it was really difficult with vocationals, because each board, we were getting information from one board whereas another board wasn't really giving that much help or guidance and it was just, well as a centre you sort it out. So, we were having to mix and match from what we were getting from different boards and going, 'well that looks a good idea, we'll roll that out for all the others' because we've got some guidance there. **Senior leadership team member, academy**

Once centres had considered the guidance and were planning the assessment approaches, centres that offered both GQ and VTQ qualifications generally aimed to treat them similarly to ensure consistency in the process, where possible. Often, those who designed the processes for the different qualifications worked together to align their approaches.

There was slightly different guidance for the vocational. So, actually my colleague that I was talking about earlier, she leads on the vocational and I lead on the academic. So, we worked absolutely in tandem, but of course we made sure - so the TAG basket had to be different for vocational, the rules had to be different, because of course you're dealing with a course that is assessed in a slightly different way under the normal circumstances. So it wasn't exactly the same, but the bare-bones of it were the same. The way we communicated was the same and we tried to do it, again, all at the same time. **Deputy head of centre, sixth form college**

Combining evidence and arriving at final TAGs for VTQs

In general, those involved in TAGs for vocational qualifications felt they were in a strong position to determine grades given that they were largely based on the same pieces of work that would have contributed towards grades in a 'normal' year. In these cases, much of the students' work had already been completed and marked and so judgment was relied on less.

We were in a situation with IT for example, where the students had sat the external exam in January, because they did still happen. They'd all got merits and distinctions. Their assignment work got sent off for standards verification and we were like well they've done the qualification. This isn't a 'holistically what do we think this student might be getting', they've done the qualification, which obviously with the A-levels was a different ball game. **Head of centre, further education establishment**

However, judgement was used in awarding TAGs for VTQs in some circumstances, particularly where a qualification-level TAG was required. Teaching staff considered the conditions in which the assignments had been undertaken, and the degree to which they were different to a normal year.

But we've got the four units and we've done them, and we had to go, 'yeah, but think about the conditions that the students sat those in'. [...] because things had been done at different points, and we're going, [...] 'yeah, but remember that happened then' and 'that's when they were doing that assignment'. **Head of department, sixth form college**

As discussed in the section entitled '[Opinions on changes to the assessment arrangements in January](#)', many of the teaching staff we spoke to reported that, for GQs, SLT predominantly designed the overall approach to TAGs, and individual departments had a role considering the specific evidence to use. This was slightly different in the context of VTQs. Because of the specific requirements of each

qualification, individual departments had a much larger role in designing the process for determining TAGs. Some departments found this quite challenging.

It was ours as individual departments. So, for example with the sport and the performing arts, normally we'd have to have video evidence and everything else and we were going to struggle with that and also with data protection and students filming. So, each department had to come up with their own way of managing that, which was difficult, it was really difficult. **Senior leadership team member, academy**

The same SLT member explained that some AOs wanted to moderate the evidence but issued unrealistic deadlines, which required the centre to negotiate with the AOs to be able to deliver the course.

We had to do a mix of both as much as possible, which is, and this is where I've got another issue with the exam boards. Because [AO 1] and [AO 2] wanted to moderate externally, which I get, but they were pushing for so early to try and moderate that we wouldn't have been able to deliver, so we had to keep saying to the exam boards, you've got to give us time to deliver the content so we've got the evidence rather than push for the moderation so, so early. **Senior leadership team member, academy**

In our interview sample we spoke to two individuals delivering functional skills qualifications. Following the policy decision by DfE, the Ofqual [information for functional skills](#) qualifications detailed 3 ways students could gain a result: 1) assessments could take place at centres in line with public health guidance; 2) assessments could be taken remotely online and 3) where neither option was available and the students needed the grades to progress, TAGs could be awarded. Taking assessments was the preferred option rather than using TAGs to determine a result.

Both the teaching staff we spoke to wanted those of their students with sufficient evidence to receive TAGs, because the level of disruption their students were experiencing and their specific circumstances meant that accessing assessments would be difficult. They described going through similar evidence collection and evaluation processes as for other VTQs, and submitted TAGs for their students to their AOs. However, there were different outcomes for the two TAG submissions. For one, all TAGs were accepted and approved by the AO.

Yeah, because the learners who wanted the qualification [were] over the moon that they've got it. I mean there's some learners, because the Open University deadline is August/September time, they've now got their level two and they're able to do it. **Teacher and/or tutor, further education establishment**

However, for the other tutor none of the TAGs were accepted by the AO. The tutor explained that, as professionals, the department had looked carefully at who should be put forward for TAGs and the reasons for this. They were surprised when these

were rejected and to receive the advice from the AO that students should complete remote exams despite continuing difficulties the students faced in accessing them (e.g., digital poverty and childcare responsibilities).

So we submitted our list of learners to [AO] and they rejected each and every single one of them because all of the barriers that were in place in January when we started the process, some are still in place, but they don't care essentially. They've thrown them all back and we've appealed and they still have rejected it. [...] They just rejected all the learners straight off and said well, nobody needs to shield anymore, you can sit the exam remotely. **Head of department, training provider**

The same tutor continued that she felt this was unfair for the FSQ students compared to those completing general qualifications, as she felt they were being treated differently.

And it just, it's wholly unfair that these learners are, have been, disadvantaged so much compared to learners, you know, that are doing GCSEs or A-levels and again it comes back to that thing of why are they being put on a different playing field to these guys? When essentially you have to do a functional skill within an apprenticeship because it has to be an equivalence of a GCSE. So why aren't they being treated as an equivalent of a GCSE? **Head of department, training provider**

Contrasts between Awarding Organisations

Although Ofqual provided guidance on how to award TAGs, much of the practicalities surrounding this, for example the amount of evidence to collect, was left to the discretion of centres (under guidance from the exam boards and AOs). Centres often enter students for different subjects and qualifications, and therefore often use multiple AOs. This meant that some interviewees were able draw on their varied experiences and make comparisons.

The quality and quantity of assessment materials provided by the different AOs was also a topic of direct comparison. Several of the teaching staff we spoke to noted that exam boards in GQ took different approaches in their provisions to assist with assessment and grading. Some provisions were particularly helpful, such as being able to digitally generate assessment papers. Teaching staff also reported that some of the guidance documents to support grading were useful, whereas others were too large in volume and complexity for centres to usefully extract information from them.

[Exam board 1] are great in that they have [software], and you can say I want questions on this, this and this and it produced a paper for you. [exam board 2] have something similar, but they charge you for it. So then it's are we going to pay for [exam board 2] to generate this, or shall we just physically sift through papers and do that. **Head of centre, further education establishment**

A few of the teaching staff we spoke to also perceived differences in both the quality of communication and the process of collecting evidence for external quality assurance employed by the AOs.

Yeah, timely streamlined communication, simplification of the process. I know that our exam board had a more complicated process than others, because I've got friends teaching [exam board 1], for example, and I know that we had more hoops to jump through and more of a burden of evidence to provide, so in the end, yeah, we went nuts and we gave [exam board 2] something like 21 pieces of evidence and said, 'there you go then, there's your evidence, have fun', because I got cross! **Teacher and/or tutor, further education establishment**

A few teachers told us that some AOs continued to deliver external moderation after the January announcement, but others did not. This continued support was perceived to be helpful.

For the majority of the boards where their external moderation tried to continue, [...] that helped as well to cement 'yes, well yes the exam board have said our assessment is to national standards' [...]. I'm not happy with [named exam board] because they decided to completely step back. And so luckily all our [named exam board] deliverers are experienced, if we'd had new members of departments in that they would have struggled. **Senior leadership team member, academy**

Other considerations in the TAG process

As well as the evidence-collection that directly contributed towards TAGs, the teaching staff and students we spoke to often discussed a number of other issues related to fairness, the arrangements in place to account for reasonable adjustments and special consideration, and the challenges of assessing specific types of students, such as private candidates. The personal experiences of teaching staff and students are also important, in particular the range of pressures that teachers experienced, and the well-being of students.

Fairness and minimising bias

Issues of fairness were hugely important to the teaching staff that we spoke to. The importance of ensuring fair grades for all students was reflected on by almost all interviewees, with particular efforts described to ensure that students with protected characteristics were not disadvantaged during the process. Ofqual released [guidance to support teaching staff in making fair and unbiased judgements](#), which centres used as part of their training. It is worth noting that several processes were implemented in centres to ensure objectivity and fairness during the marking of the assessments and deciding the final TAGs, which were discussed earlier in the section entitled '[Evaluation of evidence and Internal Quality Assurance](#)'.

Information, training and awareness

We asked our interviewees whether they received information or training on ensuring fairness for students and minimising bias during the TAGs process. All of the teaching staff indicated that they did receive training on making objective judgements and minimising bias through a range of reported approaches.

Many staff reported attending training sessions that focused on maximising objectivity and minimising bias throughout the assessment and marking process. For some the training offered was extensive, and often embedded within wider internal training on the whole TAG process.

We most definitely had training on many, many points, specifically bias, and there were probably, I've lost count actually, but I think somewhere in between four or five, maybe even six different training meetings where the processes and procedures would be gone over, what was expected of us and any questions and answers, of which many of us had many. [...] And also the literature that was made available through attachments in Microsoft Teams, in messages and attachments to emails with very clear instruction from management that it was essential, that it was critical that you read this and understood it. And we and the

candidates also had to provide some online signatures to indicate that we had taken part in the training, or that we had viewed the videos of the procedures.

Teacher and/or tutor, tertiary college

For others, the sessions were relatively brief but were accompanied by ongoing informal discussions within the departmental team.

The training on the bias took place in a 15-minute meeting after school on a Monday just before one of our many moderation meetings **Teacher and/or tutor, selective**

Sometimes, centres provided teaching staff with training content they could review in their own time. One centre provided teaching staff with documents to read on bias (this was in place of face-to-face training). In another centre, video-recorded training sessions were made available to staff.

Yeah, so that was one of the training sessions that I did with staff. I've taken the lead in digesting all the big documents and stuff and then we'd do and like you've done with me, we'd record it so that if anyone wanted to go back to it or anything they could do. I did a distilled version for them. **Head of centre, further education establishment**

For many teaching staff, reducing bias was already integral to their usual processes. Consequently, training on unconscious bias was a helpful reminder rather than novel information.

We're also part of the Institute of Physics. There's an improving gender balance programme that they're running at the moment. It's really about getting more girls to study physics, but equally get more boys to study things like arts and English and stuff like that. And we'd already had a lot of training earlier in the year on unconscious bias, because that was picked up as one of the things that we needed to work on. **Head of department, comprehensive**

One interviewee suggested that additional examples of bias would have been useful to support training and understanding.

It was really good what we received and we really appreciated that, but I think what people want are examples and so it would have perhaps been really good to have some small case studies. **Deputy head of centre, sixth form college**

Reasonable adjustments, special consideration, and access arrangements

The terms 'reasonable adjustments', 'special consideration' and 'access arrangements' are used to describe changes made to an assessment to make it more accessible for a student. Access arrangements is a broader term generally

used within schools and colleges that describe any adjustment that allows a candidate to access the assessment without changing the demands of the assessment. They are agreed before GQ assessments through applications to JCQ. We use this term below to be consistent with the terminology used by teaching staff.

Reasonable adjustments refer to changes made to the assessment and how it is delivered for a candidate with a disability. A reasonable adjustment may be unique to a candidate and may not be included in the list of access arrangements. Special considerations are usually post-assessment adjustments made to marks or grades for reasons, such as illness, injury, bereavement or other indisposition at the time of the assessment.

The published guidance stated that where possible, reasonable adjustments and access arrangements should have been in place when evidence was generated. Where this was not possible, centres should take this into account when making their TAG judgements.

Access arrangements

Our interviews suggested that for most evidence that contributed towards the TAGs, students were granted the same access arrangements that they would have received if they had sat normal assessments. Most of the teaching staff we spoke to described how SENCOs played an important role in supporting teachers in setting up assessments for SEND students.

So, they were all taken into account, so the students that are allowed 25% extra time, they were given that for every assessment, the same as students that needed larger font size and things like that, so we made sure it was very clear and our SENCO sent out on a very regular basis a reminder, student A needs this, student B needs this, just to make sure that they are getting it. And so, yeah all the reasonable adjustments were fully taken into account. **Senior leadership team member, academy**

Because of the logistics of assessing students outside of a normal exam timetable, and the limitations on resources, some centres did struggle to provide all the required access arrangements.

And then of course because of doing the assessments, we couldn't do them in exam halls so it wasn't a normal situation. We could have had for instance as a school, English TAG assessments going on at the same time as science and one of the option subjects, meaning that suddenly you've got to find qualified scribes or teaching assistants or invigilators to supervise all of these students in different rooms doing different exams, who all need their exam access arrangements at the same time. It was really challenging. **Head of department, academy**

When this was the case, the access arrangements were often accounted for in the grading by teaching staff.

