

RESEARCH AND ANALYSIS

Teacher Assessed Grades in summer 2021: Surveys

Analysis of teaching staff and student surveys

ofqual

Authors

- Nathan Case
- Ellie Keys
- Stephen Holmes

With thanks to

- The teaching staff and students who gave their time to share their experiences and views.

Contents

Authors	2
With thanks to	2
Executive summary	5
Introduction	8
Method	11
Survey design.....	11
Respondents and geographical coverage	13
Information provided to respondents	13
TAG teaching staff survey results	14
Demographic details	14
Judging TAGs.....	17
Considerations for judging TAGs	21
Judging individual student TAGs.....	29
Agreeing TAGs	53
Internal quality assurance (or standardisation)	55
Final thoughts on submitted TAGs following internal quality assurance.....	60
Centre declaration responsibility	65
Final thoughts	66
TAG student survey results	73
Demographic details	73
Initial thoughts and feelings.....	77
Disruption	80
Details on TAG evidence	82
Your views on fairness of TAGs.....	92

Reasonable adjustments	96
Balance between carrying out assessments to support TAGs and further learning of new content.....	99
The future.....	101
Final thoughts.....	103
Discussion.....	108
Confidence in accuracy and freedom from bias.....	108
Perceived fairness.....	108
Perceptions of consistency across centres	109
Guidance	110
Evidence used.....	111
Impact on teachers and students.....	113
Preparedness for next steps.....	113
Limitations, demographics and survey bias	114
Annex A – Information screen on teaching staff survey	115
Annex B – Information screen on student survey.....	118

Executive summary

In January 2021 the government confirmed that summer 2021 assessments could not go ahead as planned due to the 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 judgement 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 determining TAGs, and students receiving TAGs. This report details the surveys of teaching staff and students that were run following submission of TAGs to awarding bodies (on 18 June for GQs and some VTQs) and completion of most external quality assurance activities, but prior to the release of final qualification results (on 10 August for AS and A level and many level 3 VTQs, and 12 August for GCSE and many level 1/2 VTQs).

Separate surveys for teaching staff and students were open from early July to early August 2021. We received 1,785 responses from teaching staff and 550 responses from students. Because this was a self-selecting sample of respondents who chose to complete the survey, and because we identified over-representation of some demographic groups, care should be taken when extending the findings reported in this report to the national population of teaching staff or students involved in the TAG process.

Overall, teachers expressed high confidence in both the accuracy of their own submitted TAGs and their belief that these were free from bias, with most rating their confidence close to or at 100 on a confidence scale of 0 to 100. These reported measures were at similar levels to those reported for the centre assessment grades (CAGs) used for awarding in 2020. Students had slightly lower confidence, although most (53%) did think that the TAGs would be fair for them overall.

Almost all teaching staff were aware of the Ofqual guidance on making objective judgements, but this, together with equivalent guidance from awarding organisations (AOs), was considered less useful than it had been when we asked the equivalent question in 2020. This may have been because respondents felt that more effective steps to minimise bias had been taken this year compared to last. The majority of centres looked at data from previous years to try and identify potential bias, and almost three-quarters of respondents reported that formal staff training on bias had

taken place, around double the previous year. Special educational needs co-ordinators were also frequently involved.

Respondents reported that tests taken under exam-like conditions were the main source of evidence used to determine TAGs in most GQ subjects, while practical assessments or coursework tended to be the main sources of evidence for VTQs and the more creative general qualifications. TAGs for GQs were usually determined based on 4 to 6 individual pieces of evidence. There was little difference here between centre types. Students did feel that they had been over-assessed though, with 57% saying too much time was spent on assessments at the expense of further learning.

Evidence for TAGs was generated mostly after the announcement that assessments were cancelled in January, with 78% of GCSE and A level teachers collecting most or all evidence from after this point. For VTQs, more evidence came from earlier in the course, with only 47% generating most or all evidence after the announcement, reflecting their generally modular structure and the more continuous nature of assessment typically used.

Teachers generally reported that they had delivered most of the content for the courses they taught (median values of 90% for GQs, 75% for VTQs). However, despite the high median, 25% of GQ respondents stated that they had delivered 75% or less of the content. Students believed they had been taught a little less content than the teachers indicated, but this difference may reflect content individual students missed through, for example, illness, or content that was taught but the students perceived to be not effectively taught.

Determining TAGs for all students was most frequently rated by teaching staff as being 'slightly difficult', but certain types of students were typically judged 'difficult' to decide for. These were students with inconsistent performance, students new to the centre, those missing more content than others, or to a lesser extent (and based on a much smaller sample of respondents), private candidates. These ratings were also reflected in the numbers of students for which teachers stated they were unsure of their grade; a median value of 10% in our sample.

Reference to the centre policy was rated the most important element of the internal quality assurance (IQA) process, followed by comparison of TAGs to previous years' data. Whilst most teaching staff (92%) thought that everyone who was needed was involved in the IQA process, those who thought other individuals could have been involved most frequently mentioned exam boards or examiners, and some free text responses regarding extra sources of information also suggested that better assessment materials, mark schemes, grade descriptors and training from exam boards would have been very helpful.

There was an almost equal split of teachers reporting that initial teacher TAGs had or had not been changed following IQA within the centre before submission. This may,

of course, reflect differences in the way initial grades were determined before IQA, but may also reflect differences in the use of past results during IQA in the centre.

When the whole process was considered, TAGs were perceived by teachers to be marginally more fair than the CAGs were in 2020. When compared to normal assessments, the most common response was that they were 'about the same fairness', though where a difference was reported, exams were more likely to be considered fairer than TAGs.

Teaching staff were asked for three words to sum up their experience of determining TAGs. An analysis of their responses suggests that stress and workload were major issues. This was supported by analysis of the open-response questions, where a great number of comments were about the amount of effort the process had required, with a need to create, run, mark and moderate bespoke assessments before TAGs could be decided for GQs. This generated a great deal of stress and unhappiness. Teachers also reported that they had spent around twice as many working days completing the TAG process compared to the CAG process in 2020.

In open responses students tended to report a mixture of relief and pleasure that normal assessments had been cancelled, and uncertainty and fear when considering their initial response to the cancellation. They also indicated high levels of uncertainty and worry in anticipating their final TAGs.

While teachers thought that students were fairly well prepared for the next stage of their lives (51% 'very well' or 'well' prepared vs 21% 'very poorly' or 'poorly' prepared), students were slightly less optimistic (46% 'well' prepared vs 33% 'not well' prepared). There was little difference between Year 11 and Year 13 students.

Fears about comparability between centres also came through in written responses from both teachers and students, with most reporting thorough processes in their own centres, but worrying about what other centres were doing, often based on things they had heard from contacts. A few students also referred to variation across subjects or classes in their own centre, with variable help or pre-warning of test content. Teaching staff also felt they the support they had received, particularly from exam boards in the form of guidance and assessment materials had not been sufficient.

Overall, the outcomes in their own schools and colleges were viewed by teaching staff as reliable and fair, but responses to open questions suggested that the work involved to achieve this was considerable and there was no desire to repeat the process in the same form. The findings from these surveys, and the subsequent interviews we carried out, were used to help inform the contingency planning for 2022. They also support our evaluation of the impact of the assessment arrangements that were used in 2021 on both teachers and students, helping us to understand what they did in practice and their experiences.

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 (later JCQ published guidance along with 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). 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-exam 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 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 assessment 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 surveys. A [separate report describes the interviews](#) we carried out.

Finally, it is important to remember that these survey responses were received before the qualification results based on TAGs were given to students. Students should not have been aware of their TAGs. Therefore, all questions in the student survey regarding views on final TAGs may involve some assumptions by respondents. At the point the survey was live, 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.

Method

We ran separate online surveys to capture information on approaches to, and views of, TAGs for teaching staff and students. The surveys were open for respondents from 6 July to 7 August 2021. This window fell between the completion of most external quality assurance (EQA) activities by the awarding organisations (so that most centres knew whether their TAGs had been accepted) and the results days in mid-August. We did not want respondents to be focussed on the results themselves, but rather to focus on the process they had gone through to determine those results.

Survey design

The teaching staff survey was divided into several major sections, which were automatically presented or skipped for individual respondents based on their answers to routing questions regarding which parts of the centre judgement process they were involved in. We divided the process, and therefore the survey, up into the following main sections:

- **Demographic details:** all respondents were asked about their centre type, job role, years at their current centre, and years in the teaching profession
- **Judging TAGs:** respondents who indicated that they were involved in the judging of TAGs were asked how many TAGs they judged, for what subjects, and for how many classes and students
- **Considerations for judging TAGs:** respondents who indicated that they were involved in TAG meetings or discussions were asked about what was considered when planning the judgement of TAGs
- **Judging individual student TAGs:** respondents who indicated that they were involved in the judging of TAGs were asked to consider one specific qualification and subject for which to answer questions around the practicalities and process of judging TAGs
- **Agreeing TAGs:** respondents who indicated that they were involved in agreeing TAGs with other staff members were asked about any difficulties they faced
- **Internal quality assurance:** respondents who were involved in the internal quality assurance process were asked further details about this
- **Final thoughts on submitted TAGs following internal quality assurance:** respondents who saw the submitted TAGs were then asked whether there had been any changes to the TAGs, and why, as well as how fair they thought the submitted TAGs were

- **Centre declaration responsibility:** respondents who had signed the centre declaration document were asked about the fairness of TAGs
- **Final thoughts:** all respondents were asked about the time they spent working on TAGs and how well prepared they felt students are for their next steps

The student survey was also divided into several major sections, which were also presented or skipped based on answers to routing questions. There were the following main sections:

- **Demographic details:** all respondents were asked about their centre, school or college year, and subjects studied
- **Initial thoughts and feelings:** all respondents were asked how they felt about the exam cancellation notice at the time
- **Disruption:** all respondents were asked about the level of disruption they faced to their learning due to the COVID-19 pandemic
- **Details on TAG Evidence:** all respondents were asked a range of questions about the evidence used in determining their TAGs, as well as their level of awareness and engagement in completing and selecting these pieces of evidence
- **Your views on fairness of TAGs:** all respondents were asked their views on the fairness of TAGs and how this relates to normal assessment years
- **Reasonable adjustments:** respondents who indicated that they received reasonable adjustments were asked further details about these
- **Balance between carrying out assessments to support TAGs and further learning of new content:** all respondents were asked about this balance and whether they thought it was appropriate
- **The future:** all respondents were asked about their preparedness for their next steps
- **Final thoughts:** all respondents were asked their thoughts now TAGs were completed and for anything else they wished to share

For questions where keyboard input was required, some basic data validity filtering was undertaken. For example, typographical errors when listing subjects were fixed and implausible answers to numerical questions were omitted, for example stating more than 50 years' experience in the teaching profession.

In this report we describe responses to all questions, but cover some in more detail than others to keep the length of the report reasonable. We also carried out analysis of responses by sub-group, and report differences between, for example, teaching

role, centre type or student year group where we found substantial and relevant differences, not for every question.

Respondents and geographical coverage

Teaching staff who had been involved in any part of the Teacher Assessed Grades process, and students who were receiving qualifications using a TAG, were invited to respond to the relevant survey through a series of announcements via a range of communication channels including social media advertisements and dissemination by teaching unions and associations. The surveys continued to be publicised while they were open and we monitored the response rate.

Ofqual regulates qualifications in England. The aim of the surveys was therefore to gain a picture of the experiences of teaching staff and students in England only, and this was made clear to respondents. However, we did not collect geographical information on the survey and so it is possible that there may have been some responses from other parts of the United Kingdom.

Information provided to respondents

Having followed the links in the announcements, potential respondents saw an information screen (Annex A for the teaching staff survey and Annex B for the student survey) detailing the purpose of the survey, who it was intended for and the specifics of data handling, to help them decide whether the survey was relevant to them and they wanted to complete it. They entered the full survey when they confirmed that their centre had completed the submission of Teacher Assessed Grades to awarding organisations (or, if they were a student, that they were receiving one or more qualifications through TAGs) and that they wanted to continue to the full survey. If they did not confirm this, the survey ended.

Following completion of the survey, respondents indicated if they would like to be considered for a follow-up interview to explore their experience in more depth. This strand of the research is [detailed in a separate report](#).

TAG teaching staff survey results

In total, 1,785 teaching staff completed the TAG teacher survey. Partially completed responses were not saved by the survey platform.

Nearly all questions were optional, except the key routing questions that were used to determine which sections each respondent saw based upon their involvement in different parts of the TAG judgement process. Since we did not force an answer to be entered for the non-routing questions, respondents were free to not answer, or to answer only partially, and so the number of responses varies across questions. We state the total number of responses for each question as (N = xxx) and where appropriate, the number of responses for options provided in questions as (n = xxx).

We present the results in sections relating to different aspects of the process with individuals answering sections depending on their involvement.

Demographic details

All respondents to the survey were presented with the demographics section – though some questions were optional.

Q. Which of the following options best describe your centre?

Respondents indicated their centre type, and the counts and percentage of the sample are shown in Table 1.

Category	Count	Percentage
Secondary comprehensive	617	35
Academy	434	24
Independent school (including city technology colleges)	341	19
Secondary selective, for example, grammar or technical	139	8
Further education establishment	76	4
Sixth form college	74	4
Alternative provision or pupil referral unit	22	1
Free school	18	1

Category	Count	Percentage
Secondary modern	16	1
Special school	14	1
Training provider	6	less than 1
University technical college	6	less than 1
International school	5	less than 1
Tertiary college	4	less than 1
Other	13	1

Table 1: Number of respondents by centre type (N = 1,785)

For some questions that follow, we provide an analysis by centre type. To provide suitably robust statistics, some centre types are combined, giving the following analysis categories:

- Secondary (n = 633) – secondary comprehensive, secondary modern
- Academy (n = 434)
- Independent (n = 341) – independent school (including city technology colleges)
- College (n = 154) – sixth form college, further education establishment, tertiary college
- Secondary selective (n = 139)

Q. Which of the following options best describe your role?

Respondents indicated their role, and the counts and percentage of the sample are shown in Table 2.

Category	Count	Percentage
Head of department	712	40
Teacher, tutor, or trainer	519	29
Deputy or assistant head of centre	180	10
Deputy or assistant head of department	107	6
Exams officer	63	4
Head of centre	62	3

Category	Count	Percentage
Other senior leadership team role	58	3
Key stage leader	44	2
Pastoral leader or head of year	13	1
Data or quality manager or analyst	13	1
Special educational needs co-ordinator (SENCo) or other special educational needs (SEN) specialist	10	1
Other	4	less than 1

Table 2: Number of respondents by role type (N = 1,785)

Q. How many years have you held a position in your current centre?

The median length of time the respondents (N = 1,782) had held a position in their current centre was 5 years, with half of all respondents being at their current centre for between 3 and 10 years. These results suggest that we have a sample of respondents who are mostly well established within their centre.

Q. How many years have you been in the teaching profession?

The median number of years the respondents (N = 1,782) had been in the teaching profession was 16 years, with half of all respondents having been in the profession for between 10 and 22 years. We had a significant number of staff who were well established in the profession and, by comparison, we had relatively small numbers of respondents who were new to the field. Indeed, 8% (n = 142) of respondents had 30 or more years' experience whilst only 2% (n = 28) had 2 years or fewer. We note that this bias may be related to the substantial number of respondents who held various senior roles.

Judging TAGs

Q. Did you judge TAGs for individual students you taught directly?

All survey respondents (N = 1,785) were asked if they had judged TAGs for individual students as a compulsory routing question for this section. Those who answered yes were then asked the questions that follow in this section. Those who answered no were routed to the next section.

Eighty-six per cent of respondents indicated that they had judged the grades of students that they taught directly. Of the 14% who did not judge grades themselves, only 18% listed their role as a teacher, tutor, or trainer.

Q. Which general qualifications were you involved in judging TAGs for (excluding those for which you were only involved in Internal Quality Assurance activities)?

Respondents were asked to indicate which GQ subjects they were involved in judging TAGs for. A total of 1,522 respondents judged TAGs for at least one GQ subject. The 1,977 responses for GCSEs are shown in Figure 1 and the 1,169 responses for AS or A levels in Figure 2. Note that, as respondents were able to select more than one subject, the total number of responses is higher than the number of respondents.

The higher project qualification (HPQ) is included with GCSE subjects in Figure 1 and extended project qualification (EPQ) and core maths (Level 3) are included alongside the AS or A level subjects in Figure 2.