Because they were doing assessments they applied, so they had extra time, they had laptops, all the usual assessment because we were doing it. On a couple of occasions if it was classroom-based assessments they may not, in which case it was taken into account in the grading. The overwhelming majority we managed to put the access arrangements in place, because it was all assessment based.

Teacher and/or tutor, private candidate exam centre

SEN departments also took a part in determining how to evaluate evidence for assessments for which the appropriate arrangements were not put in place.

And the TAG data, after it had been seen by heads of department, was then put through the SEN department. [...] What they [students] won't have necessarily had extra time with is any classroom-based assessment that preceded January. Because obviously you're working to a, you've got a 35-minute lesson and a 30-minute test [...] But I think, I have got a lot of confidence in our SEN department, our head of SEN has been the head of SEN for about 12 years, so I think it was in a set of capable hands and I have no doubt that reasonable adjustments were made. **Teacher and/or tutor, selective**

Sometimes the access arrangements had to be made in a way that was perhaps not ideal due to logistical difficulties.

Students who needed extra time, because we couldn't get them in anything other than lessons we either had to reduce the content of the exam to make it an exam that actually could be sat in 45 minutes meaning that the extra time could be absorbed into the lesson, meaning you have to cut content somewhere and do that content somewhere else or you have to ask students to do their extra time significantly after they'd done the assessment. Even if that just meant that they had to come back at lunchtime or they had to do it in the next lesson that's still not how they would have sat exams in the real thing. **Head of department, academy**

The students we spoke to who had access arrangements reported that their needs were fully accommodated into the assessment process.

I'm partially blind in one of my eyes, so my reading time tends to be a bit slower. [...] So, I was allowed extra time off the exam board for that. So, with the exams and the mocks, I know that we got in the large paper. So, my teachers all obviously still granted me that. [...] For the May mocks, I was still, me and quite a few other people still sat in a different room. So, it was all catered for, nothing of that changed. If some people needed the computer or a laptop to write, you still got a computer or a laptop. If you needed a reader or a scribe, all that was still done. [...] And people who normally stayed at home, did it at home. **year 13 student, secondary**

Special consideration

Special considerations are usually dealt with by the AOs, however, in summer 2021 [JCQ guidance](#) suggested that centres should take individual student circumstances such as illness into account by selecting alternative evidence for the student. This was usually one of the many considerations described in the interviews around selection and weighting of evidence for determining TAGs.

We had special considerations for students who were off or who had COVID or who had bereavements and we had to take those into account. So, if there was a student who for example performed really poorly in an assessment [...] we would disregard that and find a different piece of evidence instead, one that is more reflective of how they actually normally achieve. So, we did take that into consideration obviously because you have to, don't you? **Teacher and/or tutor, sixth form college**

Because most centres had multiple pieces of evidence, they were often able to exclude from consideration those where the student had been disadvantaged. We also earlier heard how centres sometimes arranged extra assessments to gather additional evidence for those students who had missed one or more of the scheduled assessments.

Many centres requested students and parents or carers to submit a form for extenuating circumstances that was then reviewed by individual teachers or SLT (including the pastoral team where relevant).

After this set of exams, they sent an email out and said, 'let us know about any special consideration that you have had, [...] has there been anything that has affected you recently or more generally across the period?' And then one of the Deputy Heads in the school, who wasn't involved in any real department looking at things, they independently then looked at these. Any that they felt were [...] worthy were on the spreadsheet for us and it basically just said, 'give consideration'. **Head of department, independent**

Because of lockdown and the pandemic, many centres said they received a substantial number of requests for consideration of extenuating circumstances and this did create an administrative burden.

Students are normally required to request consideration at the time of the assessment. However, the use of evidence gathered before normal assessments were cancelled meant that in some cases there was a perceived need to apply special considerations retrospectively, to account for students who had not made a request at the time.

People are supposed to have requested this at the time, [...] this person needs it applied back to February 2020, like it says here you need to apply for it at the time of doing the assessment and I think the rationale was they didn't know it was

going to be part of their assessment at the time they sat it, so they wouldn't have applied for it at the time. [...] It did mean that we had a lot of stuff going back to say this person needs to be covered. **Teacher and/or tutor, selective**

Although not in alignment with the JCQ guidance, some centres appeared to deal with special considerations through some kind of mark adjustment.

But the decision couldn't be made centrally by the school because their argument quite rightly was it might not have affected them in some cases, in some assessments. And then of course the 5%, up to 5% was actually for every assessment overall. Like it wasn't oh this kid got, out of 130 marks they got 100 marks, we should give them 5% extra, that's not the way it works. You have to make that decision for each individual assessment. Were they affected on that day. **Head of department, academy**

Issues relating to specific types of student

In this section we explore the TAG process for students with little or no evidence, students with evidence of inconsistent performance, and students for whom the centre had little or limited data on prior performance, particularly private candidates.

Some of the teaching staff we spoke to described instances where students with more limited attendance had not produced the evidence that typically contributed towards TAGs in their centre. These instances were felt to be particularly challenging to teachers making the final grading decisions. Where this did happen, teachers explained that they evaluated the limited evidence they did have and applied their professional judgement.

That was more tricky because you want every student to have something, but there were some that because their attendance has just been so poor, and it wasn't just the pandemic, it was a general thing, they wouldn't have got a grade [...] so we used our professional judgement at that point. **Head of department, comprehensive**

The issue of fairness was brought up by many of the teaching staff we spoke to. Some thought it was fairest to award a TAG based only on the work that had been produced. Conversely, another wondered if that approach was fair to those students who had produced all of the work required.

We had a couple of students who became non-attenders so didn't sit any of their minis [exams focussed on a particular topic] but [...] they've still been awarded a grade because we were able to use their mock and NEA [non-exam assessment] or assessment that they'd done. So we felt that that was fair [...] in that circumstance. **Deputy head of centre, academy**

I had students who hadn't given in all the pieces of work and so had a lack of evidence, but were still coming out with these quite good grades. And that was a

really hard thing to either think about or get your head around and then think about 'OK, is this fair to the other students?'. **Teacher and/or tutor, sixth form college**

Some of the teaching staff we spoke to also identified students with inconsistent performance as particularly difficult to determine a TAG for. The approach to dealing with this varied. In one case, it was a labour-intensive effort of going through each piece of evidence again, this time with colleagues, to come to an agreement on an overall grade based on professional judgement. Whilst in another case, the approach was to allow the students to re-sit specific assessments to gain better consistency and improve their grade.

In every single case where [...] there was inconsistency, where there was, was a difficult decision for that subject, we went back, we literally had the grade descriptor printed in front of us and we cited in the student's portfolio of evidence where they had, where there was evidence of those students meeting that standard. **Deputy head of centre, academy**

But for two [students] it [the evidence] was much more inconsistent [...] they were two of the students who re-sat [...] the kind of candidates that we thought should have another kind of go at it really. **Teacher and/or tutor, independent**

Similarly, students with no ongoing performance data were mentioned. This lack of evidence was generally because the student had transferred from another school and no information about them had been received.

There were two students who for their year 10 and then their mock grades had no information. [...] It was clear that one of the students had transferred to us from another school in October, but their teacher assessed grade was a U and there was no budging because actually there was no work, there was no substance to be able to say that there was anything else ... that I feel was an oversight actually... why have we got this information missing for this child, I was disappointed it was only at this stage it was identified that actually we probably could have asked the previous school have they got anything to back it up? Could they have a grade? **Head of centre, comprehensive**

Private candidates

Private candidates were required to register with a centre for their TAGs to be determined. Because they had not attended the centre previously, unless they were resitting, they lacked a "track record" with the centre. One interviewee noted that they had two private candidates at their centre but that the process for assessing them was largely similar, bar one minor alteration, to their other candidates.

We pulled him in, and he only did three GCSEs with us. And we got him in during our exam timetable and he sat the papers with everyone else. [...] Another

candidate, there was one, I think it was English literature where he'd studied a different text, so we had to devise a different paper for him, so that we could get a data point that was fair for him, which we did. [...] But otherwise they [...] sat papers alongside our students, which made it dead easy. **Senior leadership team member, independent**

One interviewee came from a centre that works with private candidates as part of their normal practice. They described some of the considerations that they made when working with private candidates to ensure that they were assessed on content they had been taught.

We insisted that all students did assessments through us, even if we had evidence from tutors, we made sure that we had at least one considerable piece of evidence that was gathered at the centre, usually more than one. [...] We contacted all students and we asked them what they had covered, whether they'd had any disruption, so that we could establish that we weren't testing on things that people hadn't done. [...] We told them that they were going to do a paper at the centre, we put that in place and that was non-negotiable, because that was the one they'd all done the same topic on, so we told them straightaway they were doing that. The second paper they did at the centre, although we did a complete paper, we allowed them to choose the topic because they'd all covered different ones. **Teacher and/or tutor, private candidate exam centre**

It was possible for private candidate to have seen different past papers for practice or mocks, which could make designing common assessments difficult.

There was one occasion I had to rewrite all the assessments the night before, because, having just got the tutor evidence in, I realised that the tutor evidence included the paper that they were supposed to be sitting the next day. **Teacher and/or tutor, private candidate exam centre**

Pressure experienced by teaching staff

There were several sources of internal and external pressure described by teachers/tutors throughout the process. Almost all of the teaching staff we interviewed described the substantial burden arising from the need to create, administer and mark their own assessments in addition to their already busy teaching workload. Furthermore, many felt a tension between their personal responsibility for awarding fair grades that supported progression, and pressures from SLT to ensure that grade profiles were not substantially different to those of previous year groups. Although not as direct, a final pressure discussed by interviewees was felt from parents/guardians and media sources, however these were often managed robustly by centres.

The burden of determining TAGs

As we have seen, in deciding and collecting the evidence they would use to support TAGs, most teaching staff we interviewed were involved in the creation and marking of assessments, internal standardisation and moderation processes, and determining TAGs, all of which needed to be delivered within what was perceived to be a relatively short timeframe. This was in addition to their normal administration and teaching responsibilities for other year groups. All teaching staff reported a strain on their time and resources, and experienced significant stress.

We did have quite a burden on us, that's for sure. I think essentially what it feels to me like is that the government have said 'right, we're not going to do exams. [...] You guys come up with a Teacher Assessed Grade' [...]. And then left it entirely to us on top of our already pretty heavy workload anyway. So, I think that the worst thing is that we've had to basically make our own exam series, coordinate it all, timetable it all. [...] And then do all the normal stuff that we do. Teaching, revision, coaching kids and making sure that we give kids who need access arrangements, making sure that they get that as well. [...] Not enough time, not enough information and virtually no support. **Head of department, university technical college**

We had the reality of a business studies teacher with 150 sets of exam papers times three and she's a lone teacher in our school because of our small context. [...]. I was arguing I needed more time to test the data, to quality assure the data, to get all of our teachers out working with other schools. So the pressure was quite immense. **Head of department, academy**

Many teaching staff felt let down by exam boards as they believed they had taken on this additional and complex 'assessor' role with little support and no financial compensation.

Now I'm sat in a room marking an exam which is actually going to produce a grade. Well exam boards pay people quite a lot of money to do that and the teachers have just done that for nothing and carried a huge, I actually think, a very significant pressure and we're yet to see how that ratchets when parents appeal. **Deputy head of centre, selective**

Nearly all interviewees involved in GQs expressed disappointment about the assessment materials that exam boards had made available in support of the TAG process. Teaching staff expected exam boards to provide some new, unseen assessment materials with which they could make robust grading decisions and felt that this had not been delivered. Instead, teachers felt that exam materials were published relatively late with few unseen questions that could be utilised by teachers/tutors.

There is I suppose the feeling, the negativity seems to be stemming towards the exam boards if I'm going to be honest, for lack of clarity, lack of guidance, and I suppose a feeling that the staff were doing the exam boards' job. That's how they felt about it. Because all this promised guidance and past example questions and stuff just appeared to be all the past papers that the staff already had. In some instances, the exam board had chopped them up so they could say all questions on this topic are here, all questions on this topic are here. **Head of centre, further education establishment**

What was slightly disappointing was having waited for the exam boards to release material [...] I think myself and a large number of other colleagues were utterly dismayed to see that the released material was only past paper questions stretching back a decade or so and it was exactly what we already had, so there was nothing new and it was a huge volume. I don't remember the exact number, but I think I counted something like 156 separate documents just for biology alone and I thought if we've got to trawl through all of that now at this late stage to try and construct assessment material someone's having a laugh. **Teacher and/or tutor, tertiary college**

As discussed previously in the section entitled '[The design of the post-guidance assessments](#)', because the exam board materials were not perceived to be as useful as hoped for, departments instead chose to create their own assessment materials. This was felt by many to be a challenging task. Some acknowledged the specialist knowledge and skills required, others mentioned the time pressures to produce assessments leading to difficulties.

It was the lack of unseen papers that really made it difficult, because of the amount of assessments we needed to do and because we wrote them, mistakes happened. There's times I had to apologise to parents and students because I did make mistakes, just for the sheer amount of workload and fortunately, hopefully it was all caught at an early stage and it hasn't affected their final grades because I was able to take account of it, apologise, redo stuff if need be and all worked, but where people are placed under that amount of workload is going to make it, they were luckily minor, but things could have been much worse all the way through.

Teacher and/or tutor, private candidate exam centre

A few of the teaching staff we spoke to took some positives from the task of writing their own assessments. They felt it was a useful personal development opportunity to learn more about the process of writing assessments, something that they would not have experienced otherwise.

What was good about this process is for my own personal development is that I really looked at how exam papers are developed and how they create them. And it's like a two-year process that we're trying to jam into a couple of months whilst also teaching. Which, made things pretty difficult. **Head of department, university technical college**

Further reflection on expertise that staff gained through carrying out the TAG process are detailed within the section entitled '[Expertise gained from the process](#)'.

Overall, all teaching staff reported that the workload from the TAG process was significant, and had to be carried out alongside other responsibilities. This was true from class teachers to SLT.

Our Easter holiday, we probably managed to have one week where we didn't do work, maybe and then a week where we were working and didn't feel that we could call upon other people because we felt, looking at them, they needed a rest, so the burden fell on us. And we were so, it was right that we were heavily involved, but we, every week we lived and breathed it with all of the teaching staff and with the students and with the parents and even the half term, we've also been working in the half term. **Head of centre, comprehensive**

As a result of all of these issues and the workload that teaching staff had faced, many of the teachers were aggrieved that centres had still been charged substantial fees by exam boards.

I think I was very disappointed with the exam boards, I was very disappointed they didn't provide any new material and they didn't provide any 2021 material. They're still asking for the full, pretty much the full fee, for having had the teachers write the exams. **Head of department, academy**

External quality assurance

External quality assurance (EQA) for GQs was carried out by exam boards. They checked all centre policy documents. For sampled centres, senior examiners acted as assessors who looked at the evidence and judgements for a random subset of students to confirm that the TAGs were appropriate. EQA approaches for VTQs varied, depending on the qualification design. Some followed a similar evidence-sampling approach to GQ, and others adapted existing quality assurance or verification processes that are used for internally-assessed coursework in normal years.

Distinct from the process of judging TAGs, there were some reported difficulties with the process of submitting evidence for EQA. Because this submission was often required within tight timescales, this caused staff stress.

We were told you'd find out what your sample was on the Monday. They emailed our exams officer at 20 to nine at night on the Monday night with a message that said Ofqual have been very clear that you have to adhere to the 48 hour window for evidence, so all evidence needs to be in by 10.00am on Wednesday. It was one of those like we're putting the blame on somebody else, because this isn't our rule. Then they went oh because there was an outcry and Twitter said you can have 48 hours from now. And then the following day they went OK, we realise it probably wasn't appropriate to say 48 hours started at 10.00pm at night,

you've got 'til Thursday I think they gave us in the end. **Head of centre, further education establishment**

It also seemed that several teachers found the process lacked clarity and put them in a difficult position whereby they had to chase students for pieces of evidence.

I do know the pressure that it put people under to produce evidence in 48 hours. But it was English for GCSE [...] I think it was further maths for A-level. And having that evidence to hand was, again it's not something that we're used to doing. [...] We had given the students the papers back with all the feedback and guidance and marking so that they could use them for the next assessment. And I'm not sure that it was made sufficiently clear early on that all this stuff needed to be kept. So there were certainly panicky phone calls going out to students to say can you get this piece of work to us because you are one of the ones that have been picked, and we don't have the evidence. **Teacher and/or tutor, selective**

Many of the teaching staff we spoke to who had submitted samples of students' work were frustrated by what they saw to be a lack of guidance about what AOs wanted in the sample. This was particularly felt where centres had used a range of different pieces of work as evidence.

The other thing to say about the sampling part of that quality assurance process was when they asked for the sample they didn't actually tell us what they wanted. So, do you want all the student work over the whole two years or one year? Whatever it is, do you want just April and May? Do you want the question papers? And, you know, so they didn't [tell us], so we just went, 'give them everything', because then at least they've got what they need if they need it. But I don't know if that was OTT and not needed, I don't know, so that was really unhelpful. **Deputy head of centre, sixth form college**

Where there were many pieces of evidence, particularly paper-based evidence, that contributed towards the TAG, teachers found the process of uploading the contents of the sample particularly time-consuming. Although EQA for GQs was carried out by one exam board per centre, many centres also had TAGs for VTQs with various AOs. One teacher reported that, because AOs had different systems for uploading the samples, the centre had to devise separate processes for each AO that quality assured them. Overall, for many centres we spoke to, EQA was perceived to be a fairly burdensome process.

In the sample [EQA] process we had one way of working with [AO 1, which] we dealt with first, and then [AO 2] had another way of working. And all the things that we'd done with [AO 1] wouldn't quite work in the same way with [AO 2]. So it just took hours. But of course, we had to upload everything digitally. [...] GCSE resit students: because you also need to learn to write, and you need to practice writing as a student because these exams are written (if we were to have exams). So, most of their work was actually handwritten. Be that on exam papers or in

books or whatever, or on file paper, so every single piece had to be scanned and then downloaded onto the portal. In music technology, that was particularly challenging because of the file size, as I'm sure you can imagine, so yeah, that was hard. **Deputy head of centre, sixth form college**

The EQA process in some qualifications was sometimes perceived to be duplicating earlier external moderation of internally-assessed coursework in VTQs. External quality assurance therefore felt an unnecessary use of time and effort to them.

Every subject, every level had to have an external moderation before the TAGs were inputted. And centres then got called for the external quality, and we're thinking, 'well, we've already been moderated why are we getting called for the 48 hours, got to get this evidence in, when we've already been moderated?'. So that caused a lot of anguish I know, we weren't called for any extra evidence, but I know of centres especially within our LA that were called for extra and it just caused so much stress because we'd already had a full moderation process. **Senior leadership team member, academy**

Personal responsibility

In almost all interviews it was clear that teaching staff felt a personal responsibility for ensuring that students were awarded grades that were fair but also that allowed them to progress.

I was exhausted. I was broken when I finished school. It was a massive responsibility to carry and it wasn't just an individual student's grade, because every student's grade in every subject matters. [...] There's been a lot of tears shed about this whole process [...]. It was really hard and you were working with human beings at the forefront of decisions that are going to make a big deal on a lot of young people's lives. **Deputy head of centre, academy**

This could be even more difficult for qualifications taught by a single teacher, as there was no sharing of responsibility within a department.

We had spent this whole week putting them together, and I, as a single course manager did that all myself. So all the responsibility for those grades are on me. The buck doesn't get passed anywhere else. It stops with me. So I took that very, very seriously. Which is obviously, very stressful. **Teacher and/or tutor, sixth form college**

The high stakes of the assessments, particularly for students taking A levels who were hoping to progress to university was also considered a source of pressure.

I think there were one or two students in the whole year whose outcomes, their UCAS choices are very heavily dependent upon these TAG grades, so I cannot possibly categorically say there was no pressure at all. I think there's a, it's not a direct pressure, it's quite subtle pressure and it might not be brought to bear by, it

certainly wasn't brought to bear by any parents that I'm aware of, I'm fairly confident in saying that it was never levered by students either, but I think there's this almost invisible pressure on staff, because if the knowledge is in your head that you know that a particular candidate needs B, B, B for their first choice, you have to think very carefully about, is that influencing my decision **Teacher and/or tutor, tertiary college**

But the opposite could also be true, that staff were aware that over-generous TAGs could be detrimental to students in the long run.

It's the A levels which is the bigger problem where if there's a mismatch between schools the wrong people are getting to university. [...] The ones who do get in on inflated grades won't then cope with the course and will drop out and have wasted a course place for somebody who could have got it and that's the real worry. **Teacher and/or tutor, selective**

Pressure from SLT

We saw earlier how SLT quality assured the TAGs and sometimes requested teaching staff to review or adjust them. The staff we spoke to often felt enormous pressure on their professional judgement from SLT.

SLT were crunching the numbers [...] the pressure became more, just greater. They would just come back and say 'we need some more grade 5s' or 'we need more grade 6s' or 'we need more grade 7s' or whatever. And because I'd re-marked them really thoroughly, I just knew that I didn't have any more. If I was to do that it would be, I'd have to have added classwork or added homework. Oh yeah, I'd have had to have added other things. And I felt that if I was adding classwork for one I'd have to have a look at it for all students. [...] If the evidence wasn't there the evidence wasn't there and I wasn't prepared to lose my integrity to make it up and say it was. **Teacher and/or tutor, comprehensive**

There was, I wouldn't say, the word wasn't pressure, but it was very heavily stressed to us that we needed to be able to justify the grades that we had given. And I mean, very heavily stressed, that we needed to be able to do that in case we were audited or our course was audited. **Teacher and/or tutor, sixth form college**

Teachers also felt pressure from their Head of Department, although this was often pressure from SLT passed on through the head. Limited time led to pressures on class teachers.

I think one of the other things that I felt, I felt pressurised just to agree with the Head of Department [...] Like in terms of time constraints, in terms of, 'you're kicking up a fuss', in terms of, 'let's just get on with it'. **Teacher and/or tutor, selective**

External pressures from parents, students and the media

The majority of interviewees said that they did not experience direct or indirect pressure from parents or students during the TAG process. In most interviews teaching staff reported receiving support from their centre to protect them from pressure. Centres generally provided teaching staff with robust standard responses to address any external contact.

They came to me and they basically got a standard reply, which said if you continue to behave in this way you're likely to disadvantage your student's result, thank you very much for your input kind of thing...everything's anonymised and your son's grades will be determined by how he's performed. We did get more requests than we usually get for actual adjustment, but that may be because there's been more difficult circumstances around in families. **Deputy head of centre, selective**

Others referred all contact to a designated senior member of staff.

Anything we got we were told, send it straight to them, don't engage with it so that there's no dialogue, and then after all the grading and stuff was done, again, it was if anybody communicates anything with you, they were told that it needed to go through a central email address and that kind of thing so that whatever message went out was consistent and there was no personal [communication]. **Head of department, sixth form college**

Pressure from parents was to be treated as malpractice, and many centres were clear with their staff or the parents themselves that this was the case.

Our school was very, very clear to parents very early on that any kind of pressure or querying about grades or anything like that would be considered malpractice and therefore the school would sort of do something. They never said what but that they would do something about that. And I think there was only one instance where a parent tried to provide some sort of evidence base to the teachers to show that with private tutoring they were achieving X, Y and Z during their tutor sessions and stuff. To be honest we just shrugged it off and we just ignored it. As it transpired we had a similar quality of work from that student so it didn't mean anything to us but I do think there was increased pressure perhaps. **Head of department, academy**

In some centres an effort was made to promote TAGs externally as 'centre' assessed rather than 'teacher' assessed grades. This was to shift the responsibility away from the teachers onto the centre as a whole.

Our head was very keen on calling them centre assessed grades. And all of his communication to parents, to us, everything because he was trying to emphasise the point that this is not down to one individual teacher. It's the school that's doing it and so on. And so from the start he was very certain in all his communication to parents that any questions go to him only. We were told to forward anything on.

Teacher and/or tutor, academy

However, a small number of teaching staff did report pressure from parents to award specific grades. These teachers described requests and questions relating to students ahead of results day, with parents expressing concerns that their child's mental health or progression could be adversely affected by disappointing grades.

From parents, we actually had a parents evening right at the end of the spring term, so in March. So basically parents knew that it was Teacher Assessed Grades and two parents did say what does my daughter have to do to get an A*. Or is she going to get an A* or could she get an A? **Teacher and/or tutor, independent**

And then you also get parents who say 'well his mental health is really fragile, if he doesn't get this grade what might happen?' So there's all those things going on as well and then we're just not, it isn't what we do, it's not what we're used to.

Teacher and/or tutor, selective

The reporting of TAGs in the media was also a pressure described by one interviewee. There was a view that public confidence in teacher professional judgement was largely influenced by information reported by the media.