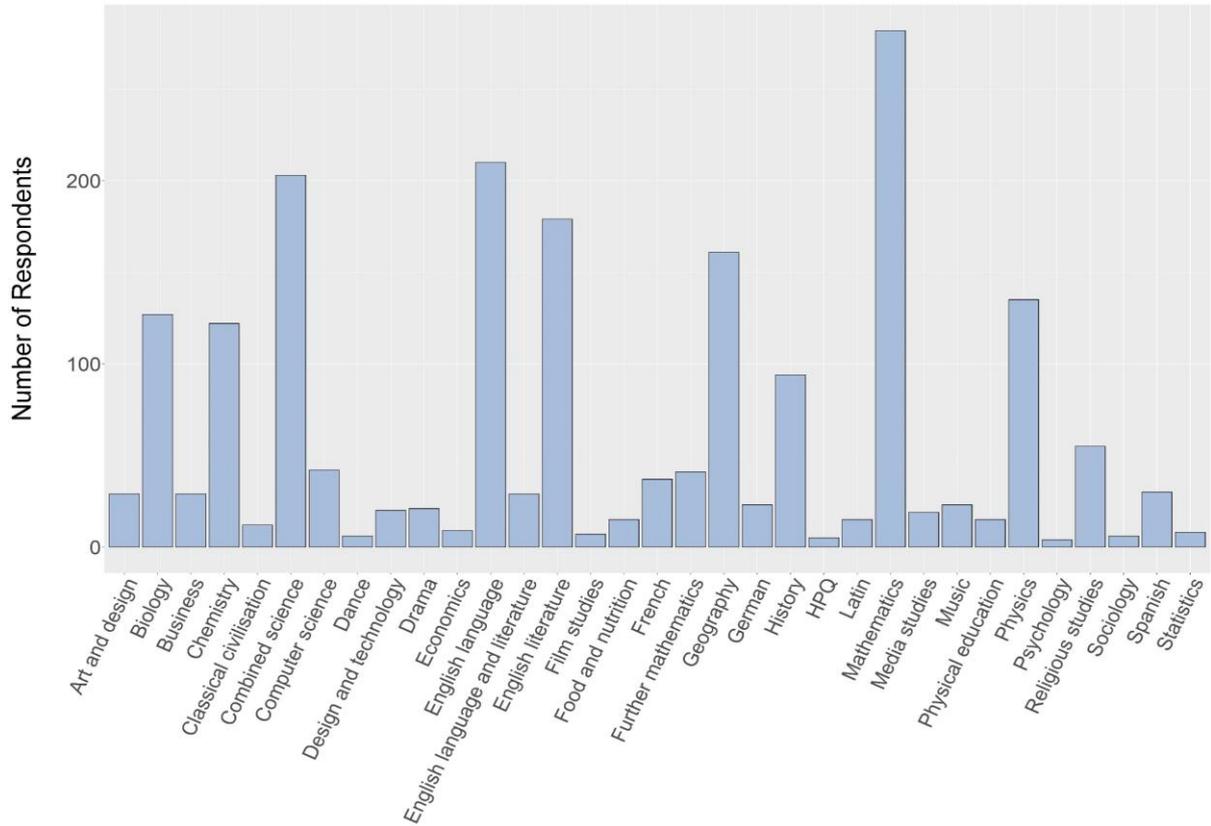


Figure 1: Number of respondents by GCSE subject (n = 1,522)

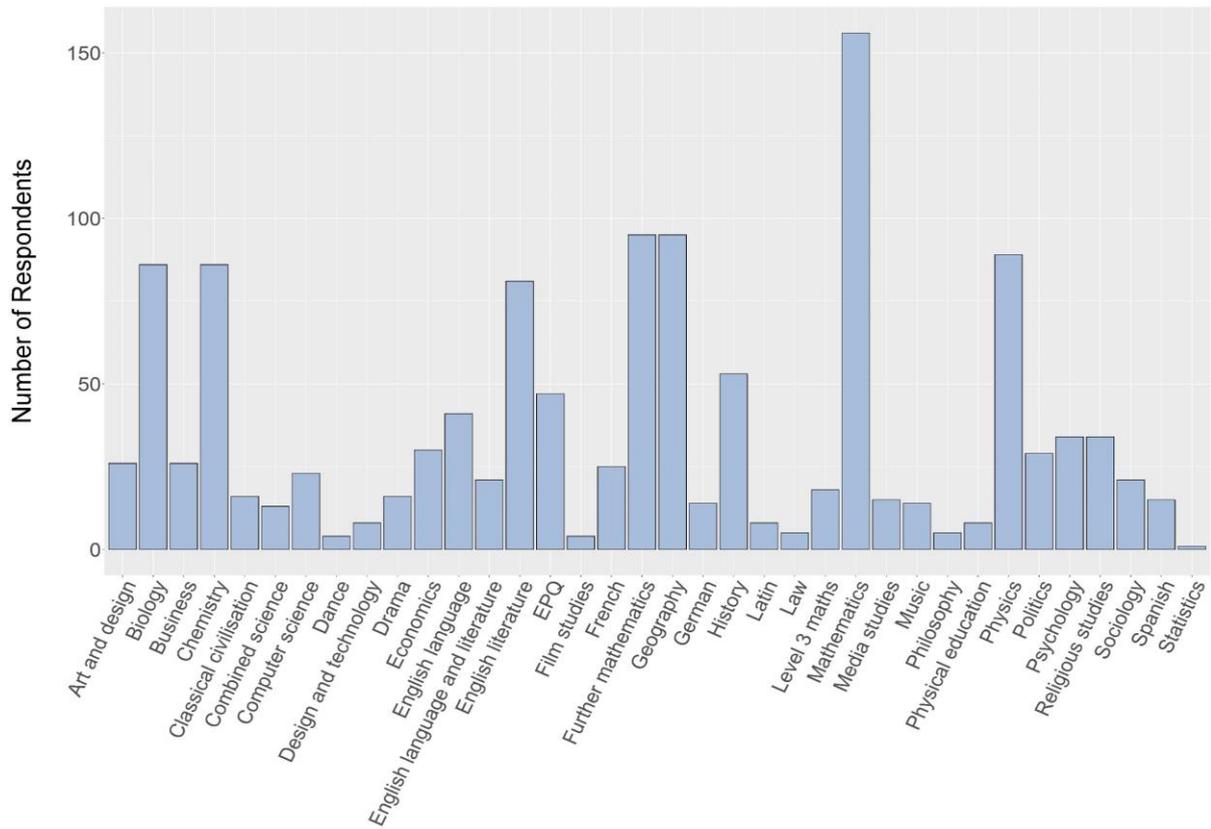


Figure 2: Number of respondents by AS/A level subject (n = 1,522)

As with the 2020 CAG survey, mathematics was the most common subject listed for both GCSE and AS or A level. For GCSE, maths was followed by English (including English language, English literature, and combined English language and literature – while not offered as a GCSE, due to the grid layout this option was likely selected where staff produced TAGs for both language and literature) and the sciences (combined sciences, physics, chemistry, and biology). Geography and history both also feature prominently. The picture is broadly similar at AS or A level with the additional appearance of further mathematics in the most common subjects for which TAGs were generated.

Q. For which types of vocational and technical qualifications did you judge TAGs for this summer, if any?

Respondents indicated which VTQ types they judged TAGs for. Because of the large number of qualifications available we did not ask for specific qualification titles. A total of 164 respondents indicated that they judged TAGs for at least one VTQ type, providing a total of 218 responses shown in Figure 3.

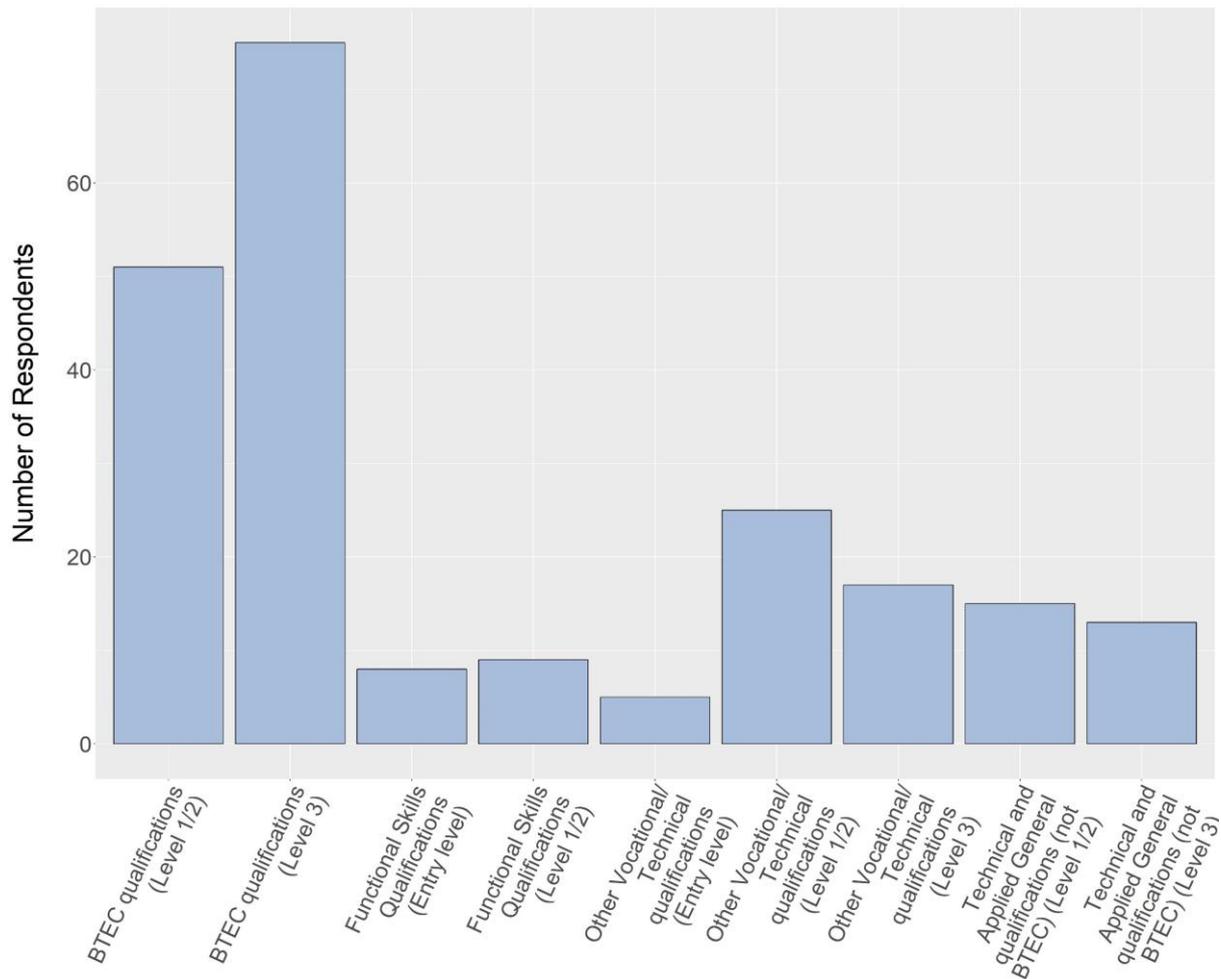


Figure 3: Number of respondents by VTQ type (N = 164)

The most common VTQ qualification response was BTEC Level 3 (34%, n = 75) followed by BTEC Level 2 (23%, n = 51). In total, Level 3 qualifications were the most frequent response (48%, n = 105), followed by Level 1/2 (46%, n = 100) and entry level (6%, n = 13). The distribution of VTQ qualification responses is in-line with that from the 2020 CAG survey.

Q. How many classes did you judge TAGs for (excluding those for which you were only involved in Internal Quality Assurance activities)?

Respondents entered the number of classes for which they judged TAGs (N = 1,494). The median number of classes was 3, with half of all respondents judging TAGs for between 2 and 5 classes. Thirteen (1%) respondents indicated that they judged 25 classes or more. These were senior staff that we presume counted classes that they had not directly taught but instead had been involved in generating or moderating TAGs for.

Q. How many individual students in total did you judge TAGs for (excluding those for which you were only involved in Internal Quality Assurance activities)?

Respondents were asked to enter the number of students for which they judged TAGs (N = 1,524). The median number of students was 50, with half of all respondents judging TAGs for between 28 and 90 students. There were 36 (2%) respondents who stated that they judged TAGs for 300 students or more. As with the previous question, most of these outliers were senior staff such as heads of departments.

Considerations for judging TAGs

Q. Were you involved in any of the following: (i) meetings or discussions, or (ii) training or sharing of information with colleagues, about how to make TAG judgements?

All survey respondents (N = 1,785) were asked whether they were involved in meetings or discussions about how to make TAG judgments or involved in any training or sharing of information with colleagues about how to make TAG judgments as a compulsory routing question for this section. Those who answered yes were then asked the questions presented in this section, while those who answered no were routed to the next section.

Ninety-five per cent of respondents indicated that they were involved in TAG meetings, discussions, training, or information sharing. Of the 81 respondents (5%) who indicated they were not involved in any of these activities, 57% were teachers, tutors, or trainers, 21% were exams officers, and 11% were Heads of Department.

Q. Did you consider/look at any of the following in any meetings/discussions before or during the judging of TAGs?

Respondents were asked to select either yes or no for several items to indicate whether they had been considered during the meetings/discussion before or during the judging of TAGs (N = 1,704). The percentage selecting yes for each option is given in brackets:

- Guidance on the process from relevant awarding organisations or exam boards (97%)
- Grade descriptors (94%)
- Previous years' student outcomes (2019 or earlier) (91%)
- Grade worthiness standard (90%)
- How to select evidence based only on content taught to students (90%)
- Previous years' candidate work (63%)

Except for 'previous years' candidate work', all items received a positive response of 90% or higher, showing that they were extremely well-used. The positive response percentage for a candidate's previous years' work was significantly lower, though the majority (63%) still indicated that it was used. This trend was also observed in the 2020 CAG survey.

Q. Was the Centre Policy document used as part of the training and planning?

Respondents were asked if the centre policy document (required for GQs and some VTQs) was used as part of the training and planning process (N = 1,698).

Respondents could answer yes, no, or not sure. The vast majority (87%) of respondents indicated that the Centre Policy document was used as part of the training and planning of the TAGs process. There was no substantial variation to this result based on centre type. The 10% of respondents who were not sure may comprise those who did not see the centre policy directly and were not sure whether the training for the TAG process came from this document or not, or staff delivering only VTQs, where the document was not always required.

Q. Were you aware of any steps taken by your centre to protect against unconscious bias in the whole process?

Respondents were asked if they were aware of any steps taken by their centre to protect against unconscious bias. They were asked to select from one of 4 options:

- Yes – there were effective steps taken
- Yes – there were some partially effective steps taken
- Not sure – there may have been some steps taken but I was not aware
- No – no steps taken

The percentage of respondents giving each response is shown in Figure 4.

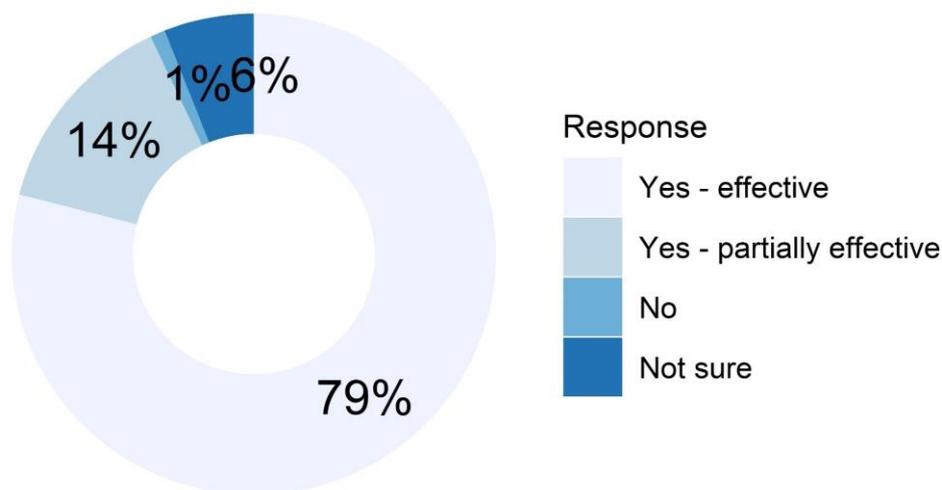


Figure 4: Percentage share of respondents for steps against bias (N = 1,700)

In total, 93% of respondents indicated at least partially effective steps were taken by their centre to protect against unconscious bias, with 79% of all respondents indicating that these were effective. Only 1% of respondents indicated that no steps were taken to protect against unconscious bias.

In comparison with the 2020 CAG survey, this is a marked increase. Particularly, the percentage of respondents who indicated that effective steps were taken increased from 68% to 79%. This increase appears to have come directly from the drop in respondents who were not sure if steps had been taken – down from 17% to 6%.

Q. Did you/your centre look at previous years' data to reflect on potential systematic under- or over-grading (e.g., for different groups of student such as those with protected characteristics), including last year's outcomes?

Respondents were asked if their centre looked at previous years' data to reflect on potential systematic under- or over-grading (N = 1,701). They were asked to select from one of the 5 options, with the percentage selecting each given in brackets:

- Yes – there was a lot of consideration (61%)
- Yes – there was a moderate amount of consideration (25%)
- Yes – there was a little consideration (7%)
- No – there was no consideration (2%)
- Not sure (4%)

Ninety-three per cent of respondents indicated that there was at least a little consideration of previous years' data to reflect on potential systematic under- or over-grading with most of these indicating that there was a lot of consideration (61%). This result is slightly down on the 2020 CAG survey, where 76% indicated that there was a lot of consideration of previous years' data. Only 2% of respondents indicated that there was no consideration of previous years' data.

Q. Were you aware of the Ofqual guidance about making objective judgements?

Respondents were asked whether they were aware of the Ofqual guidance about making objective judgements (N = 1,701). The overwhelming majority of respondents (96%) indicated that they were aware of it. This result is a slight increase over last year, up from 90% in the 2020 CAG survey.

Q. Was the Ofqual guidance about making objective judgements useful?

Those respondents who indicated that they were aware of the Ofqual guidance about making objective judgements were then asked if they found this guidance useful. They were asked to select one of the following 4 options:

- Yes
- No
- Not applicable – we didn't use this guidance
- Not sure

The percentage share of these responses is shown in Figure 5.

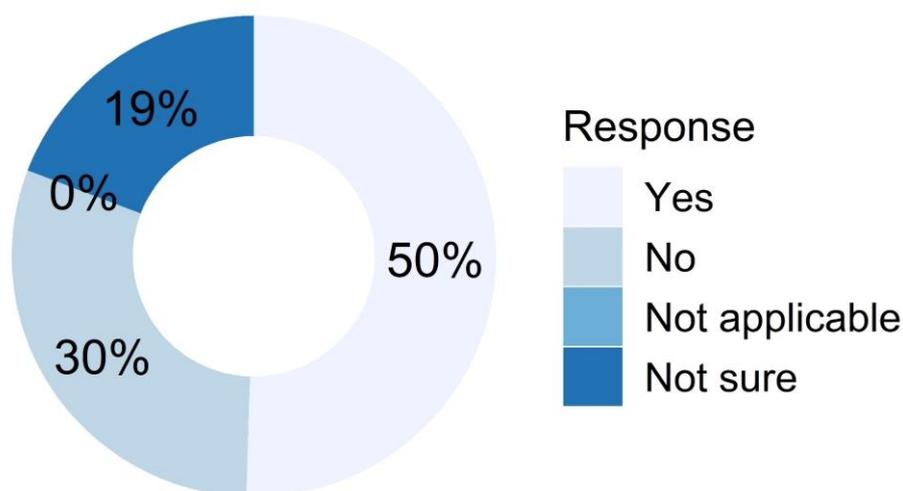


Figure 5: Percentage of respondents who found the Ofqual objective guidance useful (N = 1,625)

Almost all respondents who were aware of the Ofqual guidance indicated that it was used (over 99%). Fifty per cent indicated that the guidance was useful, 30% indicated that it was not useful, and 19% were not sure. This represents a significant decrease in usefulness compared to the 2020 CAG survey, where 84% of respondents indicated that the guidance was useful. We note that in the CAG survey, the option of 'not sure' was not provided. However, even after accounting for this, the proportion who found it useful is still much lower (62%) than last year (84%). It is worth noting that this guidance was not new in 2021, and many staff may have used it the previous year, therefore possibly reducing ratings of its usefulness.

Q. Were you aware of any guidance about making objective judgements provided by the awarding organisation/exam board?

Respondents were asked whether they were aware of the guidance about making objective judgements provided by awarding organisations (N = 1,698). A large majority of respondents (78%) indicated that they were aware of the guidance. This is an increase over the 2020 CAG survey (66%); however, it is significantly lower than those who were aware of Ofqual guidance (78% vs 96%).

Q. Was this guidance from the awarding organisation/exam board about making objective judgments useful?

Similar to the questions regarding the Ofqual guidance (Figure 5), those respondents who indicated that they were aware of the awarding organisation guidance about making objective judgements were then asked if they found this guidance useful. They were asked to select one of the following four options:

- Yes
- No
- Not applicable – we didn't use this guidance
- Not sure

The percentage share of these responses is shown in Figure 6.

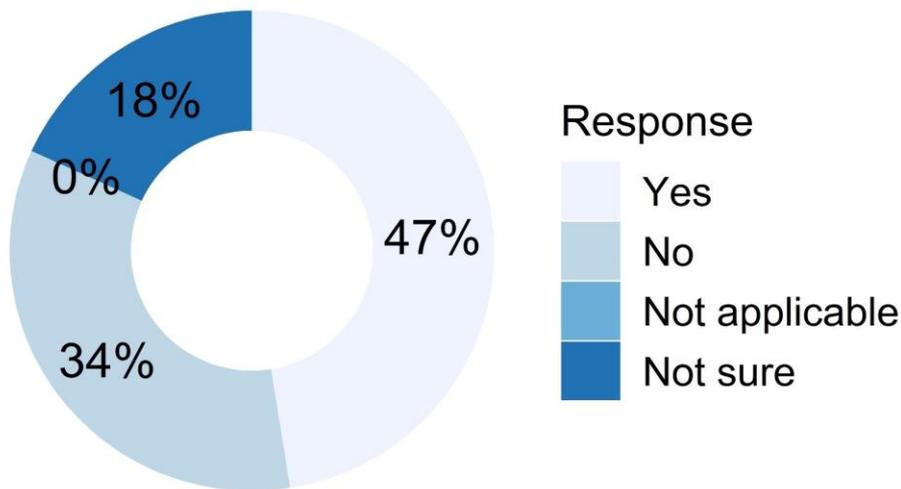


Figure 6: Percentage of respondents who found the awarding organisation objective guidance useful (N = 1,320)

The distribution of answers indicate that respondents found the awarding organisation guidance on making objective judgements to be of similar, though slightly less, use than the Ofqual guidance. Forty-seven per cent of respondents indicated that the awarding organisation guidance was useful, 34% indicated that it was not, and 18% were not sure.

Again, as with the Ofqual guidance (Figure 5), this result is markedly different to the 2020 CAG survey where 85% of respondents who were aware of the guidance found it useful – even when accounting for the lack of a ‘not sure’ option last year.

Q. Were any of the following included as part of discussion/training around making objective judgements and the avoidance of bias?

Respondents were asked to select from several options that may have been included in the discussions and training they were involved in around making objective judgements and the avoidance of bias (N =1,371). The following options were presented, with the percentage selecting each given in brackets:

- Staff training on bias (70%)
- Input from a special educational needs coordinator (SENCo) (57%)
- Other resources (for example, online) not provided by Ofqual or the exam board/awarding organisation (29%)

- Input from other specialists on diversity or reasonable adjustments (23%)
- Input from special educational needs and disability (SEND) experts (20%)
- Academic research (12%)

The 2 most frequently reported inputs were 'Staff training on bias' (70%) and 'Input from a Special Educational Needs Co-ordinator' (57%). Except for 'academic research' (12%), all other items were used by between 20 to 30% of respondents.

In comparison with the 2020 CAG survey, we see a doubling in the use of staff training (up from 35% in 2020 to 70% in 2021) and a halving of the use of academic research (down from 23% to 12%). Other inputs are broadly similar across the 2 surveys.

Q. Would more information/resources have been useful when considering the issue of making objective judgements and the avoidance of bias?

Respondents were asked if more information or resources would have been useful. (N = 1,684). The most common response to this question was 'no' (46%). Thirty-four per cent indicated that more information and/or resources would have been useful whilst 20% were not sure.

In the 2020 CAG survey, there was no 'not sure' option for this question. Therefore, to draw a direct comparison, we removed these responses from this year's data and found that 43% of the remaining respondents answered yes (57% answered no). This is very similar to the CAG survey (41% yes and 59% no).

Accompanying this question was a free-text field allowing respondents to add additional comments to their answer. We analysed these 283 responses and collated them into a range of different categories. Whilst most respondents provided additional comments on what information and resources could have proved useful, others used this opportunity to highlight what worked well or not so well at their centre and others expressed their feelings on the TAGs process more generally.

The most common response (29%, n = 92) was that the respondents had issues with the broader guidance that was provided by, for example, the Department for Education, Ofqual, AOs, and JCQ. Notably this related mostly to timing (for instance, guidance was released too late), vagueness (for instance, guidance was not subject specific), or inconsistencies across different external bodies.

Comments stating that awarding organisations should have done more were the next most frequent response (19%, n = 53). Particularly, it was felt that the GQ exam boards should have provided better assessment and marking materials, for example,

more detailed grade descriptors or new, unseen exams. Whilst this does not seem directly related to bias, and is perhaps indicative of wider frustrations with AOs, perhaps a perception that better materials may have reduced the amount of teacher judgement required and therefore the potential for bias. However, this is not clear on the basis of this survey.

Fifteen per cent (n = 42) of respondents commented on the difficulties teachers faced in assimilating the guidance given to them. They particularly highlighted the idea of information overload, commenting that there were too many sources of information producing too much guidance, and workload difficulties (for example, balancing the reading of guidance whilst judging TAGs and continuing to teach their students during the pandemic). Again, this may not directly relate to reducing or mitigating bias.

Thirty-seven respondents commented on certain pieces of evidence or methods used at their centre that they felt either helped to reduce bias (10%, n = 29) or introduced, or increased, bias (3%, n = 8). Examples given that respondents felt helped to reduce bias included relying purely on objective data collected via highly controlled assessments and using techniques such as blind marking. Others discussed using moderators (both internal and external) or swapping assessment marking with other centres. Examples that respondents felt potentially increased or introduced bias included not using techniques such as blind marking or moderation. Additionally, some respondents felt that allowing members of the senior leadership team to alter TAGs, despite having used rigorous methods to arrive at those grades, could have introduced bias.

Eight respondents (3%) felt that exams should have gone ahead or that external marking of assessments should have been undertaken. Some respondents noted that it was impossible to remove all bias, particularly with teacher assessment, (6%, n = 17) whilst others felt that teachers did not need training on bias and that the guidance produced was insulting and patronising (4%, n = 11). Similar sentiments were observed in the 2020 CAG survey. Views about how the potential for bias was, or could be, mitigated in the TAG process were clearly complex and varied.

Judging individual student TAGs

This section was only completed by respondents who had indicated that they judged individual student TAGs (see page 17). Respondents were asked on the survey page to think about a single qualification and subject for which they judged TAGs - the qualification and subject with the most students or the qualification and subject they thought was most representative of their overall experience. While most questions were asked of respondents for both GQs and VTQs, a few questions were only presented to those determining TAGs for GQs. This is indicated for the relevant questions.

Q. Please tell us the category of the qualification you will be telling us about judging TAGs for

Respondents (N = 1,538) were asked what type of qualification they were thinking of, either General Qualification (GQ) or Vocational or Technical Qualification (VTQ).

The vast majority (97%) of respondents selected judging TAGs for GQs. Although 164 respondents indicated that they assessed TAGs for vocational or technical qualifications (Figure 3), only 52 (3%) respondents chose a VTQ qualification to answer further questions on. This is likely due to the majority of VTQ responses coming from centres that offer these alongside GQ qualifications.

Q. Please type in the qualification and subject for which you will be answering the questions in this section

Respondents were asked to type in the qualification and subject they were answering about in a free text field. Most GQ respondents (69%) indicated that they were thinking about a GCSE subject. This was followed by A level (including AS) at 25%, EPQ (less than 1%), and Pre-U (less than 1%). Four per cent of respondents did not provide a qualification type or level in the response.

The qualification types given by VTQ respondents were varied, however, numbers were very low for most types. The 3 most common responses were BTEC Level 3 (23%), BTEC Level 1/2 Tech Award (21%), and BTEC Level 2 (9%).

We used the most frequent subjects or subject areas to analyse subject-level differences for some of the questions that follow. The subject areas for GQ, together with the number of respondents for each were: the sciences (n = 337), maths (n = 317), English (n = 234), geography (n = 174), history (n = 99), modern foreign languages (n = 67), religious studies (n = 59), computer science (n = 37), art-related subjects (n = 30), music (n = 23), drama (n = 22), psychology (n = 21), and business studies (n = 20).

As with qualification type, the subjects provided by VTQ respondents were highly varied, and most had a very low count. Health and social care was the only subject area with more than 4 respondents (n = 7). We did not analyse VTQ responses by subject area.

Q. For your selected qualification, approximately how much of the course do you think you were able to teach?

Respondents were asked to enter how much of the course they felt they were able to teach as a percentage. Answers were restricted to integers ranging from 0 to 100 and are grouped into bins of 5% width. The result from GQ respondents is shown in Figure 7 and from VTQ respondents in Figure 8.

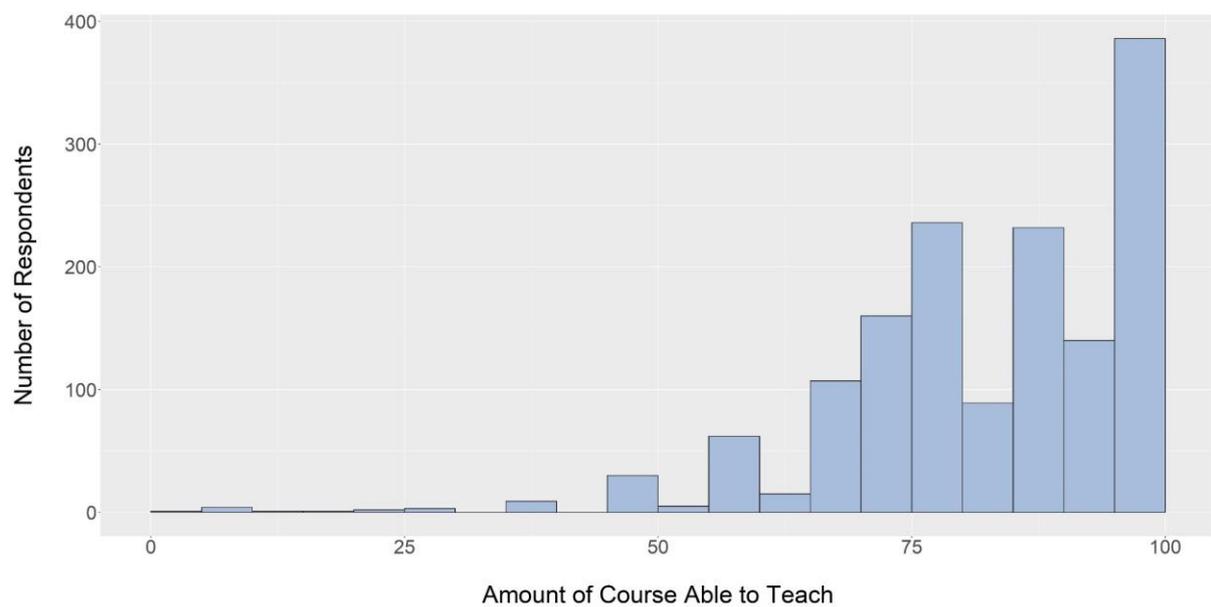


Figure 7: Number of GQ respondents by percentage of course they were able to teach (N = 1,483)

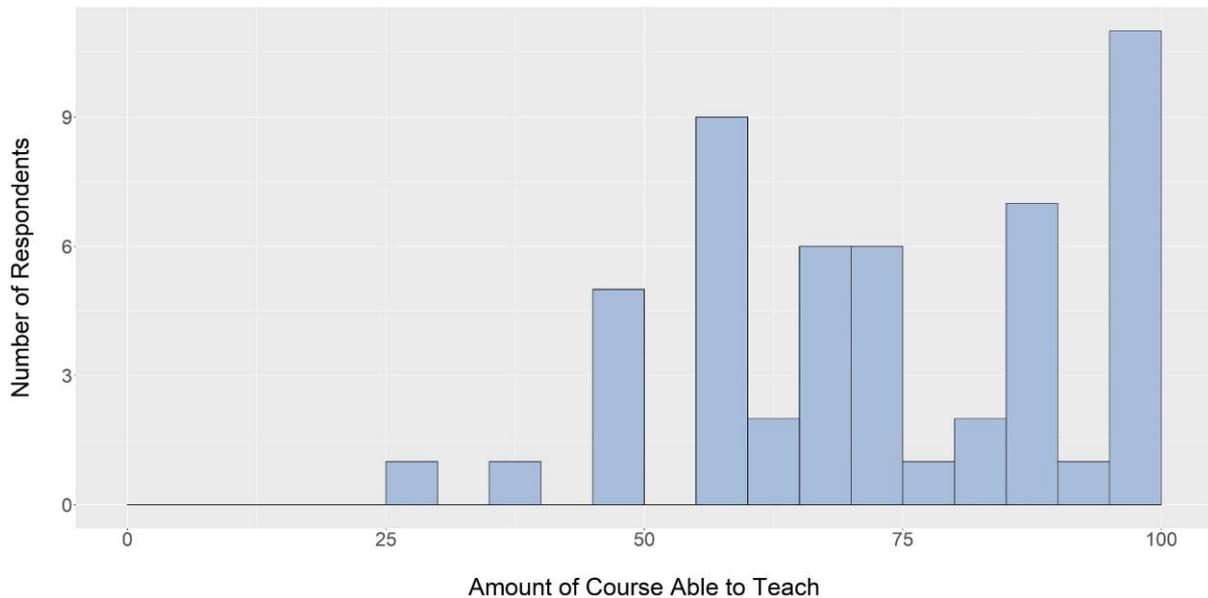


Figure 8: Number of VTQ respondents by percentage of course they were able to teach (N = 52).

For GQ respondents (Figure 7), the median result was 90% of the course being taught, with half of all respondents indicating that between 75% and 100% of the course was taught. Three per cent of respondents indicated that they thought they had only been able to teach half of the course or less.

For VTQ respondents (Figure 8), the median result was 75% of the course being taught, with half of all respondents indicating that between 60% and 90% of the course was taught. Fourteen per cent of respondents indicated that they had only been able to teach 50% or less of the course. The VTQ respondents clearly tended to indicate that they were able to teach less of their course than GQ respondents. For this and other questions in this section, it is worth remembering that the VTQ respondent sample was quite small.

For GQ respondents, we were able to undertake a subject analysis. We found that, generally, the percentage of content taught was similar across all subjects, with medians ranging from 75% to 90%. The one exception to this was the art-related subjects, where the median percentage of content taught was 70% and the upper quartile (80%) was below the median across all subjects (90%). We note, however, that this was based on a relatively small number of responses (n = 30).

We also investigated the effect of centre type. We found that respondents from independent schools and secondary selective schools tended to be able to teach slightly more of the content than their counterparts – especially those from secondary and academy schools (90% versus 85%). We note, however, that the median amount of course respondents were able to teach is relatively high across all centre types.

Q. What proportion of the evidence you used, on average, came after the announcement in January that summer assessments would be cancelled and that TAGs would be required?

Respondents (N = 1,476 for GQ and N = 52 for VTQ) were asked to identify what proportion of evidence used came from after the January announcement that exams would be cancelled. The following options were presented, with the percentages of GQ and VTQ respondents selecting each given in brackets:

- All of it (GQ: 25%; VTQ: 12%)
- Most of it (GQ: 53%; VTQ: 35%)
- Around half of it (GQ: 19%; VTQ: 31%)
- A little of it (GQ: 3%; VTQ: 23%)
- None of it (GQ: 0%; VTQ: 0%)

For GQ, over three quarters of respondents indicated that most or all the evidence used to determine TAGs was generated after the January announcement that summer assessments were cancelled (combined total of 78%). Nineteen per cent indicated that around half of it was generated before the announcement. In total 4% indicated that only a little or none of the evidence was generated after the announcement.

Two GQ subjects contrasted with this overall pattern, namely the art-related subjects and drama. In both subjects a large proportion of respondents indicated that around half of the evidence used was generated before the announcement. This is consistent with the type of non-exam assessment used in these subjects, some of which would have been in-progress or completed when the announcement came.

For VTQ, just under half of respondents indicated that most or all of the evidence used to determine TAGs was generated after January (combined total of 46%). Thirty-one per cent indicated that around half of it was generated before the announcement. Twenty-three per cent indicated that only a little of the evidence was generated after the announcement. No respondents indicated that none of the evidence was generated after it.

The percentages show that the VTQ respondents generally used evidence to support TAG judgments generated earlier than their counterparts in GQ, who clearly indicated they mostly generated the evidence after the January announcement. This is likely due to VTQ subjects incorporating a significant quantity of coursework and

also having a modular structure meaning that many assessments had taken place earlier in the course. However, within GQ subjects there was still significant variation across respondents of the relative proportion of pre- or post-announcement assessments.