So I have a PGCE in post compulsory education, I have six years of experience. I am a subject specialist in my subject which is quite rare for classics. And there were times when it felt like we weren't being trusted as professionals and that came from the media. **Teacher and/or tutor, sixth form college**

Student well-being

Many of the teaching staff and students we spoke to felt that many aspects of the assessment process were stressful to students. The students we spoke to clearly distinguished three different stages in the TAG awarding process that affected their well-being. These stages were: (1) the initial period of uncertainty between January and the issuing of Ofqual guidance in March; (2) the assessment period itself; and (3) the period after the final grades were submitted to the exam boards.

The period of uncertainty

After the initial announcement that exams were cancelled, there was a period when it was unclear to students how they were going to be assessed. Schools and colleges

felt they had to wait until the final guidance had been published before they could formally announce their plans for TAGs to parents and students.

Centres were often contacted by students and parents who were keen to get more information. Centres often felt they were in a difficult position, largely because they did not have any definitive information themselves.

I think people think that we find out before anybody else. Well, we don't, we find out at the same time as the public and so it's really hard. So you're constantly saying, 'well, we'll just tell you more as and when', and we're trying to drip feed information, but also you don't want to give out poor information, because then that can be worse actually. **Deputy head of centre, sixth form college**

The teaching staff we spoke to reported that because centres felt unable to give definitive answers to queries from students, this had a negative impact on many students' mental health.

I think the hardest thing was the students, they had so many questions. [...] They felt incredibly stressed by it. And the amount of mental health cases and stuff that needed dealing with rocketed. **Head of department, sixth form college**

This was clear in the interviews with students, who described anxiety and uncertainty about how their grades would be decided and how it could affect their lives.

So, the exams are cancelled, but we didn't know what was happening. So, it was just that uncertainty in the moment in January and February. I think it was just a blank header really and we were told to carry on working, didn't really know what we were working towards anymore, because there's no exam at the end of it. **year 13 student, selective**

And then there was also a lot of anxiety, because we didn't know immediately how we were going to be assessed. So, I was quite nervous to find that out and how that was going to affect me, and whether I was still going to get into the university that I wanted. **year 13 student, college**

Overall, centres tried to be open and to reassure parents and students as best they could while awaiting further clarity. The intention was to encourage students to continue learning and making effort, since they would be assessed in some way.

We were quite open with our parents and we carried on as if it was a normal year. We told them that they would be doing assessments, they would have assessment opportunities, we weren't sure what they were going to actually be at that point. But they needed to keep working and that it wasn't the end, we would figure something out basically. **Deputy head of centre, academy**

Students sometimes reported that during this period they felt unmotivated. However, as more details about the arrangements for assessments became available, they regained their enthusiasm to continue engaging with learning. Students reported that

they particularly focused on those subjects for which TAG evidence would be in the form of exam-like assessments.

We just kind of kept doing what we were doing before. I had, I think I stopped doing the revision all day every day until about late February and then we, I knew that obviously they were going to be exams when we went back. And we got a timetable, so I started focusing on those subjects and then doing bits of the other ones, but mostly prioritising those. **year 11 student, selective**

Students also reported paying more attention to the quality of their ongoing work, as they believed it could contribute to their final TAGs.

Although the final guidance to centres was made available in March, in some centres the uncertainty about how students would be assessed continued beyond this date. This is because it sometimes took time for centres to finalise the assessment process and disseminate that to parents and students.

I think we were notified of exams getting cancelled in January and having to wait until February/March time before we get any announcement. And then that announcement to be honest wasn't really, there was nothing substantive in that, and it was kind of: "Wait a further few months for teachers to decide what was going to happen". I don't think that was the right way to do it. **year 13 student, selective**

However, there were also students that felt relatively well informed by their centres as to the arrangements for assessments. These students also continued to feel motivated to continue with their studies, with the understanding that students would be assessed and graded in one form or another.

I think they were very good at keeping us informed of what was happening. We had like emails to keep us updated and we were told when we would find out exactly what was happening. And then we were told and that's fine. Although there was some issues with communication when it first came out, because the teachers had been told different stuff to us. So, I don't think it was communicated as well to the teachers as it was to us. **year 13 student, college**

The assessment period

During the assessment period, many students felt that the prolonged revision and busy assessment timetables affected their well-being. Students sometimes reported struggling to maintain focus and effort over long assessment periods, and suffering from the stress of the sustained pressure.

And it was this whole process was stressful, daunting, and long as well, because there were so many exams packed in, it was over six weeks, I think. And over those six weeks, it was the anticipation to these exams. **year 11 student, independent**

They constantly gave us like proper mocks, like I'm talking two-hour affairs. And to everyone who was having to isolate left, right and centre, it was quite, it was a big ball of anxiety, it was quite stressful. And even when us, a lot of the students, were like putting stuff across to the teachers like, it's stressful. **year 13 student, secondary**

The assessments used to support TAGs were felt by most students to be high-stakes and they were anxious about their performance.

My exams were something that were very important to me and I worked for them. I might have not done as well as I wanted to obviously, given everything, but like the effort was hundred percent there. [...] I don't know how to explain it, because you knew with our school that they practically were actual exams, but the feeling was also there that they're not. So, it was like: "Ah yeah, it's just another mock." You also had that voice in your head telling you: "No, with our school doing this, this is your actual exams, it's practically the same." So, I had that motivation there because I knew that I had to, but there was also a big part of my brain going: "This has been going on way too long, you're burnt out. You've had mocks here, there and there, you don't know what's going on." And it, yeah it was kind of a conflicting point of view. **year 13 student, secondary**

It made me almost more anxious, because a lot of the assessments throughout. Obviously, I knew what grades I'd been getting, and I knew they weren't the grades that I wanted, a lot of them. It puts lots of pressure on those final assessments to make sure I was performing my absolute best, because I knew if I didn't, I didn't have the grades I needed. **year 13 student, college**

This experience was not shared by all of the students we spoke to. Some students found the process of working towards and being awarded TAGs less stressful compared with sitting normal exams, largely because there were more opportunities to show what they could do.

I felt like a lot of people within my school, [were] particularly worried [about] them: the anticipation [that] they were the actual GCSEs. But I'd say that I personally found it easier knowing that they weren't the final GCSE grade [...] I thought it was better because they did emphasise that if they went badly there was other work to back you up on. **year 11 student, independent**

The fact that I was being tested regularly [...] – because I needed an A at the end of the year, so not everything for me had to be an A. I could get a B here and there, and it wouldn't be stressful for me because I knew that I could work on my next exam to get higher. **year 12 student, college**

The teachers that we spoke to also often considered students' well-being and recognised the stress that the amount of assessment could cause.

So even if they don't have an exam with you they've had an exam all morning, they're feeling washed out [...] So, I think it was really tough on the kids. And I think that the process was important for them. I think they did go through an exam process and a revision process and I think that it just got more real for them the further into the process we went. **Teacher and/or tutor, comprehensive**

The stress of normal end of course assessments was also recognised, and because of the way they were run, assessments to support TAGs were sometimes considered less stressful.

[TAGs were] fairer than the exams. Exam fear is massive, [...] awful experience going through exams and every year students get really, really upset by them and find them really stressful. So, this experience, I think genuinely, let them show what they could do without the pressure on so much. [...] some of them still got stressed out by [TAG assessments], but it was much less stressful than the 'walk into the exam room and do it under exam timed conditions'. **Teacher and/or tutor, further education establishment**

However, views and experiences were mixed on this point. One interviewee suggested that the intention in setting numerous tests was that they would be perceived as low stakes by students, who would therefore not feel as much pressure. They did not think that this strategy had worked.

The belief was that if we assess every week or every fortnight, they would become a lot of low-stakes tests as opposed to a high-stakes examination. Don't think that worked. I think they became a lot of high-stakes assessments for a lot of the pupils because they felt 'oh, I can't afford to mess this up'. **Head of department, independent**

A few teachers mentioned pastoral systems that were in place to support students throughout the assessment process. However, another centre suggested that accessing support was more difficult for students during remote learning.

We had a few students who really struggled with the whole process, and we had a few students who were generally OK. I think our students in the school are well supported, so we had very good systems with mental health and it supported the students. So I think the majority of them were fine. **Teacher and/or tutor, comprehensive**

I think a lot of them [were stressed], and especially because it was all on Zoom and they couldn't actually go to a teacher in the spring term and say you know, 'can you help me?'. You couldn't sit down with them. Which is an important part of the sixth form. **Teacher and/or tutor, independent**

The guidance was clear that teachers should not inform students of the final TAG that the centre was submitting to the exam board. Because of this, some centres chose not to disclose ongoing performance levels and sometimes the grades for

evidence that contributed towards the TAGs, so that students could not start calculating or predicting possible TAGs. Some students mentioned that this created a great deal of uncertainty for them.

It was the end of October when I got my last report, and then that was my last set of predictions and my last set of current attainment. [...] We should have got our spring report, and I think that might have been March they were due. But because the school weren't giving any indication, in anything, of what we were achieving at that time in January [...] I felt that it was unfair. Because then I was thinking 'oh god, we're going into assessments and exams and I don't actually know what I'm achieving at the moment, so I don't know how much work I need to put in'. **year**

11 student, selective

Fears for the future while awaiting results

We undertook these interviews during the period between the TAGs being submitted by centres to AOs and students receiving their grades. Many of the students that we spoke to were therefore anticipating what grades they might receive, and how this could impact their futures, at the time of the interviews.

There was just so many different emotions going on. People worrying about universities, [...] not knowing what grades you're going to get. One of my friends was going through trying to find universities because none of their offers have been accepted. [...] And it was just a massive mess of like all these emotions and no one knew what to do with them all. **year 13 student, college**

I keep looking over my notes just to keep it kind of fresh in my brain, just in case I do have to re-sits. Because everyone's in a position where they've got no idea how they've done, because it was that stressful and that rushed, and everything. [...] So, everyone's kind of more anxious for results day than we normally are. Because, obviously, it's something that's very stressful as it is, but I definitely say tensions are higher this year. Because it's that ambiguous and unknown as to what's going on, and what's happening, and how things have been treated, that no one knows. **year 13 student, secondary**

A few were worried that the EQA process could decrease their grades.

To be honest, I'm actually scared because I have high expectations, but I'm scared I'm not going to meet those expectations. [...] I'm just scared that we will get downgraded in terms of the evidence, [...] my teacher used to tell us that she was scared of proving [...] that every student deserves the grade they deserve. [...] And [we] might get grades that [we] don't deserve. That's the only thing I'm scared of. **year 11 student, secondary**

It's worth remembering that worry or concern about results and progression are a part of the experience of taking normal assessments and not unique to the TAG

process. With no point of comparison, we cannot say whether there was greater or lesser concern this year.

Overall confidence in the TAG process

This section explores overarching views of the TAG process. Many of the issues explored in the sections above, such as bias, fairness and the evidence used to produce the TAGs, contributed towards overall confidence in the final grades. These issues are not explored again in depth here, rather, we focus on how these issues impacted confidence in the overall assessment process.

The term validity is used in this section to refer to perceptions about the degree to which the grades determined through the TAG process were accurate measures of students' ability. The term consistency refers to perceptions about the degree to which the judgements were fair and comparable between students and centres.

In general, teachers and students had mixed views regarding how consistent and valid elements of the TAG process were. The discussions tended to indicate that the TAG process was perceived to have generally been undertaken fairly within their own centres to produce valid and consistent grades. However, reservations tended to focus on whether this was the case in other centres and whether the external quality assurance process would be able to correct for any variation across centres.

Perceived validity of the teacher assessments

Teachers were required to make evidence-based judgements about the grades that students had demonstrated they were working at, based on content that they had been taught. Interviewees highlighted several aspects of the TAG process that contributed to their overall perceptions of the validity of their own TAGs. These included the conditions the assessments were taken under, the types of evidence used, overall manageability of the assessments, how content coverage was dealt with, considerations for students with different learning and engagement profiles and the role of teacher judgements.

Assessment conditions

Most of the teaching staff and students explicitly commented that (at least some of) the evidence that contributed towards the final grades was collected under rigorous exam-like conditions. This included the papers being 'unseen' or novel to the students, sitting in exam halls or in similar exam-hall like conditions: undertaking the papers in silence with teachers who acted as invigilators. Where this was the case, students and teaching staff predominantly felt that this was the fairest way to ensure that performance on the assessment best reflected students' abilities.