The largest difference between centre types is with independent schools. A larger proportion of respondents from independent schools indicated that all the evidence they used was generated after the January announcement (33% versus 25% across all centre types) and a smaller proportion indicated that around half of it came from before (11% versus 19% across all centre types).

Q. To what extent do you agree or disagree that the evidence you collected to support the TAG judgements also helped to support further student learning?

Respondents (N = 1,482 for GQ and N = 52 for VTQ) were asked to indicate the extent to which they agreed or disagreed with the premise that the evidence they collected to support their TAG judgments also helped support further learning using a 5-point Likert scale ('Strongly agree' to 'Strongly disagree' along with a 'Not sure' option).

For GQs a total of 44% of respondents felt that the evidence collected to support TAG judgements helped support further learning. However, 37% felt that it did not. Twenty-four per cent held a neutral view on this question and neither agreed nor disagreed with the statement. Whilst the largest group therefore agreed that evidence collected for the TAGs supported further learning, there was a mixed picture from respondents.

We determined an 'agreement score' for each subject by assigning a numerical value to each of the response types (ranging from 1 for strongly disagree to 5 for strongly agree). Respondents who were not sure were excluded. Art-related subjects rated most highly (with a mean score 4.0) followed closely by drama (mean of 3.8). Respondents from science scored the lowest agreement value (mean of 2.8) though this is broadly in line with most other subjects.

The picture from VTQ respondents was slightly more positive. Fifty-four per cent of respondents felt that the evidence collected to support TAG judgements helped support further learning. Whereas 27% of respondents felt that it did not. We suggest these differences are because many VTQ qualifications, and the art-related subjects and drama in GQ, include coursework or practical assessment. While the coursework (or performance) is being produced some learning continues to occur.

Q. To what extent did each of the following influence your decision on the amount of evidence you gathered to judge TAGs for your students?

Respondents were asked to rate the extent to which several factors influenced the amount of evidence they collected when determining TAGs, using a 4-point scale. The percentages for each factor for GQ respondents are shown in Table 3 and for VTQ respondents in Table 4.

Factors	A great deal	Quite a lot	A little	Not at all
Provision of a robust evidence base to make judgments	71	23	5	1
Giving students the best opportunity to show what they could do	63	29	7	1
Provision of a robust evidence base in case of appeals	57	27	11	4
Decision by the head of department (or you, if you are head of department)	52	32	12	4
Your centre policy document	45	34	16	5
Decisions by senior leadership	41	34	18	7
The anticipation of external quality assurance	38	34	20	7
The interpretation of what Ofqual or exam boards or awarding organisations wanted	38	41	18	3
Perceived or actual pressure from parents or students	6	9	22	63

Table 3: Percentages of GQ respondents indicating how much influence various factors had on their decision to collect evidence (N = 1,487)

Factors	A great deal	Quite a lot	A little	Not at all
Provision of a robust evidence base to make judgments	63	31	4	2
The anticipation of external quality assurance	62	29	10	0
Giving students the best opportunity to show what they could do	60	23	15	2
The interpretation of what Ofqual or exam boards or awarding organisations wanted	58	25	17	0
Provision of a robust evidence base in case of appeals	50	35	13	2
Your centre policy document	48	37	12	4
Decision by the head of department (or you, if you are head of department)	39	37	18	6
Decisions by senior leadership	31	27	33	10
Perceived or actual pressure from parents or students	2	6	13	79

Table 4: Percentages of VTQ respondents indicating how much influence various factors had on their decision to collect evidence (N = 52)

To simplify the data, a mean score was determined for each factor by assigning a numerical score to each response, ranging from 1 (not at all) to 4 (a great deal). The most influential factor on the amount of evidence to collect, for both GQ and VTQ respondents, was ‘the provision of a robust evidence base to make judgements’ (mean of 3.6 for both). For GQ respondents, this was followed by ‘giving students the best opportunity’ (mean of 3.5) and ‘provision of a robust evidence base in case of appeals’ (mean of 3.4). For VTQ respondents, however, this was followed by the ‘anticipation of external quality assurance’ (mean of 3.5) and then jointly ‘giving students the best opportunity’ and ‘the interpretation of what Ofqual, exam boards, assessment organisations required’ (mean of 3.4). The least important factor for both GQ and VTQ respondents, was ‘perceived or actual pressure from parents or students’ (mean of 1.6 and 1.3 respectively).

The largest variation observed between GQ and VTQ respondents was on the impact of ‘the anticipation of external quality assurance’ where VTQ respondents felt it had a larger impact on their decision (mean of 3.0 for GQ respondents and a mean of 3.5 for VTQ respondents). The other factor on which the averages differed was ‘decisions by senior leadership’. GQ respondents rated this more important than VTQ respondents on average (mean of 3.1 versus 2.8). This result may indicate that

the process for general qualifications was more centrally controlled than for vocational and technical qualifications, likely because of the need for consistency across GQ subjects and the diversity of assessment designs in VTQs.

Q. Were you able to use the same evidence for judging TAGs for all of your students?

Respondents who were answering about GQs were asked whether they used the same evidence for judging TAGs for all their students (N = 1,484). Whilst most respondents (54%) indicated that they were able to use the same evidence for judging TAGs for their students, a significant share of respondents (46%) had to make exceptions for at least one student due to the teaching they had missed.

We found that there was some considerable variability to this question when analysed by subject. Music had the lowest percentage of respondents indicating that they had to make exceptions for one or more of their students (29%). Whilst English and religious studies were the only 2 subjects where the majority of respondents indicated that they had to make exceptions for one or more of their students (59% and 55% respectively). Respondents teaching drama were evenly split on whether they had to make exceptions.

There was also variation by centre type. Over half of the respondents from academies (51%) indicated that they had to make exceptions for one or more of their students. This contrasts with respondents from secondary selective and independent schools, where only 37% and 38% indicated that they had to make exceptions. Respondents from colleges and secondary schools were closer to their counterparts in academies, with 46% and 47% respectively indicating that exceptions were made.

Q. Approximately what percentage of your students did you have to make an exception for?

Respondents for GQs who answered 'no' to the question of being able to use the same evidence for judging TAGs for all students were then asked a follow-up question inviting them to enter the percentage of students they had to make an exception for. The distribution of these responses is shown in Figure 9.

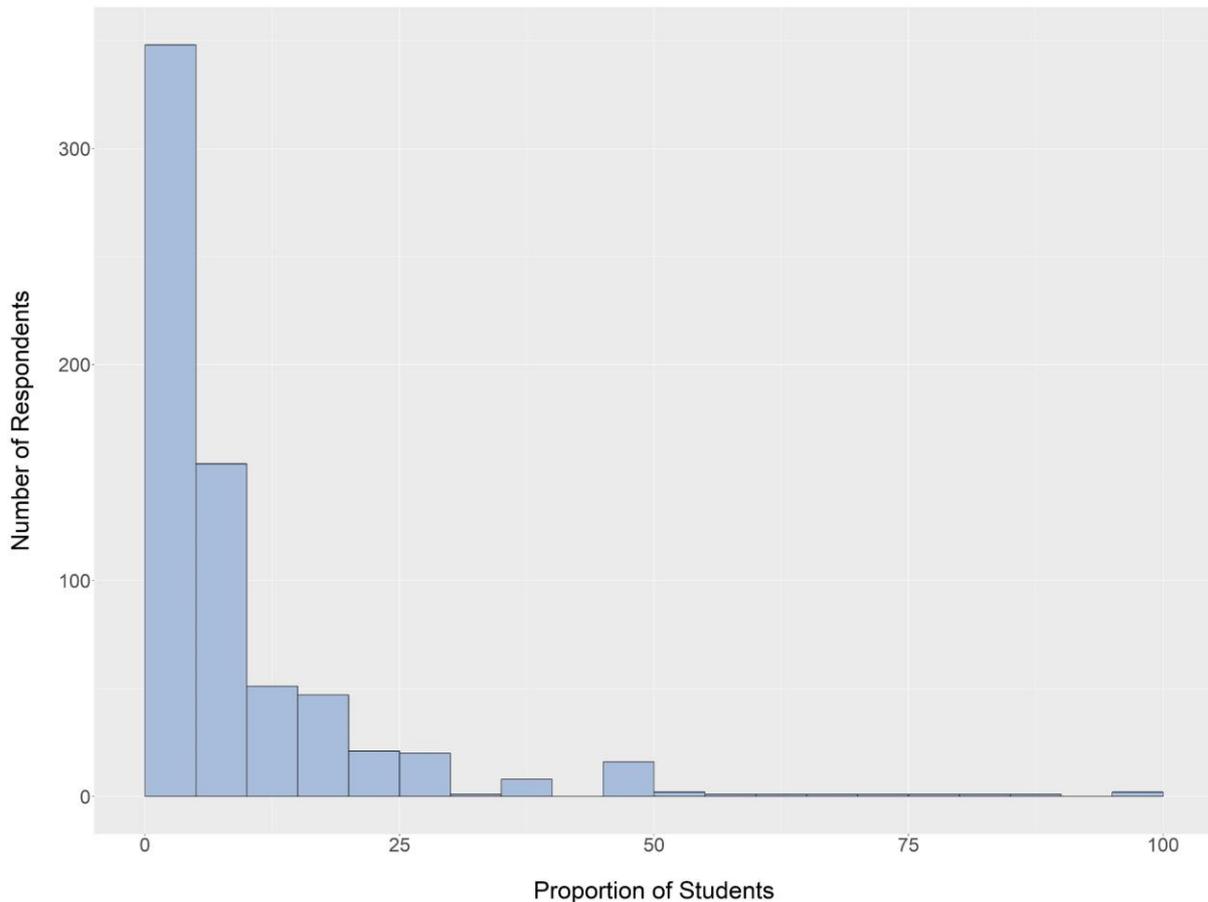


Figure 9: Number of respondents by percentage of students exceptions were made for (N = 677)

The median percentage of students these respondents had to make an exception for was 5%, with half of all respondents indicating that exceptions were made for between 4% and 15% of students. Fewer than 1 in 10 (8%) respondents indicated that they had to make an exception for more than 25% of their students.

This result did vary by subject, with the median result from respondents from art-related subjects, computer science, drama, English, music, and psychology being that they had to make exceptions for at least 10% of their students.

Respondents from colleges and independent schools indicated that, when they did have to make exceptions, they tended to have to make them for a larger proportion of their students (a median of 10%) than the other centre types (median 5%).

Q. With regard to your TAG judgements, how influential were each of the following forms of evidence?

Respondents were asked to rate how influential various forms of evidence listed on screen were in determining TAGs. They were asked to enter a rating from 0 to 100, with 0 indicating evidence that was not available, given no weight, or had no influence on their judgements and 100 indicating evidence that was given the highest weighting or had the most influence on their judgement.

The mean rating given to each form of evidence is shown in Figure 10 for GQ respondents and Figure 11 for VTQ. It should be noted that respondents may have varied in their interpretation of the scale and so the meaning of specific absolute values is not entirely clear and instead only the relative weighting of the individual forms of evidence should be used for comparison.

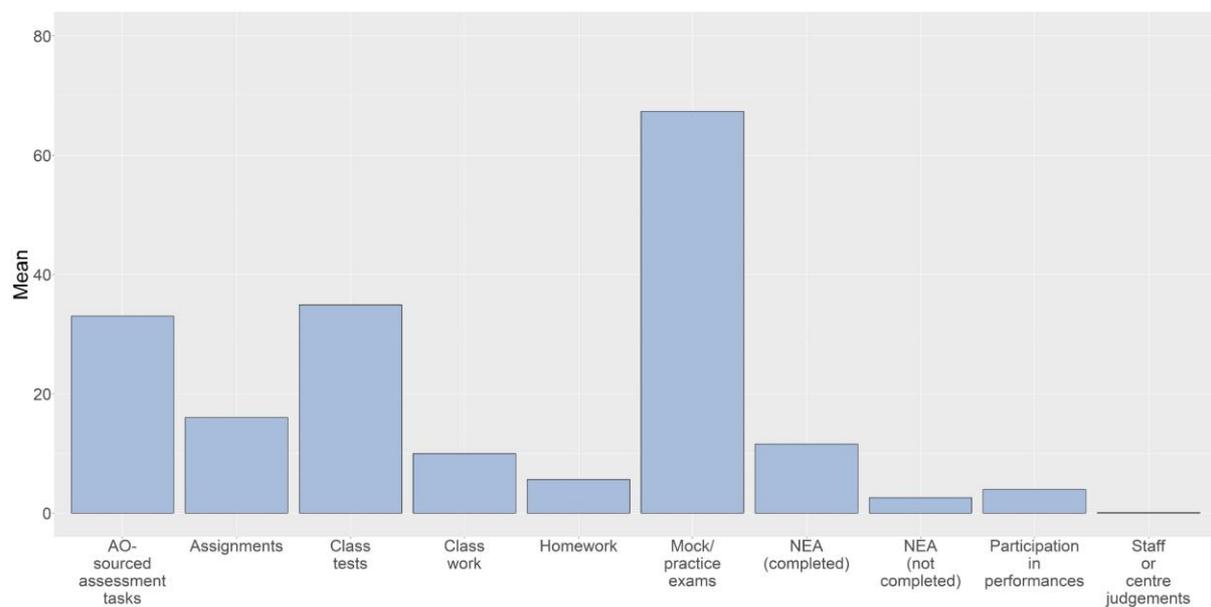


Figure 10: Mean influence rating for evidence types for GQ respondents (N = 1,483)

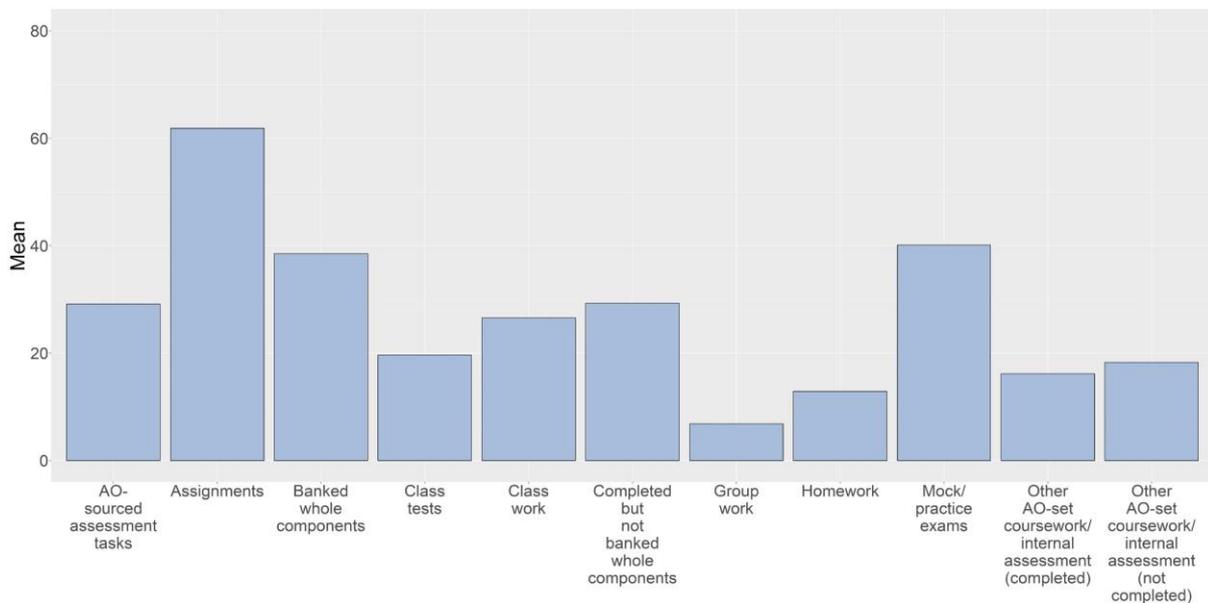


Figure 11: Mean influence rating for evidence types for VTQ respondents (N = 52).

Respondents were able to enter other sources of evidence in addition to those we listed. Two common ‘other’ responses were past exam paper(s) and custom paper(s) created using a selection of previous exam questions from different years. These 2 added evidence types may highlight a lack of clarity in our pre-defined options, particularly in the distinction between ‘AO-sourced assessment tasks’ and ‘Mock or practice exams – part or whole past papers not produced by exam boards to support TAGs’ in the GQ context. This ambiguity may have slightly distorted ratings for these evidence types. This is considered further in the Discussion.

For GQs, the most influential type of evidence used was mock or practice exams (67/100), followed by class tests (35/100), and AO-sourced assessment tests (33/100). All other forms of evidence were rated below 20/100.

For VTQs, the picture is more mixed. Assignments were the most influential type of evidence used (64/100), followed by banked whole components (51/100). ‘Other AO-set coursework or internal assessment (completed)’, ‘Mock or practice exams’, ‘Completed but not banked whole components’, and ‘AO-sourced assessment tasks’ were all closely ranked in influence at 43/100, 43/100, 41/100, and 36/100 respectively. The broadness of the evidence base for VTQs was also seen in the CAG 2020 survey and is reflective of the diversity of assessment approaches in VTQs.

When we analysed the GQ respondents by subject, we found that most subjects showed a similar trend - the dominant forms of evidence used being mock or practice exams, AO-sourced assessment tasks, and class tests. However, as might be expected, the more expressive subjects showed substantial deviations from this trend. In art-related subjects, class work was far more influential than class tests, and non-exam assessment (NEA) was more influential than mock or practice exams.

Similarly, in both music and drama, NEA and participation in performances were rated as being much more influential than for other subjects, though mock or practice exams remained an influential type of evidence.

The trend for each evidence type was broadly similar for all centre types. The only minor exception was respondents from colleges, where we saw a general increase in the importance of class work, class tests, and homework and a decrease in the importance of the mock or practice tests – perhaps reflecting the broader range of qualification types colleges determined TAGs for.

Q. How many individual pieces of evidence (regardless of their type) did you use typically to judge TAGs for each of your students?

This question was only asked of respondents who had indicated that they were describing the TAG process for a GQ. Respondents were asked to enter the number of distinct pieces of evidence and the distribution of the responses is shown in Figure 12.

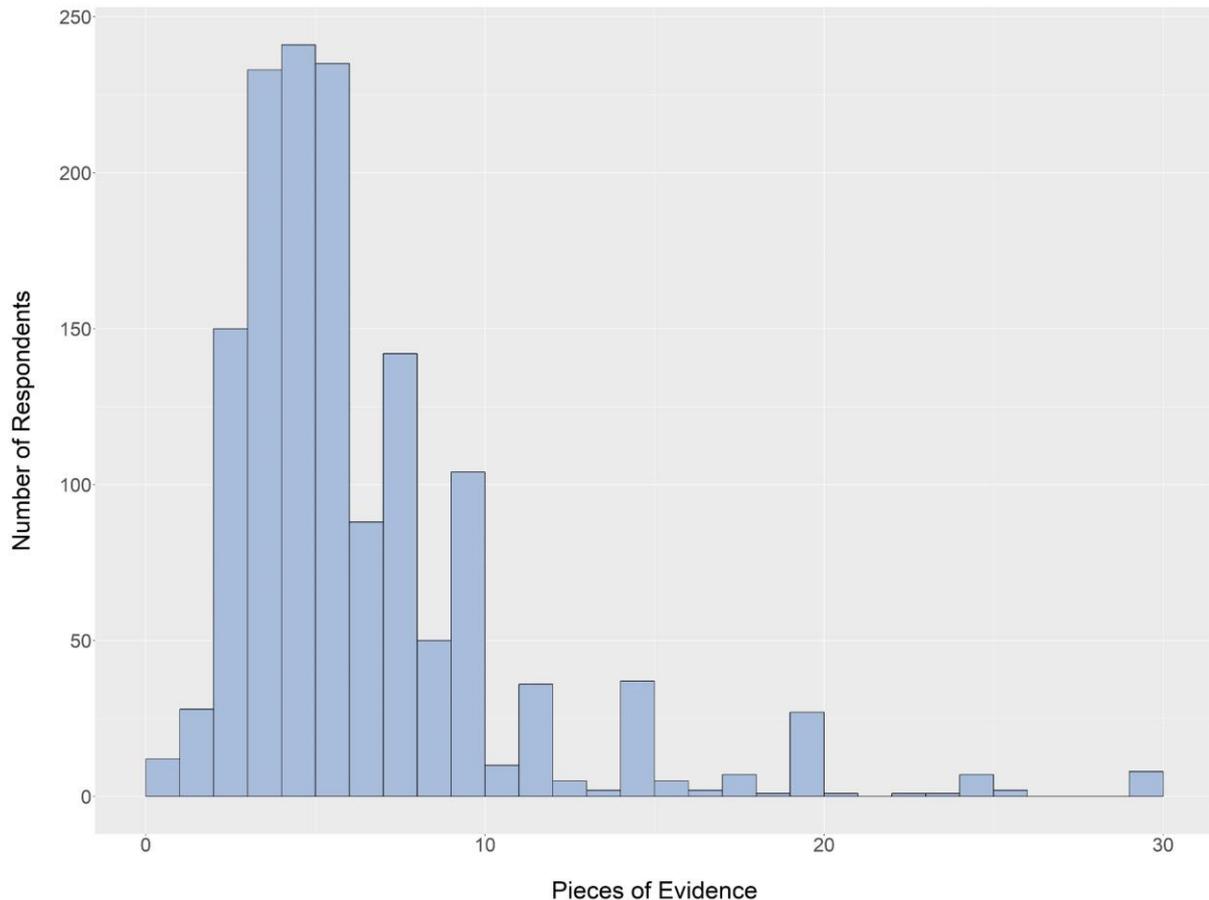


Figure 12: Number of respondents by number of pieces of evidence used (N = 1,456)

The median number of individual pieces of evidence used when judging TAGs was 6, with half of all respondents using between 4 and 8. The most common (modal) response was 5. Less than 1% of respondents (11 in total) indicated they used only one piece of evidence and one respondent indicated they used none (though, based on their other responses, we believe this was likely to be an error). Of these 12 respondents, 4 stated that they were judging maths and 4 stated they were judging an art-related subject. Conversely, 29 (2%) respondents indicated that they utilised 30 or more pieces of evidence when judging TAGs for each of their students (not shown on the graph).

Other than for the art-related subjects, we found little variation amongst subjects with most medians ranging between 4 and 7 pieces of evidence used. The art-related subjects, however, were different, with a substantially larger median value (10) and upper quartile (50). Indeed, of the 14 respondents who indicated they used 50 or more pieces of evidence, 9 (64%) were responding about art-related subjects. This is likely due to the portfolio of evidence used in art-related subjects for their non-exam assessment components. We found little variation by centre type, with the medians from all types being between 6 and 7 pieces of evidence used.

We note that the question asked how many pieces of evidence were used when judging TAGs for individual students. A low number does not necessarily indicate

that students were only given 1 or 2 opportunities to demonstrate their ability. It is possible that some respondents used the best grade a student attained from a range of assessments rather than some form of average across multiple pieces of evidence.

In addition, for those respondents who indicated they used a rather large number of pieces of evidence, it is possible that not all the pieces of evidence were weighted equally or, indeed, that all of them contributed to the final grade. The selection and combination of the evidence is explored more fully in the [companion interview report](#).

Q. How difficult were the grade judgements for the following?

Respondents were asked to rate how difficult it was to produce the grade judgements for:

- an average across all students
- students with inconsistent performance
- private candidates
- students who joined their centre recently
- students with reasonable adjustments (or access arrangements)
- students who had missed more content than others or had been absent for long periods.

The response options for grading the difficulty were:

- not difficult
- slightly difficult
- difficult
- very difficult
- not sure
- not applicable

The response for each of these learner types is shown in Figure 13. We note that the number of responses per type of learner does not equal the total number of respondents, so we have included the response numbers for each learner type in their respective plot titles. To produce meaningful comparisons across learner types, we have excluded 'not applicable' responses from the analysis.

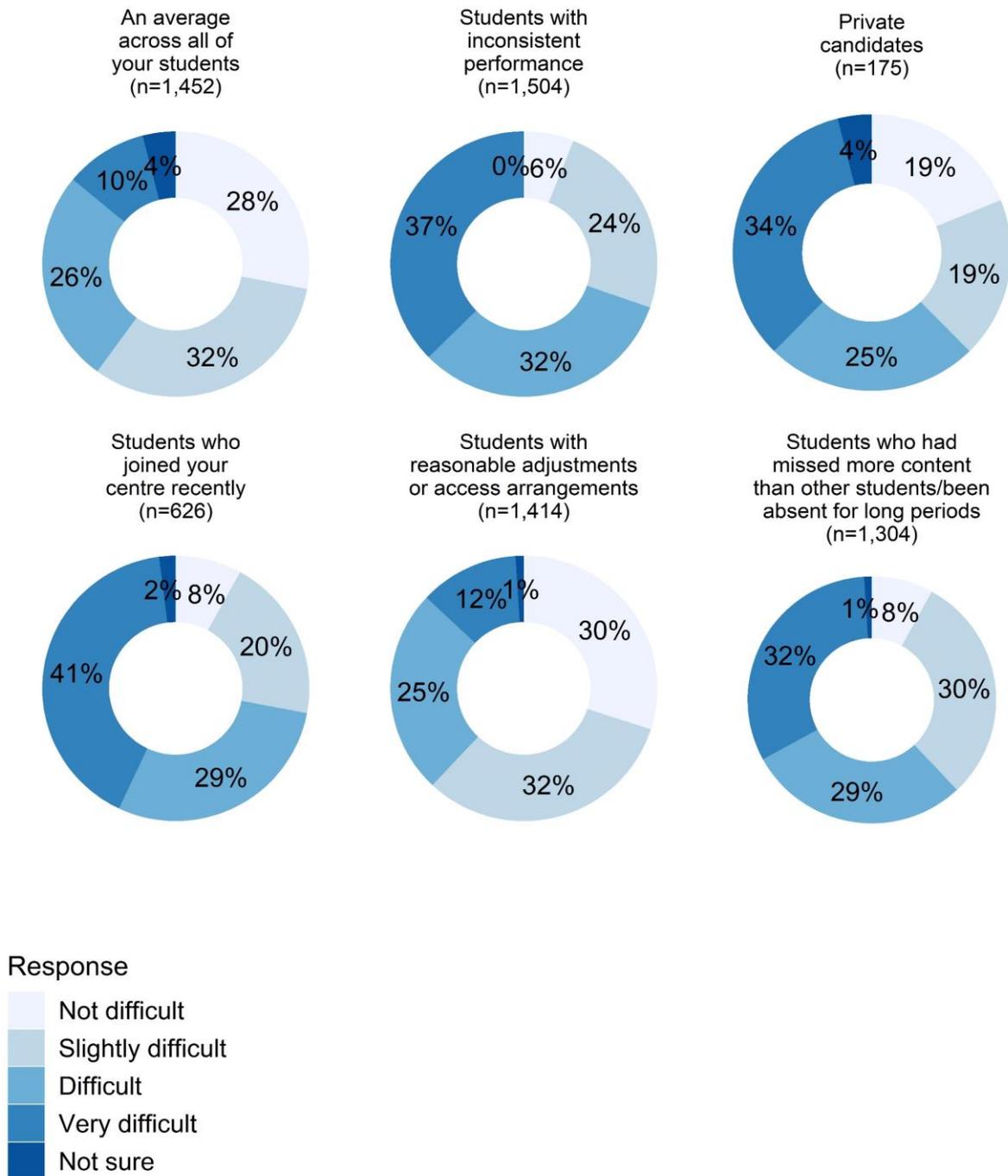


Figure 13: Percentage share of difficulty responses for each type of student (N = 1,533)

As an average across all students, 36% of respondents found the grade judgements difficult or very difficult. To compare the difficulty across each type of learner, we assign a difficulty score to each response option (Not difficult = 1 through to Very difficult = 4) and compute the mean score. The 'not sure' and 'not applicable' responses are ignored.

Students who had recently joined their centre and students with inconsistent performance were the categories respondents found most difficult to judge (mean of 3.0). Both of these categories of students therefore fall into the category of 'difficult'

on average. Students who had missed more content than others and private candidates scored similarly (mean of 2.8). The average across all students and students with reasonable adjustments also scored very similarly (mean of 2.2), being judged 'slightly difficult' indicating that the grade judgement process for most students was not considered too difficult.

Q. Did any of your students require a reasonable adjustment, such as extra time or assistive technology?

Respondents were asked to respond 'yes' or 'no' to whether any of their students required reasonable adjustments (N = 1,531). Eighty-eight per cent of respondents indicated that at least one of their students required a reasonable adjustment.

Q. How were reasonable adjustments applied in the judgment of TAGs?

Those respondents who indicated that one or more of their students required a reasonable adjustment were then asked about how this reasonable adjustment was applied (N = 1,343). They were asked to choose from the following 3 options and the percentages selecting each are stated in brackets:

- Available for students to use in all assessments (48%)
- Taken into account when judging TAGs (9%)
- Both, depending on assessment type (44%)

Where required, reasonable adjustments were largely applied to assessments (92%), although they were also frequently accounted for when making decisions on TAGs (53%). Some centres may have chosen to make adjustments to grades for logistical reasons, such as time or equipment constraints, or because TAG evidence types may not have lent themselves to the adjustments.

Q. For what proportion of students did you feel relatively unsure of the grade to give?

Respondents were asked to enter the proportion of students for whom they felt relatively unsure of the grade to give. The distribution of these responses is shown in Figure 14.

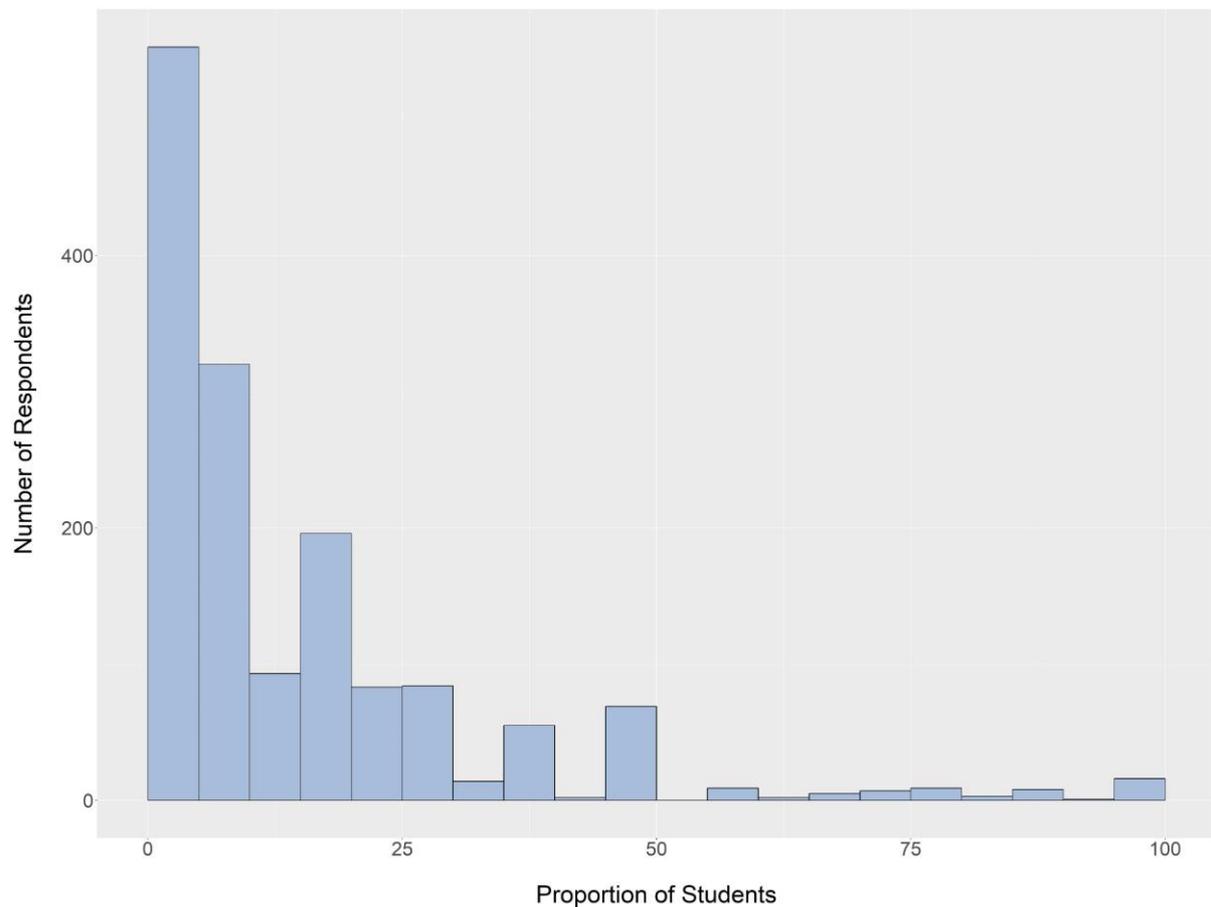


Figure 14: Number of respondents by proportion of students they were unsure of grade (N = 1,529)

The median proportion of students for which respondents felt unsure of the grade to give was 10%, with half of all respondents feeling unsure of between 5% and 20% of students. Over three quarters (76%) of respondents felt unsure of the grade for fewer than a quarter of their students, with 92% respondents for fewer than half. It is likely that many of these students fell into the categories shown in Figure 13 or were students who were close to grade boundaries.

Q. Did you feel any undue pressure on your professional judgement?

Respondents were asked if they felt any undue pressure on their professional judgement (N = 1,535). Overall, the respondents were almost equally split on whether they felt any undue pressure on their professional judgement, with 50% for both yes and no. This shows a substantial increase in the number of respondents feeling undue pressure when compared to the corresponding question in the 2020 CAG survey, where only 31% indicated that they felt undue pressure on their judgment.

Most respondents from secondary selective and academy schools indicated that they felt undue pressure (55% and 53% respectively) which is a slightly higher proportion than on average. Meanwhile, the proportion of respondents from independent schools who felt undue pressure was slightly lower than average (46%).

We also asked those who responded 'yes' to provide further details. Free text responses were received from 678 respondents. These were analysed and drawn into several common themes discussed below.

A large number of respondents (31%, n = 209) noted the pressure they felt from their centre's senior leadership team (SLT). This often centred around reducing student grades to meet previous year's grade profiles. A slightly smaller, but still substantial number of respondents noted pressure from SLT to increase grades. Both pressures conflicted with their professional judgement and the need to be fair to their students. They also mentioned having individual grading decisions checked and scrutinised by management and the burden in having to justify decisions and the additional work this sometimes required.

Externally, some respondents (21%, n = 140) noted that they felt pressure from students and their parents. Sometimes this was in anticipation of results day and any potential appeals, or knowledge of what their students needed for their next steps and the pressure this put on them when determining TAGs. However, some mentioned that they were contacted during the TAGs process from students or their parents regarding grading, what students needed for future progression or individual circumstances that should be taken into account, though it was sometimes reported that steps were put in place by the centre to minimise the effect of this, or that it was simply ignored.

The other main source of external pressure was public perception of teachers (7%, n = 47). Respondents felt that there was a general lack of trust in their judgement and respect for their profession, both publicly and from government. They felt as though they would be blamed regardless of outcome, either for grade inflation or for not giving students the grades the students may have wanted or felt they deserved, and this added to their overall feeling of being under pressure. The emotional burden of

being responsible for student futures, and the general desire to be fair were mentioned by 18% (n = 125) respondents.

Pressures resulting from the TAGs process itself were also mentioned many times (23%, n = 156). This centred on the substantial time and workload pressures respondents faced because of the need to create, run, and mark assessments. The difficulty in making the grading decisions was also mentioned, particularly surrounding use of grade descriptors and specific types of students (Figure 13), and sometimes restricted capacity to exercise their professional judgement because of the more numerical approach taken by their centre. Some noted that grading students was not something they had been trained for or felt comfortable doing, and several less experienced teachers felt particularly uncomfortable with doing so. A final pressure that came up quite frequently was the comparison across schools, with teachers unsure whether the more rigorous processes they felt they had applied might disadvantage their own students.