We leant on the fact that we had a high degree of confidence in what we could control. I mentioned this right at the start, the fairness in our centre was that the students all went on the same day, they sat the same exam, they were in exam

conditions, there were invigilators. We were following the JCQ guidance in terms of running our exams. So everything we did was controlled and we marked what we had. **Deputy head of centre, academy**

[What I] actually ended up doing was a bit like I was taking the proper exams. [...] We came out at the same time as everyone else. There was no unfairness at all and there was no way that anyone could have communicated what was going to be on that paper. **year 11 student, selective**

This experience was not shared by all though. Some of the teaching staff we spoke to reported how sometimes assessment conditions were not sufficiently controlled during their own assessments, and they suspected that some students' grades were superficially high because of this. Several teachers suspected that students had already studied questions and mark schemes, or practiced answers for some or all of the assessment materials, particularly where the assessment materials were past papers that could be accessed online, or where the centre chose to assess students in a staggered sequence (such that students taking the assessment first could inform those taking the same assessment later on).

So, they all did it on a Wednesday because they had [...] six Wednesdays in a row, but five counting as the grades that went towards making the judgement, and of course you'd have period one you'd have one lot doing it, they'd just go and tell the next lot what's on the paper, so by the time you'd got to period four they'd had much more time to revise the thing! **Teacher and/or tutor, selective**

We could see it in the results where the students had got lucky and they'd gone onto [the awarding board's website] and they'd found the exemplars and their mate down the road had given them a bit more help, we could see papers where that had had an impact. And we could see the papers that they found harder. So whilst it was frustrating and there were anomalies in there. We'd got students who'd got 5s in their mocks, all of a sudden were pulling out 8s and 9s in their minis and that was hard. That was hard to rationalise because some of the answers were word for word what was written on an exemplar that was out circulating round. **Deputy head of centre, academy**

Several students recognised the limitations of these less-controlled assessments. Students reported that there were opportunities for pre-preparing for the assessments by looking for past papers and mark schemes on exams boards' websites.

Yeah, the exam boards have given us good exams, but they put them online for everybody to use, including the students. So, a lot of people went on those exam board pages, got the test, searched up all the answers for them, and the next day took the test. How is that an advantage? Or some people did the exams online, and then they could have cheated, we don't know if they cheated. [...] And the fact that the test was already online, people could have searched up the answers,

looked up everything, and then the next day done the test. And a lot of people have done that for a fact, because I know that they have spoken to me about it, that they've done it as well. **year 11 student, independent**

Because of this availability of materials online, some students felt that this turned the assessment into one where memorisation of the 'correct' answers was rewarded, rather than skills and knowledge.

There was also definitely an element of memorisation there, and that's not my strong point. Because I know a lot of people were literally looking through the questions, looking at the mark scheme next to it and going I'm going to memorise that answer because I think that's going to come up on it. So, it's a bit difficult really because there's that aspect of memorisation. It's going to be massively different as to how people performed just based off that. **year 13 student, college**

Because it became, it wasn't really about how good you were at that subject, and how good you were at doing that, it was more about you can, who's the best at finding the test online, or who has the most friends who can tell them the answers and stuff. **year 11 student, secondary**

Another student said that, because of the social distancing restrictions, teachers were unable to circulate around the classrooms to provide sufficient invigilation. This student felt that this created opportunities for students to cheat and collude.

We just did ours in the classroom probably most of them in a normal seating plan. Sometimes we'd move the tables around to try and make it harder to like copy people, but it was just in the normal classroom. And obviously because of COVID the teacher can only sit at the front, they couldn't move around the classroom to check people weren't cheating. So if you were at the back I'm sure you had a massive advantage to people who maybe would be at the front. So that was also quite a big problem. **year 11 student, secondary**

A few of the teaching staff mentioned how they purposefully avoided rigorous and highly controlled exam-like assessments because exams had been cancelled. In these circumstances they felt that their assessments were the fairest for their students, as intended by the government.

They'd been very clear that we should not be doing exams by the back door. So, it had to be a very different way to kind of going into the exam hall and doing three exams. We couldn't do that. We knew we couldn't. It wasn't fair and it's not what was being asked from us. **Teacher and/or tutor, comprehensive**

The evidence used

Students were largely unaware of the specific process of combining and weighting the evidence and arriving at the final grade. For some this caused concern that the

pieces of evidence that contributed towards their grade might be predominantly collected pre-announcement, such as their results in Autumn mock exams. They felt it was unfair to use evidence that was captured so early, at a time when they did not feel fully prepared, and when they didn't know the work would count towards their final grade.

I think obviously, if the circumstances had been different, maybe I would have done better [...] like from our mocks in October, maybe if we knew for sure that the exams weren't happening before, then we would have tried harder, because we knew these [the mock exams] would count more. **year 11 student, secondary**

A couple of students further felt frustrated that the assessments that contributed towards their grades had limited their ability to show their full breadth of knowledge because they were reduced in scope from normal assessments.

I would have actually done all four of those [topics], but I actually only got to show my skills in one of them, which I think was a bit, that was one of the unfair things for me. Because obviously I'd done so much work on conflict poetry from October. I was doing it since then. I'd done loads of essays, loads of work. [...] for me it was disappointing, because I would have rather have done at least three [topic questions], to be honest, to show everything. **year 11 student, selective**

This sentiment was shared by a few of the teaching staff we spoke to, who felt disappointed for students where the process implemented by their centre only allowed for a limited amount of evidence to contribute to their final grades.

I had hoped it would be done so that those students would be assessed on the entirety of the work that they'd covered to date and given them a fair crack at the whip in terms of the range of material, rather than it ending up being quite a narrow selection that happened at the end. **Teacher and/or tutor, selective**

Ultimately though, teachers overwhelmingly felt that they were able to appropriately evaluate and weight the evidence in the context of the circumstances in which it was collected. They felt this resulted in fair and accurate grades within their centres.

Manageability of the assessments

Interviewees also often considered the degree to which the assessments were manageable for students. A few students commented on how the series of smaller assessments allowed them to better focus on the content for each test and therefore perform better.

Yeah, I thought that was really good actually. [...] Because it helped with the focus in revision, because normally when you revise it's just so overwhelming, all the stuff you need to do. So I think having smaller chunks to focus on, and then

once you'd done that you could kind of put that all aside, and then focus on the next bit, and I thought that was good. **year 11 student, secondary**

Lessons became [...] revision periods before the exams. [...] Basically we had a few chances to do quite well. [...] It's essentially the idea of giving us the greatest opportunity to show what we can do that was really useful. [...] I think overall it was a fair approach. We got multiple chances on the evidence. **year 13 student, selective**

A couple of teaching staff mirrored these sentiments, highlighting how the assessment environment and set-up was much more manageable for the students. They perceived that this enabled students to perform better.

For instance, students sat these exams in their classrooms. So immediately you're in a familiar space that you're already associating with that subject. And, there has been research done on how space impacts memory. [...] And then the fact that they were sat in much smaller chunks. So, one of the things that happens every year with any exam, presumably in any qualification, is that students run out of time. Well, that wasn't a thing this year because instead of having to manage their time in two-hour chunks for instance, they were only having to manage their time in 45 minutes to an hour chunks. **Head of department, academy**

Content coverage

We specifically asked students about their perceptions as to whether content coverage was dealt with fairly or not. Views were mixed around the extent to which students were assessed only on the content they had been taught. Around half of the students agreed that they were assessed only on taught content.

But overall with actual content that we had learnt, it was fair in that regard, because teachers weren't sort of 'oh yeah, well we're just going to vaguely teach you, stick you into a normal exam, have you answer a normal exam question and expect you to answer it to an amazing degree like you've learnt it for two years', because you haven't, you've learnt it in, what, two weeks, and you've just got a vague idea of it. So, it was fair in that respect, I guess. **year 13 student, secondary**

This sentiment was not shared by all students. Some commented that content coverage was not dealt with fairly, largely because they felt they were tested on content they did not know.

I remember I sat down in my first physics exam, I opened the first page, I have never seen that topic in my life. I can't remember what it was, but we have not learned it. I'd go on the next page and the next page and the next page, I don't know any of this. **year 11 student, independent**

It is important to consider here that it may sometimes be difficult for students to disentangle content taught to the class and content they had learnt effectively. Students could have missed some content if they were absent through illness, and some students might have engaged less well with content taught remotely. A couple of students did reflect how disruption to learning and remote learning likely resulted in some students learning the content less effectively, or not at all.

I think in that situation where students are working at home, you can't assume that they're going to have an environment that is anything like it is at school. You know, you have to kind of assume that they're not going to be able to work as efficiently and I think a lot of people will say that that is what happened. **year 13 student, secondary**

But with different people being in school at different times, some people being taught this topic better than others: in that regard it was quite unfair. So, for topic A, I might have only been in school for 60% of it; whereas my mates might have been for topic A in school, 85% of it. So, in that respect of different people getting different levels of teaching, I think was quite unfair. And especially with people having to do it online or people just not being there. **year 13 student, secondary**

Fairness across students with different patterns of learning or effort

A common concern for the teaching staff we spoke to was around the degree to which the assessments were valid measures of ability for all students. A few staff felt that the assessment methods benefitted some of their students, particularly those who struggled with end of course formal assessments.

But there are definitely pupils who benefited from the system, ones who work consistently for example, but also those students who don't do well in exams, who can't revise huge amounts of information all in one go and do full length papers. Although all of our assessments were really rigorous, they did have cut down content and therefore the papers were shorter and that really helped some of them to achieve better. **Head of department, independent**

Several teaching staff also felt that the TAGs process disadvantaged students who accelerated their learning and engagement towards the end of the course. This was particularly the case where a range of evidence was used from across the course.

And there were a minority of students who probably would have done better if they'd have been sitting an exam, because they would have risen to the occasion, and I know it's a bit of a stereotype, but, for example, I've got two or three boys who are a little bit lazy but smart and would cram up before an exam and do very well. **Teacher and/or tutor, selective**

Some students who go through the year or two years coasting a bit, going through the motions and they pull it out of the bag in the exam. They are the students who probably would have suffered from this system. [...] what's good for the goose is not always good for the gander. It's good for some but not for others kind of thing. **Deputy head of department, tertiary college**

A similar sentiment was expressed by one of the teaching staff we spoke to but regarding their whole cohort, because TAG evidence was collected earlier than the scheduled normal exams, so students missed learning and revision time.

[Candidates did the assessments] a bit earlier than they would have done and they probably, to be honest, I think a few of them may have got a grade or so higher had exams gone ahead, because we'd have had time to revise more, time to train in the exam. So, I think they were, because so much of that last term was taken up by assessments that they lost that, so it wasn't just a case of, so from about Easter it was just pretty much all assessments, so they lost the content and they did lose time that way and because the exams should have been delayed this year they would have been later, they probably took them about six weeks earlier. **Teacher and/or tutor, private candidate exam centre**

Because TAGs were based on evidence that was collected over a potentially long period of time, it was rare for teachers to feel that they didn't have sufficient evidence to make a valid judgement. However, evidence was limited for some students.

The ones I struggled with are those who didn't have pieces of evidence that I could use. That was really hard. So those that had struggled in and out of lockdown or just because they were lazy or they hadn't done bits of homework, I think that was the hardest bit for me, was trying to find, either alternative evidence, or think about the grades that I'd given, just on the evidence that they had.[...] I had one student in particular who had missed a lot of external, or internal work for me, but the work he had done was of a good quality. So that was really hard for me to work through as to what to give him as a justification. Is it fair? **Teacher and/or tutor, sixth form college**

The role of teacher judgement

As discussed within the section entitled '[Professional judgement and/or holistic approach](#)', professional judgement was applied by teachers when looking across the evidence to determine the final grades for their students. They felt this was important to ensure that the final grades were a true reflection of students' performance, particularly for those students whose performance fluctuated across the various pieces of evidence. A couple of teachers reflected that this close knowledge of the students was important when applying their professional judgement.

So, this idea of trusting your professional judgement. Because yes, we had the data, but we also knew the students because we'd taught them for two years. So,

we had what we knew of them from the first year as well and that feeds into your professional judgement. But one or two of my colleagues, a member of staff left [...] at the end of the first year [and] the class got passed on without passing the marks on. [Then a] new person picks them up, doesn't know them from Adam. All of a sudden, sets an essay in the run up to Christmas, one kid might have underperformed, and all of a sudden, you're looking at it going 'oh well that kid's a D'. Well they're probably not. So it wasn't a great system if you hadn't taught them for a full two years. **Deputy head of department, tertiary college**

The application of professional judgement could also be viewed as important to determine fair grades for certain kinds of students, such as the case below.

I have one girl in particular this applied to, she was a strong mathematician, and she was good in class, she could do it all, but she just could not do exams and tests. She just went to pieces, and it all just went wrong. So this system worked brilliantly for her, because I could look at what she'd done in an exam, but also look at my professional judgement. Whereas if she'd have done GCSEs I think she would have found it a lot more difficult. **Teacher and/or tutor, independent**

Students often reported feeling comforted by the perception that their teacher could exert their professional judgement and take a more holistic view over a range of evidence. Those who commented felt that the role of teacher judgement should enable a fairer reflection of the students' abilities.