Respondents also noted that pressure came from both the requirements of the task, but also a lack of support (18%, n = 125). They felt that there was pressure from government (for example, Ofqual and the Department for Education) and awarding organisations to get things right despite there being, in their opinion, insufficient support or planning from those external agencies. These pressures centred on difficulties because of the timing and quality of issued guidance, and more often, the quality of the materials (grade descriptors, assessment materials) provided by awarding organisations.

Q. Were you aware of steps taken by your centre to protect you against external influences (such as parents and students) on your TAG judgements?

Respondents were asked if they were aware of any steps taken by their centre to protect them from external influences (N = 1,533). The following responses were available, and the percentage giving each response is stated in the brackets:

- Yes – there were effective steps taken (65%)
- Yes – there were some partially effective steps taken (15%)
- No – no steps taken (4%)
- Not sure – there may have been some steps taken but I was not aware (16%)

In total, 80% of respondents indicated that there were at least some partially effective steps taken by their centres to protect them from external influences. These results are similar to the 2020 CAG survey, though the percentage of respondents

indicating that partially effective steps were taken increased slightly (+5%) whilst the percentage of respondents who were not sure decreased by the same amount.

Q. How confident were you that your TAG judgements (prior to any Internal Quality Assurance activities) were as free as possible from bias?

Respondents were asked to rate how confident they were that their TAG judgements, prior to any Internal Quality Assurance activities, were as free as possible from bias from a scale of 0 (no confidence) to 100 (high confidence). Answers were restricted to integer values and have been grouped into bins with widths of five in Figure 15.

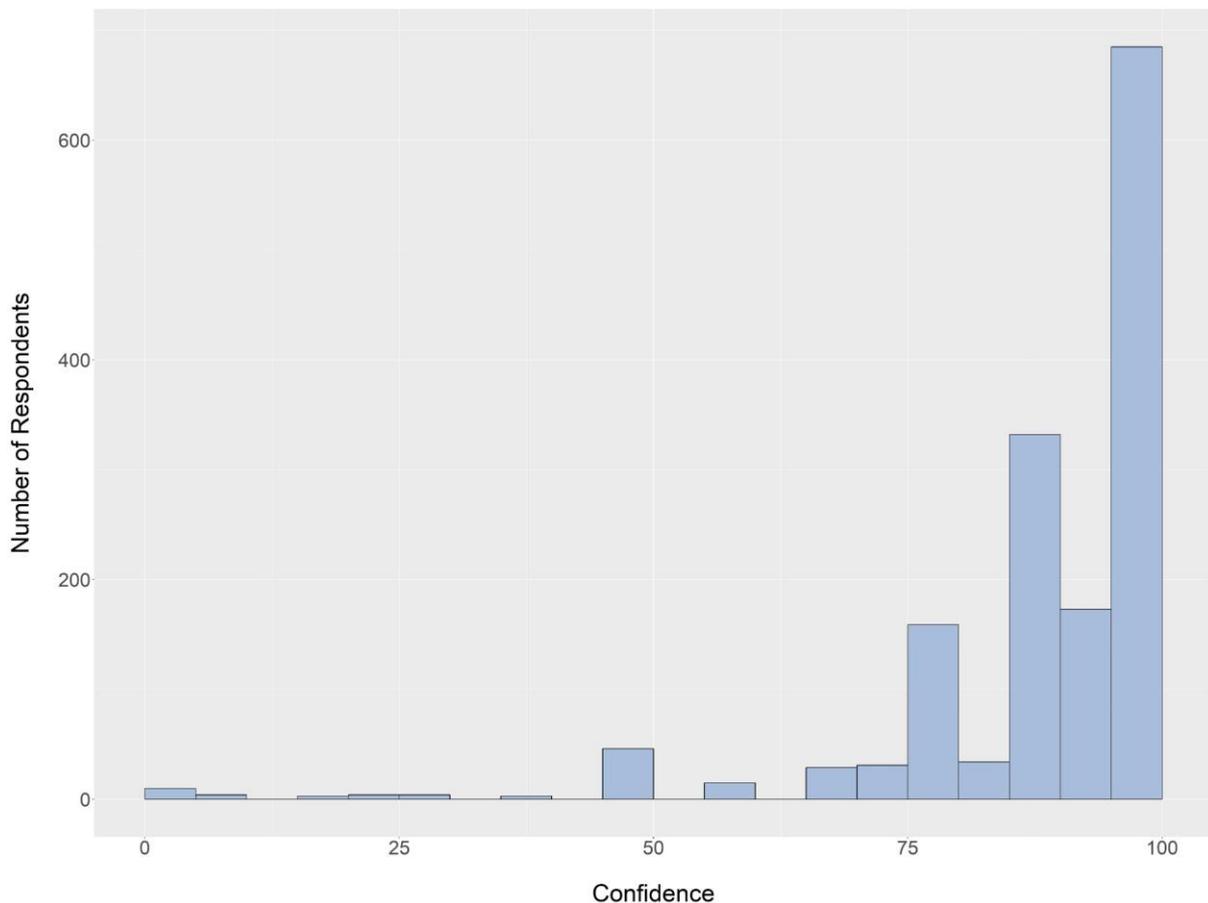


Figure 15: Respondent number by confidence in judgement free from bias (N = 1,532)

The median confidence was 95, with half of all respondents reporting confidence between 90 and 100. Over 90% of respondents rated their confidence at 75 or higher, with 40% rating their confidence at 100. We note these are confidence ratings, and low confidence does not necessarily mean the presence of bias just as a high confidence rating does not guarantee the absence of bias.

Confidence in the absence of bias was high on average amongst all types of GQ respondents. All subjects had a median confidence of 90 or higher and there was no variation to confidence based on respondent centre type.

Q. How confident overall were you in the accuracy of your TAG judgements (prior to any Internal Quality Assurance activities)?

Respondents were asked to rate how confident they were in the accuracy of their TAG judgements, prior to any Internal Quality Assurance activities, from a scale of 0 (no confidence) to 100 (high confidence). Answers were restricted to integer values and have been grouped into bins with widths of five in Figure 16.

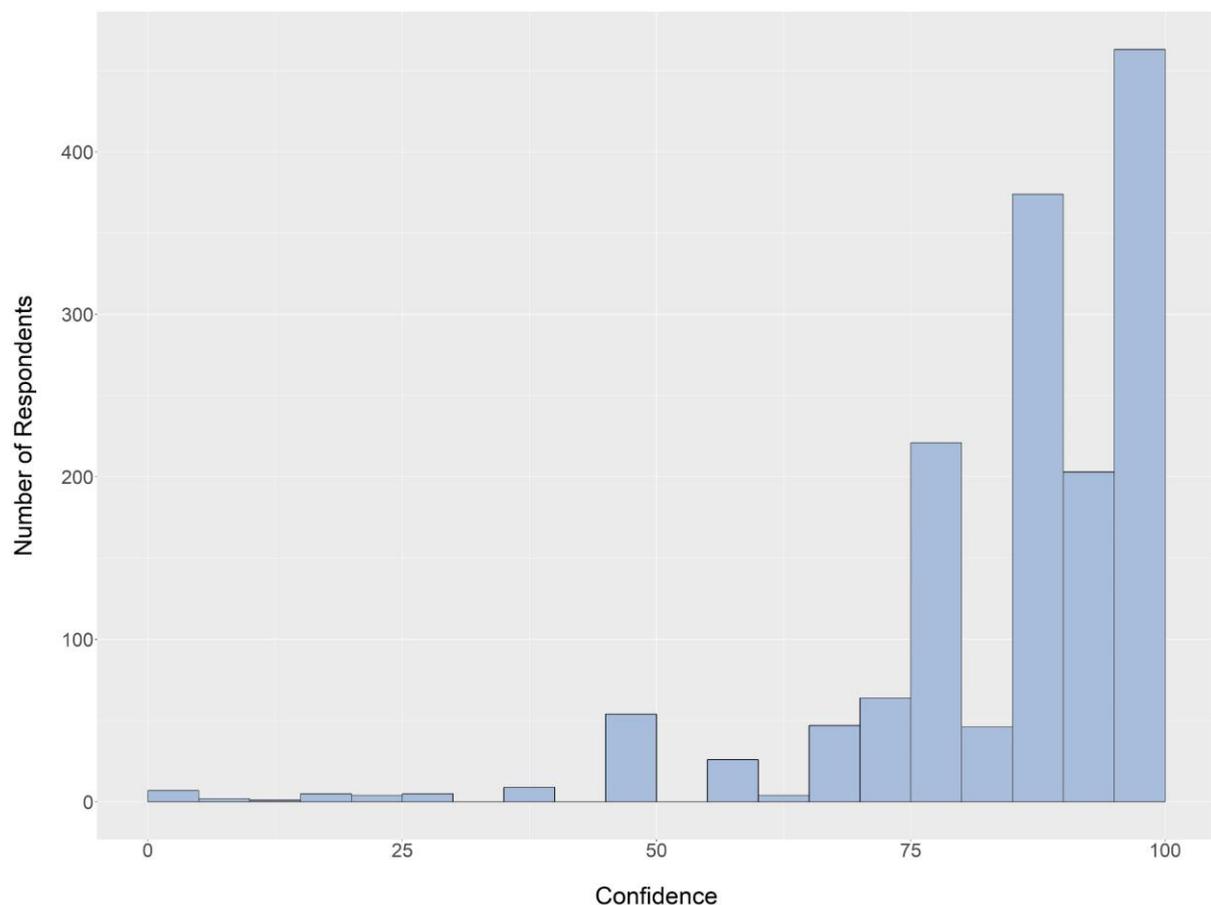


Figure 16: Respondent number by confidence in judgement accuracy (N = 1,535)

The median confidence was 90, with half of all respondents indicating confidence between 80 and 100. Over 85% of respondents rated their confidence at greater than 75, with 27% rating their confidence at 100. Again, we note that these are confidence ratings and do not represent the actual accuracy of the TAG judgments.

Confidence in the accuracy of the TAG judgements was, on average, high across all GQ subjects. Indeed, all subjects had a median accuracy rating of 90 or higher. We note that the subjects with a median confidence of more than 90 tended to be those with smaller respondent numbers and those with higher respondent numbers tended to equal the median across all subjects (90). The significance of any differences by subject is therefore unclear. There was no apparent variation to confidence based on respondent centre type.

Q. Please write down up to three words that summarise how you felt about the experience of judging TAGs

Respondents were asked to provide up to 3 words summarising how they felt about the experience of judging TAGs. While the examples given on the screen seen by respondents suggested 3 unlinked words, some respondents typed in phrases or full sentences, sometimes significantly longer than 3 words. We processed the responses by removing all 'stop words' (for example, 'I', 'and', 'the') and for fairness we selected only the first 3 words remaining from each response. Words which appeared at least twice in the responses to this question are presented in the form of a word cloud in Figure 17.



Figure 17: Summarising judging TAGs feeling word cloud (N = 1,520)

The top 10 most frequently used words were:

- stressful (n = 514, 34%)
- time-consuming (n = 231, 15%)
- exhausting (n = 202, 13%)
- unfair (n = 150, 10%)
- stressed (n = 102, 7%)
- pressured (n = 90, 6%)
- difficult (n = 82, 5%)
- pressure (n = 82, 5%)

- fair (n = 82, 5%)
- pressurised (n = 60, 4%)

It is clear that the overall sentiment from this word cloud is negative, with 42% of respondents mentioning the word stress, or one of its derivatives, in their response and 9 out of the 10 most frequent words providing a negative sense.

This contrasts with the corresponding word cloud in the 2020 CAG survey, where we saw an approximately equal split between positive and negative sense words.

Additionally, whilst some of the negative words are repeated from the CAG survey (for example, stress, pressure, and unfair) we find that new, strongly negative words (for example, exhausting, time-consuming, and difficult) are present in this list that were not before. We consider this further in the discussion.

Agreeing TAGs

Respondents who indicated that they were involved in agreeing TAGs were shown this section. Those who indicated that they had judged TAGs were automatically shown this section, whilst those who indicated they had not were asked to confirm if they had had any involvement in agreeing them.

Q. Did you produce the first set of grade judgments for individual students or were you an additional person checking/agreeing these initial grades?

Respondents were asked about the role they had in agreeing the TAGs (N = 1,650). They were given 3 options and asked to select all that applied. The percentage selecting each option is given in brackets:

- Original person making judgements (63%)
- Shared responsibility for making the original judgements (46%)
- Checking or agreeing someone else's initial judgements (41%)

The most frequently selected role was 'original person making judgement' (63%). Note that respondents were asked to select all that applied from the 3 options, hence the sum of the 3 options being greater than 100% and the indication that some respondents carried out multiple roles.

Q. How easy or difficult was it to agree all of the individual TAGs with the other member(s) of staff?

Respondents were asked to rate how easy or difficult it was to agree TAGs with other members of staff, using a 5-point scale from 'very easy' to 'very difficult', as well as 'not sure'. The percentages for each response are shown in Figure 18.

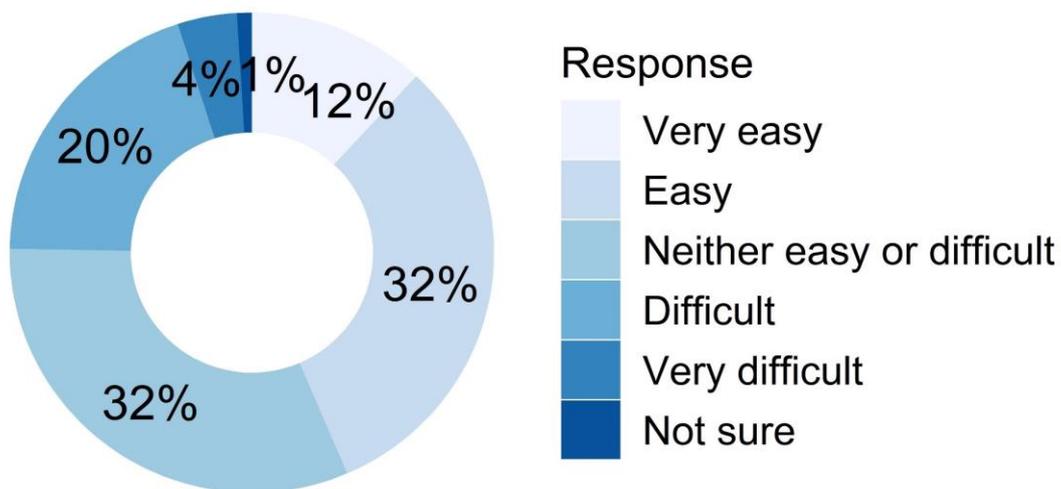


Figure 18: Percentage of respondents per difficulty option on agreeing TAGs (N = 1,648)

A total of 43% found the process of agreeing TAGs with other member(s) of staff easy or very easy, whilst a total of 24% found it difficult or very difficult.

Q. Did you experience any of these difficulties when agreeing the TAGs with the other member(s) of staff?

Respondents (N = 1,594) were asked to select all of the difficulties they faced when agreeing TAGs with other member(s) of staff from the following list of options, with the percentage selecting each response stated in brackets:

- No difficulties (32%)
- Logistical difficulties in holding discussions (40%)
- Different interpretation of standard of work (36%)

- Different emphasis on different sources of evidence (26%)
- Different views on content to be assessed (16%)
- Different views on how the internal quality assurance process would work (15%)
- Different views on how the external quality assurance process would work (12%)

There was also an 'other' option and entries here were re-coded into an existing category or into one of the new categories below (with percentages in brackets):

- Difficulties in handling mitigating circumstances (including long periods of absence) or inconsistent performance (1%)
- Differing views or other issues with scaling to previous years' cohorts (1%)
- Issues with, or the interpretation of, guidance (including Centre Policy) and resources (for example, grade descriptors) (1%)
- Unhelpful interference, pressure, or decisions by SLT (1%)

Note that respondents were asked to select all that applied from the options, hence the percentages summing to more than 100%.

Internal quality assurance (or standardisation)

Q. Were you involved in the internal Quality Assurance/Standardisation process for at least one qualification?

All respondents (N = 1,785) were asked whether they were involved in the internal quality assurance (IQA) process as a filter question to this section. Those that answered 'yes' were asked further questions from this section while those that answered 'no' skipped to the next section. The majority of respondents (82%) indicated that they were involved in the internal quality assurance (or standardisation) process for at least one qualification.

Q. Who was involved in this internal QA process before submission of TAGs to exam boards/awarding organisations?

Respondents (N = 1,463) were asked to select all who were involved in the internal quality assurance (IQA) process before submission of TAGs to exam boards or awarding organisations from the following list of options, with the percentages selecting each option given in brackets:

- Head of department (92%)
- Class teachers or tutors (86%)
- Other members of the senior leadership team (55%)
- Head of centre (45%)
- Deputy head of centre (40%)
- Deputy head of department (39%)
- Examinations officer (31%)
- Data manager or other member of data team (23%)
- SENCo or other SEN experts (22%)
- Diversity expert (1%)

There was also an 'other' option and responses entered here were re-coded into existing categories or counted in 2 new categories listed below:

- External colleagues, markers, or advisors' (including colleagues from other schools in an academy trust) (3%)
- Pastoral Team' (for example, head of year) (1%)

Q. In your view, were there any other people that could have been involved in your centre whose expertise and knowledge would have been useful?

Respondents were asked whether there were any other people that could have been involved in the IQA process (N = 1,444). The overwhelming majority (92%) of respondents indicated that there were no other people that should have been involved in the TAG IQA process. Of the 115 (8%) who responded that there should

have been others involved, 100 provided further details. We have coded these responses into the categories shown in Figure 19.

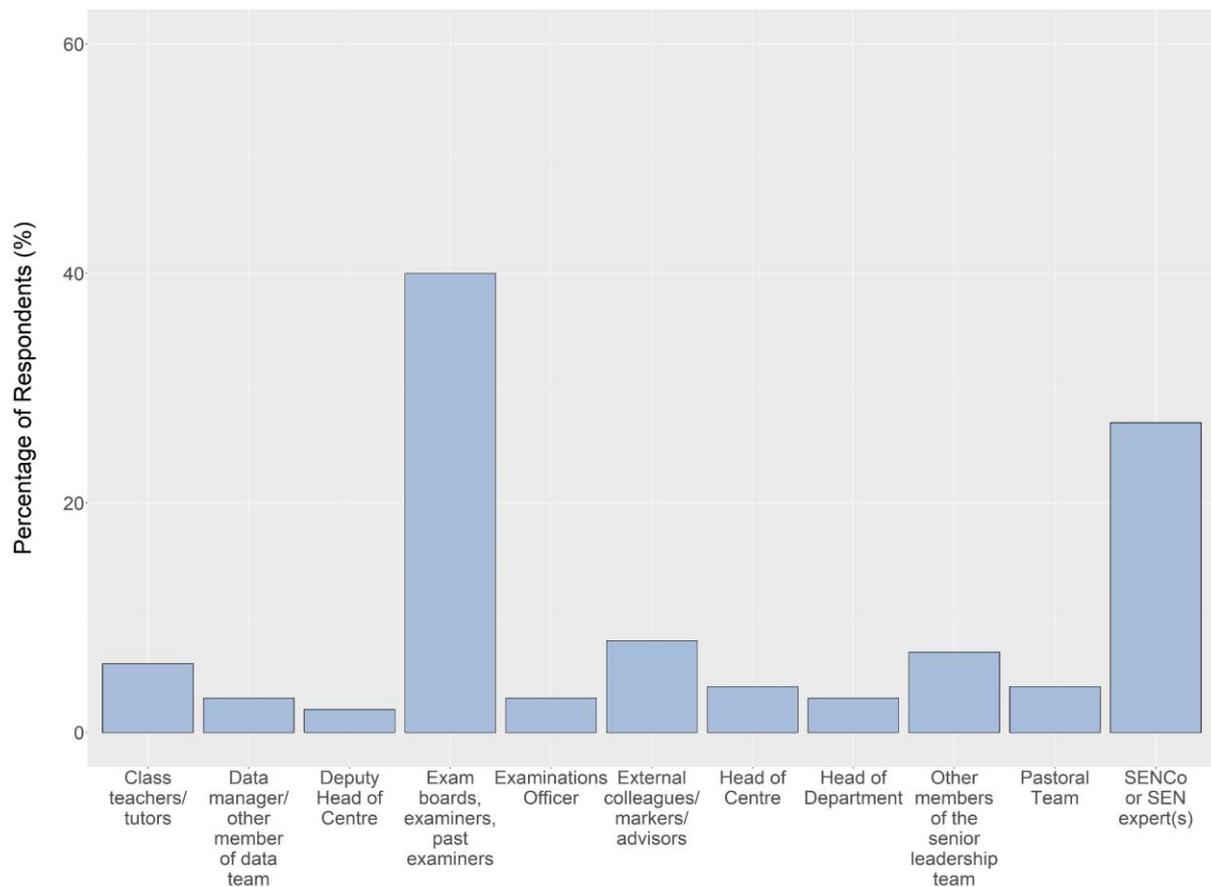


Figure 19: Who else should have been involved? (N = 100)

The most frequently suggested response for other people to be involved in the IQA process was exam boards, examiners, and past examiners (40%). This response included both external staff (for example, exam board staff) as well as internal staff who were current or past examiners for exam boards. SENCo or SEN experts were the next most suggested group (27%). Responses for all other types of staff were relatively low (less than 10% of the 100 responses).

Q. To what extent were the following important parts of the internal QA process for the TAGs?

Respondents were asked to rate how important several inputs had been to the internal QA process. The percentages for each factor are shown in Table 5.

Factors	To a great extent	To a moderate extent	To a very small extent	Not at all, Not used or Not sure
Centre policy	57%	25%	9%	9%
Comparison of TAGs to previous years' attainment data	44%	36%	11%	8%
Comparison of TAGs across different types of student within your centre	25%	36%	14%	25%
Comparison of TAGs across different subjects within your centre	18%	31%	16%	35%
Input from SEND experts or SENCos	14%	29%	24%	32%
An external agency to supply information or check outcomes this year	6%	12%	10%	72%

Table 5: Percentages of respondents indicating how important several inputs were to the internal QA process (N = 1,459)

The most important parts of the internal QA process were comparisons of TAGs to previous years' attainment data and the centre policy (91% of respondents selected 'great extent', 'moderate extent', or 'very small extent' for both). The least frequently used part was an external agency to supply information or check outcomes (28%).

To compare importance across the various aspects of internal QA, an importance rating was determined for each by assigning a numerical score to each response type (from 1 for 'not at all' to 4 for 'to a great extent').

The mostly highly rated aspect of the internal quality assurance process was the centre policy document (mean of 3.4), followed by use of previous years' data (mean of 3.3), comparison across students (mean of 2.8), comparison across subjects (mean of 2.5), SEND input (mean of 2.4), and, lastly, using an external agency (mean of 1.6).

Q. Were there other sources of information which may have been useful to your centre's internal QA process?

Respondents were asked if there were other sources of information that would have been useful in the IQA process (N = 1,415). The majority of respondents (83%) indicated that there were not any other sources of information that would have been useful. The remaining 17% of respondents indicated there were.

We provided a free-text box for respondents to specify what additional sources could have been used. We received 245 responses (including some from those who had said 'no' or had not answered the question) of which we were able to code 179 responses into the following categories:

- Training, mentoring, or support from, particularly, the exam boards – these suggestions included requests for more detailed guidance, access to more support and resources from AOs, and the provision of relevant training or mentoring (36%, n = 64)
- More detailed grade descriptors, exemplars, grade boundaries – this covered more detail from exam boards on how to grade student work including more comprehensive grade descriptors, more exemplar answers, and details on how to apply grade boundaries to papers - including past papers (27%, n = 48)
- Papers from exam boards – largely these comments expressed the need for new materials in the form of questions or whole papers, stored securely so that students could not access them, and including full mark schemes for this new material (16%, n = 28)
- Discussions or moderation with other centres – suggestions included organisation of networks of centres to discuss each other's processes, to formal cross-centre moderation activities (14%, n = 25)
- Better or more guidance from Ofqual or the government – these were general comments on Ofqual or government providing better guidance, sometimes in the form of requests for stricter rules to ensure consistency across centres (13%, n = 24)
- Better training, guidance, or data from centres – anything that could be provided within centres, including data from previous years' cohorts (4%, n = 4)

We did not code a number of responses related to the timing of guidance (by Ofqual or the awarding organisations) as they did not relate to the question being asked.

Final thoughts on submitted TAGs following internal quality assurance

Q. Did you have sight of the final grades that were submitted to the exam board/awarding organisation for the qualifications you were involved in, following any internal QA?

All respondents were asked a filter question to this section (N = 1,785). Those that answered 'yes' were asked further questions from this section; those that answered 'no' skipped to the next section. The majority of respondents (82%) indicated that they did have sight of the final grades that were submitted to the exam board or awarding organisation, after internal quality assurance had taken place.

Q. If you judged TAGs for individual students you taught, were any of your original TAGs changed following the internal QA activities?

Respondents (N = 1,461) were asked if the TAGs they had awarded were subsequently changed through internal quality assurance (IQA) processes. Ten per cent of respondents indicated that they hadn't worked on individual TAGs before the IQA process. These responses mostly came from respondents with roles such as exams officer, head of department, or a senior leadership role.

Considering just the 1,313 (90%) respondents who did work on TAGs before the IQA stage, 51% indicated that none of their TAGs were changed whilst 45% indicated that at least one of their TAGs was changed. Four per cent did not know, presumably they were not involved in agreeing the final TAGs.

Respondents who indicated that at least one of their TAGs had been changed were then also asked to tell us why they had been changed. We received written reasons for the changes from 409 individuals that we coded into the following categories.

- General moderation – These comments referred to grades being altered because of normal moderation and standardisation activities, for example, comparisons across year group or second marking. There were also some instances of fixing procedural or administrative errors. These changes had a mixed effect on grades, some increased, some decreased. (37%, n = 153)
- Centre or department prior attainment profile – Comments reflected that where current grade distributions were not in line with past distributions, TAGs were changed to better match the past results. Generally, grades went down in these instances. Responses from teachers or tutors often noted their disagreement with this approach as it overruled their judgement. (29%, n = 119)
- Borderline cases – Cases where individual students who were on the border between grades had their TAG changed. Generally, grades were adjusted up in these cases, though not exclusively. (16%, n = 67)
- Change in evidence or its weighting – These were instances where evidence was added or removed from the set of evidence used, or the weighting of such evidence was changed, after the initial TAG had been judged. Sometimes this referred to an individual student (for example, to account for missed work or inconsistent performance) or sometimes for an entire cohort (for example, changes to evidence selection or weightings for all to give a more valid set of grades). These adjustments mostly raised grades. (11%, n = 45)
- Extenuating circumstances or special consideration (including periods of extended absence) were also applied centrally to individual students. These were only upwards adjustments to grades. (10%, n = 41)
- SLT pressure to change – These were cases where the senior management requested, or demanded, that initial grades were changed by teachers or departments because of their own analysis. Respondents noted that this was mostly without stated reasons and not in line with the judgement of the teacher or tutor. Grades could be adjusted up or down under these circumstances. (5%, n = 22)
- SEND considerations – Following input from SEND specialists, grade adjustments were sometimes made, relating usually to access arrangements or reasonable adjustments. Grades were only adjusted upwards in these cases. (2%, n = 9)
- Concern for external QA – This theme includes explicit mention of a centre's concern about external QA process. Perceived pressure from EQA would lead to downwards adjustments to grade to avoid them being queried. (1%, n = 5)

We also asked those completing the free text question to enter the percentage of their TAGs that were changed. Not all did so and we received only 170 numerical responses.

The median proportion of TAGs changed was 10%, with half of all respondents indicating that between 5% and 10% of TAGs were changed. We note that this is the median only of those who had indicated that their TAGs were changed and then provided a value. It does not include the majority of respondents who indicated that none (0%) of their TAGs were changed, or those that did not provide a value.

Q. How fair do you think these submitted TAGs are compared to the grades awarded to your students following normal assessments in past years?

Respondents were asked to indicate how fair they felt the submitted TAGs were compared to normal assessments, using a 5-point scale from 'much more fair' to 'much less fair', and also with a 'don't know' option (Figure 20).

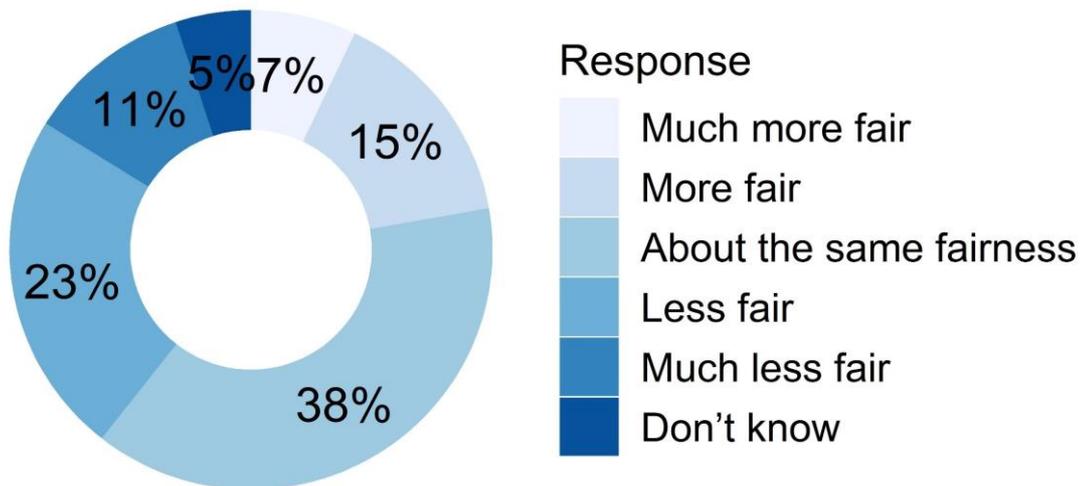


Figure 20: Percentage of respondents by fairness of TAGs (N = 1,539)

In total, 22% of respondents thought TAGs were fairer whilst 35% thought they were less fair. About the same fairness (38%) was the most frequent response. This represents a slight increase in the overall perception of fairness from the 2020 CAG survey. Note that, similar to the CAG survey last year, these responses were based

upon the submitted grades, not the final awarded ones for external quality assurance.

In the 2020 CAG survey, it was found that senior staff had a more positive view of the CAGs that had been submitted than other staff. To determine if this sentiment was matched with TAGs, we analysed the average fairness rating by role. Responses were scored from 5 ('Much more fair') to 1 ('Much less fair') and a mean average taken by the group. 'Don't know' responses were omitted.

We found that only small differences this year. Middle leaders (head and deputy or assistant head of department and key stage leader) rated the fairness of TAGs the highest (mean of 2.9). They were followed by teachers, tutors, or trainers (mean of 2.8) and then senior leaders (head of centre, deputy or assistant head of centre, and other senior leadership team member) who gave a mean rating of 2.8. All groups rated TAGs as slightly less fair than grades awarded in past years, on average.

Respondents from colleges were slightly more positive than the average picture with 30% of respondents indicating that they felt TAGs were fairer than normal assessments (vs 22% across all centre types). Conversely, respondents from Independent schools were slightly less positive, with 39% of respondents indicating that they felt TAGs were less fair (vs 34% across all centre types).

Q. How confident overall were you in the accuracy of the final submitted TAGs (after any internal QA activities) that you were involved in?

Respondents were asked to rate how confident they were in the accuracy of the final submitted TAGs on a scale of 0 (no confidence) to 100 (high confidence). A histogram of the responses is shown in Figure 21.

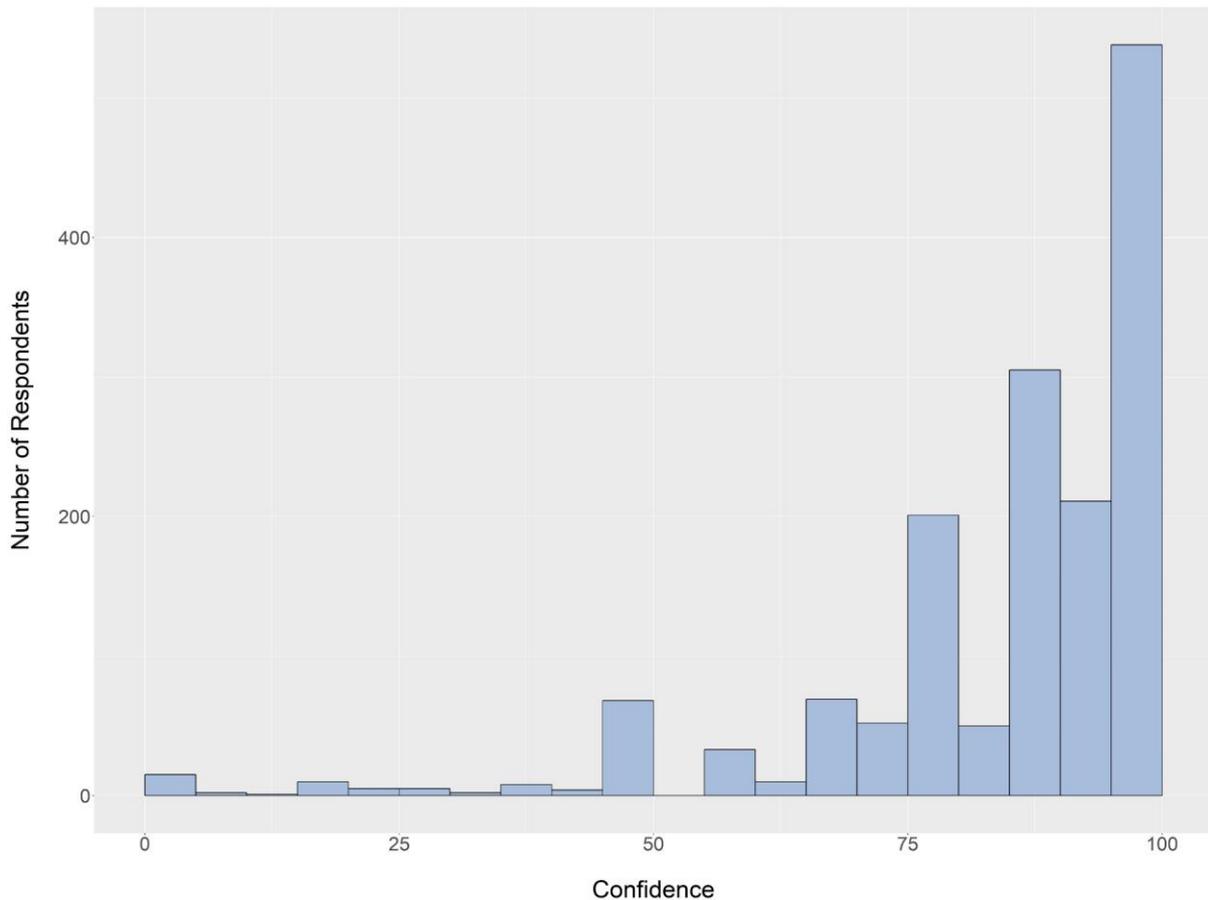


Figure 21: Number of respondents by confidence in accuracy of final TAGs (N = 1,589)

The median confidence rating was 90, with half of all respondents indicating a confidence of between 80 and 100. Just over 85% of respondents rated their confidence at greater than 75, with 29% rating their confidence at 100. This result is consistent with the results from the 2020 CAG survey.

Confidence in the accuracy of the TAG judgements was, on average, high across the board – for all subjects and centre types. We note that these results are almost identical to the levels of confidence reported before internal QA processes were undertaken (Figure 16) – suggesting that the internal QA process did not alter respondents’ confidence in either a positive or negative way.