From what I gather, the subject teachers were given the whole discretion on what evidence was used. Because I think they obviously knew best what was representative. And I think it was in the end. I think it was representative of people's ability. **year 13 student, selective**

In this situation, I would choose [being marked/graded by] the teachers [over someone external] because I know that they're going to have a better opportunity to look at everything and make a fairer judgment based on work done outside of exams, I think. **year 13 student, secondary**

Although they did not think they were personally affected, there were a few students who felt that bias could be present in teacher judgements, even if this occurred unconsciously.

There are other students that I think were maybe treated a bit unfairly by teachers. Just because they didn't get on with them on a personal level. And obviously while you should try and keep your academic judgment and personal stuff apart, ultimately there's going to be some crosstalk. They're going to affect each other in some way I think. [...] I think you know, with GCSE, A-levels, BTEC teachers, all of them are going to have, to some extent, that difficulty in disconnecting their academic judgment from their personal approaches to that student. **year 13 student, secondary**

I don't think it will be fair across different colleges, because no matter how many times you hear it, there'll always be favouritism in classes, colleges. Teachers will feel bad for students that maybe won't achieve the A grade so they'll boost them up, because I think if they gave them maybe Ds or lower they'd be very disappointed and especially if they really like the student they'll be more likely to increase the grade. **year 13 student, college, private candidate**

The teachers that we spoke to did not express any concerns about bias in their own TAGs, but there were concerns around consistency between centres that are detailed in the section entitled '[Between-centre consistency](#)'.

Perceived consistency of the TAG process

There were several aspects of the assessment process that students and teaching staff discussed in relation to consistency. We separately describe within-centre consistency and between-centre consistency.

Within-centre consistency

The main themes that emerged in relation to perceptions of within-centre consistency were around teacher professionalism and the internal quality assurance (IQA) process.

Overall, teaching staff and students felt that they trusted the professionalism of teachers within their centre. This was largely driven by perceptions that they had integrity and the relevant experience to undertake a fair and valid assessment. The majority of teaching staff we spoke to felt that the IQA processes in their centre were reliable and robust. Several teachers commented that teaching staff within their centre who had examining or moderating experience for exam boards gave them confidence in their marking and grading, and allowed them to share experience within their centre to ensure consistency.

I mark for the exam board, and lots of people did, so we were able to share best practices, this is how [awarding organisations] do it, [...] so we were able to share that experience to make sure that it's reliable, which is great. **Senior leadership team member, independent**

Fortunately, I am an examiner and my colleague who is head of department has also done some examining, so we had a little bit of an insight there in terms of trying to be consistent and moderating work and things. [...] So we'd already got moderation systems in place because of my examining experience. [...] I felt that in any given assessment a student who had a teacher that was particularly generous would not benefit over a teacher that was particularly mean when

important assessments were being marked and graded. **Teacher and/or tutor, selective**

As we saw in the section entitled '[Marking of individual pieces of evidence](#)' multiple marking, as well as checking and discussing marks, took place in most centres we spoke to. Teaching staff commented that this multi-step process, involving many different people, gave them confidence that marks and grades were accurate.

We moderated each other's work, and if there were any ones we were unsure of we'd have meetings where [...] there were five or six of us saying, 'well what grade do we think this should be?'. [...] And so, work was moderated across the whole department, as in geography and history, and then there was quality assurance within the centre. [...] I felt as a school that was very thorough. [...] we all felt we could sleep at night with the grades we'd given and that if they were challenged, we had evidence to support our decisions. **Head of department, comprehensive**

We also saw earlier that student anonymisation was sometimes used in the marking and checking process, and to determine the initial TAGs. Where this was discussed, the teaching staff felt more assured that the process was fair and unbiased, and aligned more closely with practices in normal external exams.

Working with other centres also gave teaching staff confidence, as they felt this gave them a more objective, unbiased view of student work.

I went to another school and moderated with another colleague and looked at work and we looked at the levels. And then I came back into my own school and I had another colleague who hasn't taught that particular group and then I moderated with her as well. So that we were sort of checking as we were going along to see that the levels were right and where we'd put them. [...] There was one girl that did go down. [...] That's the useful bit that you're talking with people from other centres and other colleagues who don't know the girls then you have to take away the personality and just look at the work. **Deputy head of department, independent**

While, on the whole, teaching staff were happy with the IQA process, a few teachers experienced some challenges and concerns, which made them less confident in the TAGs submitted to the exam boards. This occurred where they felt that standardisation and marking training was necessary but lacking.

The staff didn't have training as such in terms of a whole staff training at all. [...] By the time [some colleagues] were judging the TAGs and assessing students' assessment work in May, that was the first time they'd used the exam board's mark scheme and materials. And so, there was a need for training and I felt that the school didn't [manage that appropriately]. **Teacher and/or tutor, selective**

Other issues reported were where a centre's SLT prioritised profiles of historical results over teachers' professional judgement, as described earlier.

The students we spoke to did not typically comment on the IQA process within their centre as they did not have sight of it. The few students who did discuss it indicated that they were generally aware of a process that teachers were undertaking to ensure that marking and judgements were accurate reflections of students' ability. They felt this protected against any errors in judgement, both at the level of marking individual assessments and arriving at the overall grade.

I think the head of departments do look over grades as well with teachers just to make sure nothing too crazy happens with the grades. And in fairness in most schools, from what I'd like to believe anyway, there's been a moderation process of it going through heads of department, deputy heads, head teachers, and then the examining bodies are doing their own external quality assurance. So I don't think it would be that easy to get passed highly inflated grades, but obviously grades are going to be higher this year. **year 13 student, selective**

Where like my biology teacher, it's not the greatest relationship with, but then they made sure because there's been a few problems brought up with teachers in biology, different teachers were marking them. So, it wasn't a teacher you'd been taught by that was marking. So that worked quite well. I think chemistry did that as well. **year 13 student, college**

However, as discussed in the section entitled '[The role of teacher judgement](#)', bias in teacher judgments could remain a concern for some students.

A student studying for their BTECs further reflected on how, because the work and internal marking had been verified by the exam board, they felt confident that their coursework marks were accurate, and reliably reflected their performance.

For the BTECs we actually had to do a pilot study report. [...] With [pieces of coursework] having been verified by the [AOs], you have those two pieces of evidence which are solid, saying that this person has got 'this' and can achieve 'this'. [...] Putting weight on coursework is important in a situation like this because they are the only pieces of fully moderated evidence that you can get for a student's ability. **year 13 student, secondary**

Between-centre consistency

When thinking more widely about overall fairness of the teacher assessment process nationally, teaching staff and students often questioned the consistency in the approach to assessments and grading between centres. A substantial number of the teachers and most of the students we spoke to mentioned how they suspected other centres were adopting processes that could unfairly advantage their students.

These interviewees suspected (or had been told by teaching contacts or friends and family) that students in other centres were being unfairly supported. They described a variety of types of such support, such as offering more heavily scaffolded assessments, prior access to the assessment materials, being allowed multiple assessment opportunities, and disregarding poor performances even where this was more representative of the student's ability.

So we did everything [right] but I know for a fact that's not how other schools did it. I know that for some schools they cherry picked the units that they knew students would do best in or that they had not taught during remote learning or, you know, things like that. **Head of department, academy**

Well from what I've heard from my friends at other schools I'm not sure I'd say it's fair. So I know some people where they went to schools where they were given as much time as they wanted to do the exams and people were using their phones in the exams or looking at computers and they had notes and stuff. And then I know other people who had little in class assessments which I guess isn't unfair it's just a very different way of doing it. **year 13 student, college**

A few students discussed differences in content coverage across centres and the impact of this for grading standards.

You hear from other schools that they've done barely any assessments, and you think it's not really the same. Because if another school does two assessments on just two topics, and we've done everything [been assessed on all of the course content...] if they got an A for that then they're getting an A for half the subjects. **year 13 student, independent**

Reflecting on these perceived differences between centres, a couple of students felt that they would rather have taken end of year exams as normal, because this would have ensured parity in assessment conditions for all students.

I don't think it was very fair ultimately. [...] Because of it being different between different colleges, different subjects, different exam boards, there's just so many differences with it, I'd prefer to have just sat the exam and then everyone's on the same thing, everyone's in the same situation. **year 13 student, college**

These perceived inconsistencies across centres were often linked to uncertainty about the approach of teachers in other centres. Some interviewees felt that other centres would artificially inflate grades by using what they perceived to be less valid approaches.

I know what other centres did and it was way less rigorous than that. And do I feel like then, what, are my grades going to be fair, lower? Am I going to have fewer As and A*s percentage wise than another centre? **Head of department, sixth form college**

Not all teaching staff felt this way though. A few suggested that overly generous grades could be a result of inexperience rather than purposeful misconduct.

I'm confident that I haven't overestimated my grades. I've given them a fair thing. But I'm not totally confident that that's going to have happened in every school for lots of reasons. Most of them had staff that are less experienced or whatever so their judgments are perhaps not the same, rather than actually because people are out there to manipulate the data and lie. I think that's a tiny proportion if it's any at all. **Teacher and/or tutor, academy**

Furthermore, several others had full confidence in the teaching community and trusted the integrity of teachers across the country.

I think that in the majority of cases, teachers are reliable, trustworthy people and we're just doing our very best. So I'd like to think that the process is reliable but yeah, I do think there are going to be some anomalies. **Head of department, academy**

The teaching staff we spoke to also often considered how the EQA process would affect reliability of TAGs between centres. For a few, they felt that the mere presence of EQA would encourage centres to take robust and fair approaches to the teacher assessment and grading process.

I think when we knew that we would be sending a sample off, most schools probably pulled their socks up. So it's more about the impact it had when it was released what was happening as opposed to the actual checks themselves. So I think, if you like, the threat of the check was probably more powerful in making sure that people stuck to the rules. Would be my guess. **Senior leadership team member, academy**

Many others however, felt that the EQA process would be insufficient to appropriately detect and manage all occurrences of unjustified grades. The teaching staff we spoke to commented on several aspects of the EQA process that made them feel this way. Of those staff teaching subjects for which a sample was required for quality assurance, some felt that the sample requested was too small..

We spent a lot of time gathering the evidence, we spent a lot of time making sure everything was up to scratch, everything was moderated, etc. Making sure everything was in place and we were asked for three students from one subject for A levels and we were asked for maybe five students for GCSE maths, none for A level maths. I just thought 'was the exam board just ticking a box?' **Teacher and/or tutor, sixth form college**

One interviewee noted that there were a large pool of existing examiners in schools who had not been invited to assist with EQA. They felt this was a missed opportunity that would have enabled many more centres and grades to have been checked.

I mean there are thousands and thousands and thousands of schools in the country, so how can they possibly take samples and who's actually doing this work. Who's looking at these papers - because in my computer science department, I've got an examiner who's one of our teachers and he's not been employed this year as an examiner. So, there's no one to look at the samples. It just seems a nonsense. There's not going to be a QA process. **Head of department, university technical college**

Teachers noted that the lack of consistency between centres in the evidence that was used would make quality assurance fundamentally difficult, and that there would also be gaps in evidence.

I don't think it's possible to really have a particularly good quality assurance process when every school's doing different things and when we've had very little guidance from the government. You know, I can't see how you can compare the grades that we give from our school with the results from one of the schools in [the town] just a few miles away, who may have done something completely different and will have assessed on different things, given different papers. I don't understand how that could work. [...] So I don't [understand], what's their quality assurance process going to be? [...] They've taken a sample of one our grades. [...] they can say 'yes, your grades that you've given, I agree with those', but again how does that compare with other schools in different areas. I can't see that that's much of a quality assurance process. **Head of department, university technical college**

Overall, the teaching staff we spoke to had hoped that the EQA process would involve more scrutiny of evidence. This would have given them confidence that the process was able to address any unevenness nationally. Some recognised that there were limitations, but hoped that those centres that were most misaligned with previous results would be looked at closely.

The volume of work that needs to be done to quality assure to the level that we've quality assured internally is not achievable by the exam boards. And I do feel like the quality assurance process is a bit of a, like, I don't have confidence that it will make any difference. I don't have confidence. I do feel that there will be some sort of data checking exercise and I do hope that centres that are way out of kilter will have their information scrutinised. But I just do not see how there is the resource to actually, to do anything other than trust the teachers and the centres that we've done it for them. **Deputy head of centre, academy**

Overall view of the TAGs

A number of the teaching staff we spoke to explicitly indicated that their TAGs were more generous than grades in previous years, when exams had taken place. Where this was the case, they felt that this was legitimate, describing several reasons for

this. Teaching staff highlighted that the assessments were fundamentally different for TAGs compared to normal exams. For TAG assessments, students generally had multiple opportunities to show their ability, often in different ways, and there was less impact of having a single 'bad day' in an exam.

But we've probably seen a slightly larger increase in better grades, because it was less about being able to perform on the day. An opportunity to have other evidence taken into account, or if you did have a genuine reason why you were unable to sit that paper on that day, we've allowed them to take it as soon as possible afterwards or what have you, which has helped the odd student. **Head of centre, further education establishment**

The circumstances of the pandemic also benefitted learning for some students. For instance, a couple of teaching staff reflected on how some students had performed better than expected because they were less distracted from their studies by other activities, which were put on hold due to lockdowns and social distancing restrictions.

I was talking to a colleague from another school and she said actually what they were finding is that they were probably, they were coming out slightly better some of the middle range students than they might have been normally because they haven't had the distraction of being able to go out and do something else because they've been in lockdown. So all they have to do is do the work. [...] So they were finding that for some candidates they were reaching a higher level because they weren't playing football or swimming or doing whatever. They've been stuck in and they've had to do their work. **Deputy head of department, independent**

As noted earlier, in the section on '[Personal responsibility](#)', teaching staff were aware that awarding over-generous grades might lead to students making inappropriate choices. They believed that, to some extent, this would have acted as a brake on inflated grades.

Yeah, because we've tried to behave with integrity, particularly at GCSE because being a grammar school our offer at post 16 is an A-level offer and if we inflate our grades massively at GCSE, instead of having a sixth form of around 140/150, we'll have a sixth form of 180 and actually there's 30/40 students there that an A-level programme is not appropriate for. So not only did we do it because we felt it was the right thing to do, we also felt like it was the right thing to do for the students to not go to you yeah you'll be fine doing A-levels, we're going to give you loads of 5s and 6s, so you can access the sixth form and then arrive into the sixth form on an A-level programme that you're not going to manage. **Deputy head of centre, selective**

One of our biggest jobs is to get them to the next stage of their study or employment, the next stage of life and be successful and if we don't train them effectively, if we don't give them the right level of knowledge and skills then they

won't flourish in the next part of whatever it is that they choose to do. [...] what you don't want to do is set a young person up to study something [at university] that they will then find so challenging and set them off down the wrong path [...]. You've got to set them up for success at whatever point, whether that's coming to us or going on from us to the next stage, it's not just about grades. **Deputy head of centre, sixth form college**

When looking back on the whole process, some teaching staff felt that cancelling exams and awarding TAGs was the best decision for students completing a qualification in 2021. They recognised that every centre had experienced a different level of disruption because of the pandemic and therefore that a normal exam series would not have been fair to students.

I don't feel exams could have gone ahead or if they had gone ahead I don't think it would have been fair, because young people had such different experiences up and down the country, so I don't think you could have realistically run an exam series. **Teacher and/or tutor, comprehensive**

Because of concerns surrounding the consistency of awarding grades between schools and internal/external pressures discussed earlier some teaching staff felt the process could not be entirely fair.

No, and I still don't think it's the fairest decision for them, because we haven't had any mechanism in place to do any kind of cross-centre moderation. And I think whatever Ofqual and the exam boards have said, we have been subjected to pressure from students and their parents and indeed in our case from our own senior leadership team to be optimistic about results. **Teacher and/or tutor, selective**

Looking beyond the TAG process

This section explores views looking beyond the TAGs, in particular around the readiness of the 2020-2021 cohort to progress to their next year. This section also includes reflections on expertise that staff gained through carrying out the TAG process.

Readiness of 2020-2021 cohort

General preparedness for next steps

In many of their interviews teaching staff considered how prepared their students were to progress, usually in the context of moving onto A levels or university courses. Teachers commonly acknowledged that this cohort were unlikely to be as well prepared as previous cohorts because of the COVID-19 pandemic. However, many teachers believed that they had done their best to prepare students.

I honestly think we did as much as we could to teach them to the end but they've not left us as well equipped as a normal year group would have. **Deputy head of centre, academy**

Many of the teachers thought that the TAG process in 2021 had been better for students than the CAG process in 2020 because it required them to prepare for and complete actual assessments.

So I think this year's cohort will be better prepared [than last year's cohort], I think they have been through an examination effectively, they've revised, they've had that shared experience of going into the hall together, of feeling the fear and doing it anyway, of coming out and doing the autopsy as I call it of going over the paper and "oh that one, oh God did you write that?" [...] I think all of that's really important for them, because it is a rite of passage, it is a coming of age in the British education system anyway. So I think they will be in a much better place and much better equipped for education or for assessment in the future. **Teacher and/or tutor, comprehensive**

In general, students had mixed views as to how prepared they felt going on to their next year. Some were confident, others less so. This was often based upon their personal circumstances and what they intended to do next.

I'm planning on going to university in September. I don't think I'm worried about it, because although the grades have been different to what you'd expect of just doing A-levels and then exams, because I've been doing tests quite a lot, so I'm used to that. I've been doing lots of independent work particularly with my private

work, so I'm not worried about going to the next part, no. **year 13 student, college, private candidate**

I'm not sure to be honest. [...] for English literature the thing is, it's been two years since I've last done any exam questions or practised on poems and poetry. And poetry is a big part of English literature in A-levels. That's one thing I'm worried about. Another thing is, for example, the sciences I think I'm fine on them because we've learnt everything for science and maths [...], but it depends on what subject it is to be honest. **year 11 student, secondary**

Loss and gain of skills

Teaching staff often considered the skills that students had acquired, and how this impacted the degree to which they were prepared for further learning. In some centres, because of the nature of the assessments used to support TAGs, staff felt that students may have missed out on developing some of the skills associated with completing external exams.

They changed the way they revised, they didn't take that holistic view of their revision. They were like looking at the questions and they just approached it in a completely different way. They've got to relearn that now at a critical point and they're going to find their A levels hard, not just because of their knowledge gaps but also because of their skill gap in terms of the way that they prepared. **Deputy head of centre, academy**

Some teaching staff described earlier how students lacked the experience and stress of sitting high-stakes external exams, and these concerns were also felt by some of the students that we spoke to.

I haven't done an actual exam. The stress of doing a real exam that [decides] if I get into university or not is quite stressful. And I'm not prepared for that element of it. I think content-wise, I'm going to revise all year and I'll be fine, but the actual element of doing the exam, I don't think I'm really prepared for. [...] I haven't done a formal exam since my SATs in year 6. **year 12 student, college**

Various concerns around specific skills such as practical and fieldwork skills were mentioned in the context of the broader learning loss experienced by students. Although many of the teaching staff and students we spoke to recognised that some skills would be less well developed, several identified some benefits in terms of resilience and independence from navigating both the pandemic disruption and the TAG process.

I think in that respect yes they have learned their revision skills, and doing the exams, which last year's cohort didn't have, and I think it's also had to teach them to have that resilience just to keep on going, keep on pushing through. **Teacher and/or tutor, independent**

And I feel very much prepared in the sense of I've gone through so much since January with the TAGs and it's kind of what we've gone through is like a worst-case scenario basically. So, I feel like at this point going into A-level, if anything happens again—which hopefully by the time I finish it won't—but I feel like I can conquer pretty much anything at this point. **year 11 student, selective**

Many teaching staff and students described knowledge gaps caused by the pandemic disruption, where the whole curriculum had not been taught or students had struggled with remote learning. This is outside the scope of this review, and we note here that students expressed different levels of concern, depending on what they would be doing the following academic year. Teaching staff sometimes spoke about how they had done their best to fill in some knowledge gaps once the TAG process was complete.

We also note that some of the teachers suggested some younger year groups might struggle in future because so much time and effort had been dedicated to the TAG process.

I haven't just got Year 11 in the school, we've got all the other students in all the other year groups. We're trying to sort out catch up for them and, all of the staff's energy and time. Our staff room turned into a marking room. It was like a machine at times, everybody trying to do their best but inevitably all of that energy going into that means that it's got to have been taken from somewhere else. And unfortunately, the other students have, doesn't matter what you say, they have missed out because we've been so predisposed trying to get this [TAG process] right. **Deputy head of centre, academy**

Expertise gained from the process

Despite the challenges, there were some teachers who described ways in which their own expertise and skills, and that of their colleagues, had developed as a result of the TAG experience, which would benefit them in the future. Several teachers suggested that they had become more accurate markers.

I think we are now much more accurate at our marking. I think that staff, I think we can mark papers, we've got a much better understanding of the grading. We certainly look at students' answers much more closely than we would have done if we were marking mocks and things, and we've talking about that in a meeting. That is the one good thing that's come out of all of this, is we are now much better markers. **Head of department, comprehensive**

There were also several references to teachers having gained a better understanding of reporting accurate working grades and being able to use data more effectively in future.

And the reason why I think it is better is that I think in both schools it's sharpened up people's understanding about the purpose of reporting accurate current working grades that are driven by a holistic picture of everything to date the child has achieved that tell you where they're currently working at. **Head of centre, comprehensive**

It also seemed that several interviewees had created resources which would be useful for them, even in a 'normal' year.

And one of the things that's quite hard sometimes is to actually get any evidence back about where our students' weaknesses and strengths are. And I've now got a massive bank of papers that I can go back to, and look at all the maths skills... So again, they're things that we don't get back from the exam boards, they charge us if we want to buy back some of the papers from our students to look at them, but we've got a massive set of papers now right across the board. So we can look at that, and look for patterns. **Head of department, comprehensive**

That was useful in terms of there was a document that we produced, because we had refined down our content, basically we produced a spreadsheet with every single question listed out on the material that we were providing to our students and we said whether it was in scope or not. So, now I've got that list of questions that we'll add to each year to help. [...] Right, I want to find a question on centre of mass, right, where are the 20 questions on centre of mass, so we've got that. **Head of department, independent**

There was also a reported increase in confidence as teachers had been given the responsibility of judging grades which they felt showed recognition of their expertise.

So I think that those [increases] in understanding have meant that [...] there is greater confidence in what people are doing and because there has been collaboration and consideration given to them as being the experts, it's just felt like a more respectful process this time round. **Head of centre, comprehensive**

It seemed that while the TAG process was challenging, time-consuming and stressful, when offered the chance to reflect, quite a few teaching staff recognised that their efforts in summer 2021 had often benefitted themselves and their centres for the longer term.

Finally, the interviews also included discussion of lessons learnt and thoughts that teaching staff and students had on improvements should a similar process be required in future. These views and ideas were used to inform policy and in the design of contingency arrangements for 2022.

Discussion

In this section we consider several cross-cutting themes, and also some of the caveats around over-interpreting the findings.

We start by noting that we do not know how representative the reported experiences are of the entire English school and college teaching staff population, since there may be a degree of self-selection in our interview sample. There may be a variety of motivations for engaging with us, and it is not inconceivable that we spoke to a group of teaching staff who mostly felt very confident in the systems and processes their centres used. Similarly, our small sample of students requires us to treat their experiences more like case studies. We discuss this possible sampling bias further below.

Confidence in TAGs, but a heavy workload

The main finding from our interviews was that those teaching staff we spoke to expressed high confidence in their own TAGs. However, the process entailed a substantial workload and stress occurred because of this. They talked extensively about the creation of assessments and the conditions under which those assessments were taken. They described the systems they put in place for marking assessments, determining the TAGs based on those assessments, and quality assuring the outcomes. All of these controls and checks provided them with confidence in the fairness of their own grades.

A variety of approaches to determining TAGs were taken, as was permitted, but most emphasised the use of more formal/controlled assessments. Formative (less controlled) work was also regularly used to support judgements, as was the use of statistical information about previous years' results and the ability of the current cohort. Interviewees described a variety of approaches to aggregating evidence and determining initial TAGs. Some relied more on holistic judgement looking across the profile of grades or marks for the evidence collected. Others took a more analytic approach, using rules or calculations to decide the initial grade for each student.

In almost all the interviews, professional judgement was applied, and adjustments were sometimes made to the initial TAGs based on knowledge of the students. Typically, the TAG process involved several members of staff applying several layers of scrutiny to the assessment evidence that had been gathered for each student, both within departments and during internal quality assurance (IQA). Interviewees told us that this helped to ensure that TAGs reflected evidence of student performance, and minimised unconscious bias. IQA often involved SLT querying or challenging TAGs. Sometimes this involved direct requests to consider grade profiles from previous years, but in many other cases departments responded to queries by providing explanation and justification for their judgements.

However, the downside to the perceived robustness of the TAG judgements was the very high workload required to achieve this, described in all of the interviews, and the stress this caused for most of those we spoke to. A variety of sources of extra workload were identified, many of which related to the timescales in which TAGs needed to be determined. Part of the issue was that final guidance and supporting materials could not be made available immediately after the announcement that examinations were cancelled.

The assessment materials from exam boards did not contain as much new or unseen material as hoped for and many teaching staff had been waiting for these to be released before beginning to design assessments. They therefore found themselves with limited time to complete the evidence-collection, decision-making and quality assurance processes. Departments that had assessment expertise in the form of experienced examiners and moderators utilised this to help design their processes. We heard how in some centres this existing expertise was shared. The experienced examiners or moderators sometimes worked with other departments, and VTQ teaching staff who were familiar with carrying out internal assessments also sometimes helped their GQ colleagues in thinking about moderation processes.

The feeling we received from almost all of the interviews with teaching staff was that they did not simply aim to deliver an adequately robust set of assessments, but they did their best, using all of the skills and experience they had available, to collect the best evidence they could. Their aim was to produce outcomes that were as reflective of the student's abilities as they could be. However, there was a strong consensus that because this was a large and difficult job, more support and guidance from all official bodies would have been appreciated.

Consistency across centres

Those teaching staff we spoke to were confident in the processes that they and their centre had put in place to support TAG judgements, but were generally less confident about whether other centres had been as thorough and careful. Given that teachers and students alike would not have had full awareness of the processes being operated at other centres, it is perhaps unsurprising that they expressed some anxiety about the potential for unfairness.

Whether well-founded or not, reports of how other centres or teachers had determined TAGs had clearly circulated through the community and social media. From what we heard, there were concerns that other centres had bolstered their evidence of student performance by allowing multiple attempts at assessments, revealing too much detail about the content of assessments in advance, or allowing the use of textbooks or the internet in what would normally be closed-book tests. There were also widespread fears expressed by both teachers and students about other centres just being over-generous in their grading standards.

Some inconsistency between centres would have arisen unintentionally through the process. Grade descriptors provided by exam boards for GQs were one of the elements of the process that might have been expected to increase consistency between centres. They provided qualitative grade-specific descriptions for each subject of the level of performance that a student would be expected to demonstrate in their evidence to achieve that grade.

While most interviewees considered the grade descriptors too imprecise to easily apply when assessing individual pieces of evidence, they were judged more useful when determining the overall TAG for each student. This involved a holistic evaluation of the quality of all the selected evidence against the relevant grade descriptor. In these cases, the descriptor provided some assurance that the final TAG felt appropriate. However, the holistic judgement was still quite difficult, and there was a clear view that this had been more straightforward for essay-based subjects, but much more difficult to use for those subjects typically assessed using many short questions or tasks.

Decisions on TAGs were also affected by how centres interpreted the requirement to consider previous patterns of results when quality assuring their TAGs. Some centres may have been stricter than others about submitting sets of TAGs that were comparable to results achieved in previous years. To a degree this may have been partially driven by each centre's expectations of the external quality assurance process.

Differences in the grading standards applied by different centres were therefore considered possible, and there was some concern that the external quality assurance (EQA) process would not be sufficient to ensure consistency. Teachers pointed out the limited size and range of the sample of work collected for EQA, and the inherent challenges in making judgements on such large and diverse sets of evidence for each student.

Amount of testing

The approach to determining TAGs was intentionally designed to be flexible, to allow centres to decide how best to assess their students on the content they had been taught. This was necessary due to the large differences in the level of disruption that was experienced by different centres. If a highly restrictive approach had been specified, it is possible that some centres would have been unable to meet those requirements.

One consequence of the flexibility appears to be that some centres carried out a lot of assessments. This was for two reasons. One was to ensure that a sufficient bank of evidence for each student was gathered on which to base TAGs, with a view to meeting the external quality assurance requirements. A second reason was to allow students to have plenty of opportunity to show their best work. For some of the students we interviewed, the level of stress was very high because they perceived

every one of these assessments to be high stakes, given that they could count towards their TAG.

Upskilling following the TAG process

A point mentioned in quite a few interviews was that running these assessments had probably increased expertise in their department or centre in a way that would benefit them in the future. Many interviewees believed they and other staff in their department or centre had an improved awareness of assessment practices, in terms of designing assessments and carrying out more effectively standardised marking and grading than they had before. Some interviewees also reported that they felt they would make much better use of data they held following the TAG process. Other staff spoke about creating resources that would be useful in future, such as structured question banks based on the exam board assessment materials.

Our interviewees almost all stated that they would not choose to repeat the TAG process in 2022 if they could avoid it, but because of the experience they had gained they recognised that it would probably be easier a second time. Even if such a process were never repeated, this perceived upskilling was considered to be likely to lead to better teaching practices, in a similar way that teachers often take up an examining or moderating role with an exam board with the intention of improving their teaching and preparation of students for exams (Lockyer, 2018).

Sample strengths and limitations

This report reflects the views and experiences of those individuals who were willing to speak to us and so may not be representative of the national population of those who were involved in the TAG process. This is normal in any voluntary qualitative study. Because we cannot conclude that the relative frequency of different views and experiences in the interviews are representative of the national picture, we chose not to describe these frequencies in anything but the broadest terms. Despite this limitation, the research still represents a wide cross-section of views and experiences. It allows us to consider the ways in which TAG processes may have differed in practice, in the context of specific schools.

There may be a variety of motivations for wanting to share views and experiences in an interview with the qualifications regulator. For example, it was clear that some of those we interviewed were keen to tell us about the difficulties they had faced with workload and the resultant stress they experienced, so that decision-makers would be fully aware of this. Others wanted to highlight either good or poor practice in their centre or wanted to describe the lengths that they and their centre had gone to, to carry out this task in the best way they were able, to show how fair and accurate they had been.

Giving students their voice is vital and the interviews with them reveal how the experience felt to them. We spoke to only 14 students, and these interviews are more appropriately treated as case studies that allow us to explore the ways in which students may have been affected. We would have spoken to more students if we could, but unfortunately only a small number of students who had indicated interest in an interview on the earlier survey agreed to participate. We recognise that we were recruiting at the start of the summer holiday, and we are incredibly grateful to the students we did speak to for giving up their time to share their views and experiences.

Comparison of the survey and the interviews

The interviews built upon [many of the findings reported in the survey report](#), but provided more detail about the entire process within individual centres. Many of the issues analysed and described in the open response questions on the survey were clarified through discussion. For these open response questions, an overwhelming majority of comments were negative about aspects of the process. But when speaking to individuals, once they had described their difficulties with the process, such as the stress and workload involved, they then began to reflect on some of the positives that came out of the experience such as their increased expertise. These factors tended not to be mentioned in the survey.

We noted in the discussion section of the survey report that we believed we had not used the right labels to describe the exam-type evidence that teachers had used to determine TAGs. The interviews clarified some of the terminology around mocks, exam-like assessments and test materials provided by awarding organisations. This information allowed us to understand why the ratings of the importance weighting for different exam-style evidence types in the survey were lower than expected. Some of these evidence types received higher importance ratings in the 2020 CAG survey than this year, but we know from the interviews that very many TAGs were heavily based upon evidence collected in exam-like assessments.

What became clear in the interviews was that a range of terminology had been used by centres to describe tests under exam-like conditions. There appeared to be two main motivations driving this. Because centres ended up running a lot of exam-like assessments, although they were often shorter, they used a variety of names to describe them. This was to try and distinguish these from the cancelled external assessments so that it did not appear that they were bringing back external exams 'by the back door'. They also wanted to make these tests feel lower-stakes for the students to ease the pressure on them. However, it appears that students often felt just as stressed for these assessments as they would have for external exams, and there were often more of them, sometimes over a longer period of time.

Overall perceptions of TAGs

The TAGs judged in 2021 were based on evidence of actual performance collected from students, in contrast to the 2020 CAGs that were based upon a prediction of how students would have been expected to perform in future assessments had they gone ahead, informed by student's existing work. TAGs therefore required no prediction, while CAGs required both a prediction of how future learning and achievement might have changed following the school closures, and a prediction of how well each student might cope with exams.

The general view in this year's interviews was that TAGs had been based on more robust evidence than CAGs. CAGs were based almost entirely on professional judgement, since although some sources of attainment evidence were available (mock exams, tracking grades etc) these were not always taken under tightly controlled conditions, and were not subject to a process of external quality assurance. Professional judgement of teaching staff was still a strong element of TAGs, but the judgement had to be based on clear, verifiable evidence.

The teaching staff we spoke to felt strongly that their students were getting the grades they deserved, that their professional judgement had been listened to and had been fair and unbiased. That is not to suggest there were no difficulties in determining TAGs. Those we interviewed reported that it was more difficult to determine TAGs for students with inconsistent performance, those new to centres, and those unable or unwilling to complete tasks to produce evidence. They also expressed concerns about comparability across centres.

The overriding view of our interviewees was that final external examinations were the fairest and most valid assessment approach. While a desire to avoid a TAG-like process in future, because of the burden they placed on staff, influenced some of these views, exams were considered to be the best way to fully ensure consistency between centres. Students largely preferred the idea of external examinations because they felt these were free from any risk of bias, and they felt a greater sense of agency – they knew precisely when they needed to perform to the best of their ability.

References

- Arrafii, M. A. (2020). [Grades and grade inflation: exploring teachers' grading practices in Indonesian EFL secondary school classrooms](#). *Pedagogy, Culture & Society*, 28(3), 477–499.
- Brandt, N. D., Becker, M, Tetzner, J., Brunner, M. and Kuhl, P. (2021). [What teachers and parents can add to personality ratings of children: Unique associations with academic performance in elementary school](#). *European Journal of Personality*, 35(6), 814–832.
- Braun, V. and Clarke, V. (2006). [Using thematic analysis in psychology](#). *Qualitative Research in Psychology*, 3(2), 77-101.
- Carbonneau, K. J. (2020). [Teacher judgments of student mathematics achievement: the moderating role of student-teacher conflict](#). *Educational Psychology: An International Journal of Experimental Educational Psychology*, 40(10), 1211–1229.
- Cheng, L., DeLuca, C., Braund, H., Yan, W. and Rasooli, A. (2020). [Teachers' grading decisions and practices across cultures: Exploring the value, consistency, and construction of grades across Canadian and Chinese secondary schools](#). *Studies in Educational Evaluation*, 67.
- Dian, M. and Triventi, M. (2021). [The weight of school grades: Evidence of biased teachers' evaluations against overweight students in Germany](#). *PloS one*, 16(2), e0245972.
- Ferman, B. and Fontes, L. F. (2021). [Discriminating behavior: Evidence from teachers' grading bias](#). *Behavioral & Experimental Economics eJournal*.
- Holmes, S., Churchward, D., Howard, E., Keys, E., Leahy, F., Tonin, D. and Black, B. (2021). [Centre Judgements: Teaching Staff Interviews, Summer 2020](#). Coventry: Ofqual.
- Jönsson, A., Balan, A. and Hartell, E. (2021). [Analytic or holistic? A study about how to increase the agreement in teachers' grading](#). *Assessment in Education: Principles, Policy & Practice*, 28(3), 212–227.
- Lee, M. W. and Newton, P. (2021). [Systematic divergence between teacher and test-based assessment: literature review](#). Coventry: Ofqual.
- Lee, M. W. and Walter, M. (2020). [Equality impact assessment: literature review](#). Coventry: Ofqual.
- Lockyer, C. (2018). [Survey of examiners 2018: Headline findings](#). Ofqual research report 18/6449/6. Coventry: Ofqual.
- Noden, P., Rutherford, E., Zanini, N. and Stratton, T. (2021). [Grading gaps in summer 2020: who was affected by differences between centre assessment grades and calculated grades?](#) Coventry: Ofqual.

Nowell, L.S., Norris, J.M., White, D.E. and Moules, N.J. (2017). [Thematic analysis: Striving to meet the trustworthiness criteria](#). *International Journal of Qualitative Methods*, 16, 1-13.

Ofqual (2020). [Awarding GCSE, AS, A level, advanced extension awards and extended project qualifications in summer 2020: interim report](#). Coventry: Ofqual.

Stratton, T., Zanini, N. and Noden, P. (2021). [An evaluation of centre assessment grades from summer 2020](#). Coventry: Ofqual.

Westphal, A., Lazarides, R. and Vock, M. (2021). [Are some students graded more appropriately than others? Student characteristics as moderators of the relationships between teacher-assigned grades and test scores in mathematics](#). *British Journal of Educational Psychology*, 91, 865–881.

Official Sensitive: Not to be shared further without permission.



© Crown Copyright 2022

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

[View the Open Government Licence at The National Archives website](#) or write to:
Information Policy Team, The National Archives, Kew, London TW9 4DU

Published by:



Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344

public.enquiries@ofqual.gov.uk

www.gov.uk/ofqual

July 2022

[Ofqual/22/6933/2](#)