We also analysed this question by respondents’ role and years in the profession. We found no significant variation amongst role types with median confidence of 90 or higher across all roles. For years in profession, we grouped respondents into 3 categories, ‘0-2 years’, ‘3-5 years’, and ‘6+ years’ in the profession. Again, we found no statistical difference between these categories. These results suggest respondents were confident in the accuracy of TAGs regardless of role or years spent teaching.

Centre declaration responsibility

Q. Did you sign the centre declaration form(s) for submission to awarding organisations/exams boards?

All respondents (N = 1,785) were asked a filter question to this section regarding whether they signed the centre declaration form(s). The majority of respondents indicated that they did sign the centre declaration forms (56%) and were asked further questions from this section; those that answered negatively skipped this section.

Q. Which aspects of the whole process of generating TAGs were you personally involved in?

Respondents (N = 987) were asked to select all the aspects generating TAGs that they were involved in from the list below, and the percentages selecting each are shown in brackets.

- Generating teacher assessed grades for one or more classes you taught (89%)
- Internal quality assurance of the teacher assessed grades before submission (86%)
- Training and/or discussions about making objective bias free judgements for all types of students (69%)
- Initial planning discussions (66%)

Q. How confident were you that the final submitted TAGs across all qualifications you signed off were as free as possible from bias?

Respondents were asked to rate how confident they were that the final submitted TAGs were as free as possible from bias on a scale of 0 (no confidence) to 100 (high confidence).

The median confidence rating was 95, with half of all respondents indicating confidence between 90 and 100. Over 91% of respondents rated their confidence at over 75, with 43% rating their confidence at 100. These results are highly consistent with the confidence respondents felt in the TAGs they judged themselves before internal QA processes were undertaken (Figure 15) suggesting that for these respondents the internal QA process had no substantive effect on respondents' confidence (either positively or negatively). Additionally, there was no significant variation by centre type or subject in the respondents' confidence of being free from bias.

We note that these results are also similar to the comparable question in the 2020 CAG survey, suggesting little change in the confidence of respondents between the years.

Final thoughts

This section was presented to all survey respondents, with no filter question.

Q. Please give your best estimate as to how long you dedicated to work on the TAGs in total.

Respondents were asked to estimate how long they had dedicated to work on the TAGs, in days. A histogram of the responses is shown in Figure 22.

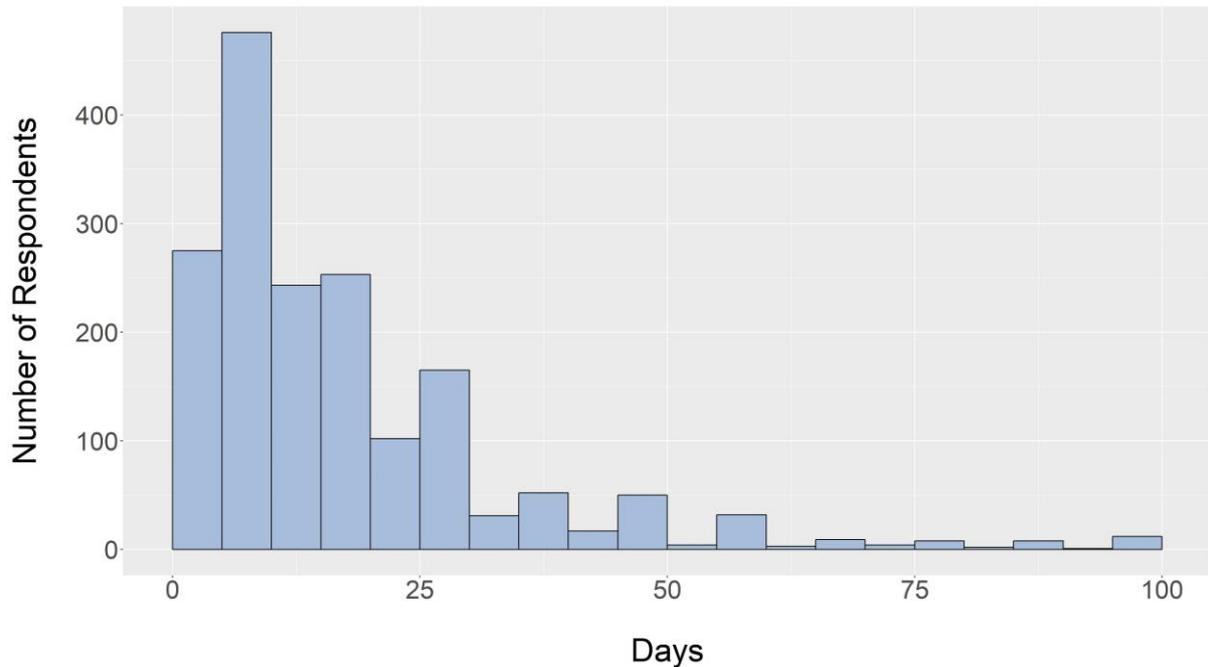


Figure 22: Please give your best estimate as to how long you dedicated to work on the TAGs in total. (N = 1,768)

The median time taken was 15 days, with half of all respondents indicating time taken was between eight and 25 days. Just over 85% indicated that they spent less than 30 working days dedicated to working on TAGs. There were 56 respondents (3%) who indicated that they spent longer than 70 working days dedicated to working on TAGs, most of which were heads of departments.

When compared to the 2020 CAG survey, respondents indicated that, on average, they spent twice as long on the TAG process than the CAG process (15 days versus 7 days). One factor that could partly explain this increase is the shift in focus to assessing students, rather than predicting future attainment, which may have required the development and marking of additional assessment materials.

The CAG survey also indicated that the time spent on deciding CAGs varied by role type, particularly, more senior roles spent around twice the time on determining CAGs than those who identified as a teacher or tutor. We found that the time spent working on TAGs also varied by role. Those who held the role of 'Exam Officer' or 'Data Manager or Analyst' tended to spend the most time on TAGs (median of 25 days). Deputy or assistant heads of centre and those with other senior leadership roles also tended to spend more time on TAGs (20 days) than the average across all respondents. We note that the overall median time spent working on TAGs is, of course, skewed due to the uneven number of respondents for each role type. It is therefore unsurprising that the median time spent on TAGs for heads of departments matches the overall median (15 days). The median time spent on determining TAGs for teachers, tutors, and trainers was 10 days.

In terms of GQ subjects, the differences between subjects were small and not significant. Additionally, there was very little difference in time spent on TAGs by centre type.

Q. How well prepared do you think your students will be for progression to employment/further learning next year?

Respondents were asked to rate how well prepared they felt their students were for next year, either in terms of employment or further learning, using a 5-point scale from 'very well prepared' to 'very poorly prepared' with a 'don't know' option. The percentage share of responses is shown in Figure 23.

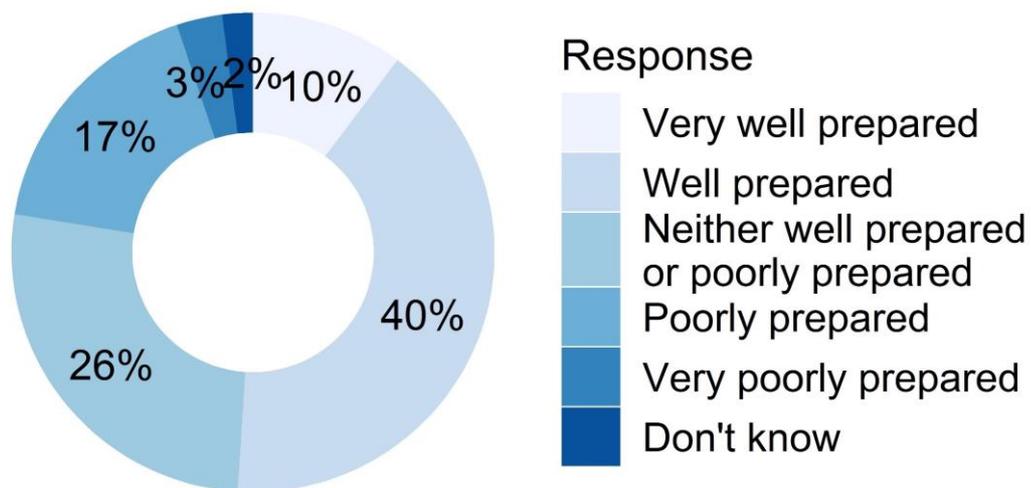


Figure 23: How well prepared do you think your students will be for progression to employment/further learning next year? (N = 1,779)

Most respondents thought their students were well prepared or very well prepared (combined: 51%) compared to 21% who felt that they were poorly prepared or very poorly prepared. Just over a quarter of respondents (27%) felt that students were neither well prepared or poorly prepared, and 2% didn't know.

We found that there was variation amongst respondents when analysed by their centre type. Respondents from independent schools were most confident with 64% indicating that they felt their students were either well prepared or very well prepared for progression. This was followed by respondents from secondary selective schools (56%). Respondents from colleges, secondary, and academy schools were less

confident with less than half of respondents indicating that their students were well prepared (49%, 48% and 43% respectively). This trend is also shown in the figures for respondents who felt their students were either poorly or very poorly prepared (independent 13%, secondary selective 15%, secondary 22%, college 25%, and academy 25%).

Q. If there is anything else you wish to tell us about the TAG process, please do so below.

Respondents were able to provide details on anything else they wished to tell us about the TAG process. Additional details were provided by 1,051 respondents. All were analysed and coded into several broad themes discussed below. We note that many of the themes mentioned in response to this question came up previously in response to other questions, but this question presented an opportunity for respondents to provide more detailed responses or to re-iterate previous thoughts.

Probably the primary concern was that of the significant workload and consequent stress that determining TAGs had produced, expressed in various forms by 46% of respondents (n = 481). Because of the need to create, run, mark and moderate assessments, many respondents described the toll this had taken on them and their colleagues. Some of these reflections also related to points that follow regarding timescales around decision-making and release of guidance or materials, and the feeling of inadequate support from AOs. Quite a number described how this whole process detrimentally affected their wellbeing and mental health and some talked about thoughts that they or their colleagues had regarding leaving the teaching profession because of this.

It was also commonly felt that teaching staff would not get any recognition for their considerable efforts, and 10% of the comments (n = 108) mentioned this and various aspects of blame or criticism teachers might face. This included criticism from the media and government for potentially inflated grades, or criticism from parents and students if they were not satisfied with their grades. Included in this were observations that the name of the process was chosen to put the blame on teachers, and that decisions had been made with no thought for staff.

Another major theme was that of perceived inconsistency of process and grading standards between, and to a lesser extent within, centres (19%, n = 203). This was a mixture of reports heard from a media or online sources, and their own discussion with contacts in other centres. There was a strong feeling from the individuals that completed the survey that they and their centre had been very thorough and done the best job they could to determine fair TAGs, but that other centres they had heard about had not been anywhere near as careful. Concerns included using less controlled sources of evidence and providing unacceptable levels of assistance to

students in the form of pre-warning of the content of tests and allowing multiple attempts at them. A small minority of respondents did discuss concerns with the process within their own centres.

Differential grading standards in terms of over-generous grades from other centres were discussed, with reflections on the unfairness of this for their own students. Some concerns here also touched on the looseness of the guidance allowing approaches that were too varied between centres. The availability of AO assessment materials online was also noted as causing inconsistency between centres and students, as some students would have a greater ability and motivation to search out and practice those test items and learn the mark schemes.

Comments regarding the decision-making process (42%, n = 440) were also common. There were many who felt that external policy decisions, such as the decision to cancel exams and subsequently replace them with TAGs, were made too late. This, they felt, left centres, teachers, and students uncertain of what was going to be required or expected of them and made planning difficult. There was also frustration with a perceived lack of planning for the 2021-2022 academic year and uncertainty with what was coming next. There were some who indicated that, even if exams were cancelled in future, the use of TAGs (at least in their 2021 form) was unsustainable and the process should never again be repeated. A small number of respondents also felt that because the process used felt quite similar to external exams, that these should have been able to go ahead as planned.

Many respondents (18%, n = 188) said that they found the communications and guidance from the Department for Education, Ofqual, JCQ and the awarding organisations to be unclear, vague, and inconsistent or even contradictory. It was noted that some documents or materials provided were either repetitive or just too long-winded to be easily read and absorbed. Some respondents noted that there should have been better cohesion with the advice given and collaboration between those providing it. Many also felt that the Ofqual guidance on what evidence to collect should have been more prescriptive to increase consistency between centres.

Many respondents also commented on what was perceived to be inadequate support provided by awarding organisations (41%, n = 433). The main concern was the provision of questions to help with student assessment only from existing exam papers, which candidates had often already seen. Respondents noted that the repackaging of these materials did not make them more useful, and in some cases the way they were provided (as non-editable documents or images) made them even less useful. Making assessment materials accessible to all was also criticised, as students could access and practice answering the questions. The release of the most recent papers that would normally have remained restricted compounded this issue since there were no 'unseen' papers left to use or select items from. They felt

strongly that AOs should have provided new, unseen, material with which to assess their students, especially since the 2021 papers would have been prepared.

Respondents also noted how, from a GQ perspective, they felt they had to take on the role of an examiner and do the job of awarding organisations without the training or recompense needed to take on this role. Additional support in marking, moderating and internal quality assurance would have been appreciated – quite a few reflected that exam boards should have been marking some or all of the assessments themselves, or at least carried out full moderation. The issue of determining grades from assessment marks alone was also discussed, particularly around the difficulty of using the grade descriptors and marking exemplars awarding organisations had provided. A large number of negative comments regarding the issue of fees paid to exam boards were also made.

A number of specific difficulties with the actual process were described by 10% of respondents (n = 106), with many talking about how difficult it was to balance the responsibility of being both teacher and examiner. The amount of assessment carried out left some with a sense that exams had been cancelled but then took place all the same because of the need for evidence. Some specific problems were also mentioned with VTQs, particularly functional skills, where students were sometimes not able to receive TAGs.

The difficulties discussed also fed into a recognition (16%, n = 168) of the stress that students had experienced, and the breakdown in the normal relationship between teachers and students. A number also talked about the impact on student learning and preparedness of both the Covid-related disruption, and the amount of time and effort spent on TAGs. This related to both school years receiving TAGs, and the school years that will receive qualifications in 2022.

Six per cent of respondents (n = 58) also talked about their expectations for grade inflation this year, and the reasons for this, both legitimate and illegitimate in their view. Mention was made of the anticipation of the results day, and particularly reflecting on possible large number of appeals.

Some respondents (8%, n = 87) discussed their ideas for potential improvements to the TAG process in case it was ever needed to be repeated. Much of these were direct responses to issues they raised, for example, requiring awarding organisations to release unseen assessment material, and the need for a much more prescriptive process to ensure greater comparability between centres. However, other suggestions were broader and included the need for much quicker decision-making with earlier release of plans and guidance. Mention was made of introducing non-exam assessment elements (for example, coursework) into subjects and establishing provision for external assessment or moderation (either with awarding organisations or local networks of centres) if exams were cancelled.

Finally, some positive aspects of the TAG process were also noted (7%, n = 70), often as part of more generally critical comments. These respondents noted that TAGs reduced stress for some students, particularly those who normally struggled with anxiety or performance in examinations. Some also felt that specific aspects of TAGs, such as empowering teachers to provide input and judgement on the performance of their students, and some specific skills learnt through the process could be put into practice more generally in future years.

We note that this was a self-selecting sample of respondents who chose to complete the survey and answer this open response question in detail. It is difficult to be certain that the relative frequency of these views in the wider teaching population is the same as that stated here. However, the number of times these ideas were expressed, suggest these were widely held perceptions.

TAG student survey results

In total, 550 students completed the survey. Partially completed responses were not saved. Similar to the teaching staff survey, most questions except some key routing ones were optional, so respondents were free to choose not to answer. Therefore, the number of responses varies across questions. As before we state the number of responses for each question as (N = xxx) and, where appropriate, the number of responses for any options provided in a question as (n = xxx). The results are presented in sections relating to different aspects of the process.

Demographic details

All respondents to the survey were presented with the demographic questions in this section.

Q. Which school or college year were you completing this summer?

Respondents indicated their school year, and the counts and percentage of the sample are shown in Table 6.

Category	Count	Percentage
Year 13	243	46%
Year 12	26	5%
Year 11	239	45%
Year 10	18	3%
Adult learner	5	1%
Other	3	1%

Table 6: Number of respondents by school year (N = 534)

This sample is representative of the year groups for which TAGs were required this year, with year 13 (n = 243) corresponding to A level and equivalent, and year 11 (n = 239) to GCSE and equivalent. The small number of year 10 and year 12 respondents will largely be students receiving TAGs for VTQs, or also AS TAGs for year 12.

For some of the questions that follow we analyse response patterns across different year groups. To do this, responses of year 10 and year 11 were combined to give lower secondary respondents taking, generally, Level 2 qualifications, as were year 12 and year 13 to give upper secondary respondents, generally taking Level 3 qualifications.

Q. From what type of centre did you receive TAGs?

Respondents indicated their centre type, and the counts and percentage of the sample are shown in Table 1. A slightly reduced list of centre types was presented to students compared to teaching staff. This was to avoid any uncertainty over particular centre type distinctions that may be of less significance to some students than to teachers.

Category	Count	Percentage
Secondary or high school (comprehensive or academy)	267	50%
College (FE college or establishment, Sixth form college, Tertiary college, University Technical College)	106	20%
Independent school	85	16%
Selective school (for example grammar school)	56	10%
Training provider	6	1%
Secondary modern	3	1%
Free school	2	less than 1%
Alternative provision or pupil referral unit	2	less than 1%
Don't know or not sure	10	2%

Table 7: Number of respondents by centre type (N = 537)

For some questions that follow we analyse response patterns across different centre types, but only for secondary or high schools (n = 267), colleges (n = 106), independent schools (n = 85) and selective schools (n = 56), since the other categories had few respondents.

Q. For which general qualifications (GCSE, AS, A level) did you receive TAGs this summer, if any?

Respondents were asked to indicate which subjects and levels they received TAGs for. We have split the data into GCSE (Figure 24) and AS or A level (Figure 25) to make the graphs easier to interpret – please note that the vertical scale is different for each graph.

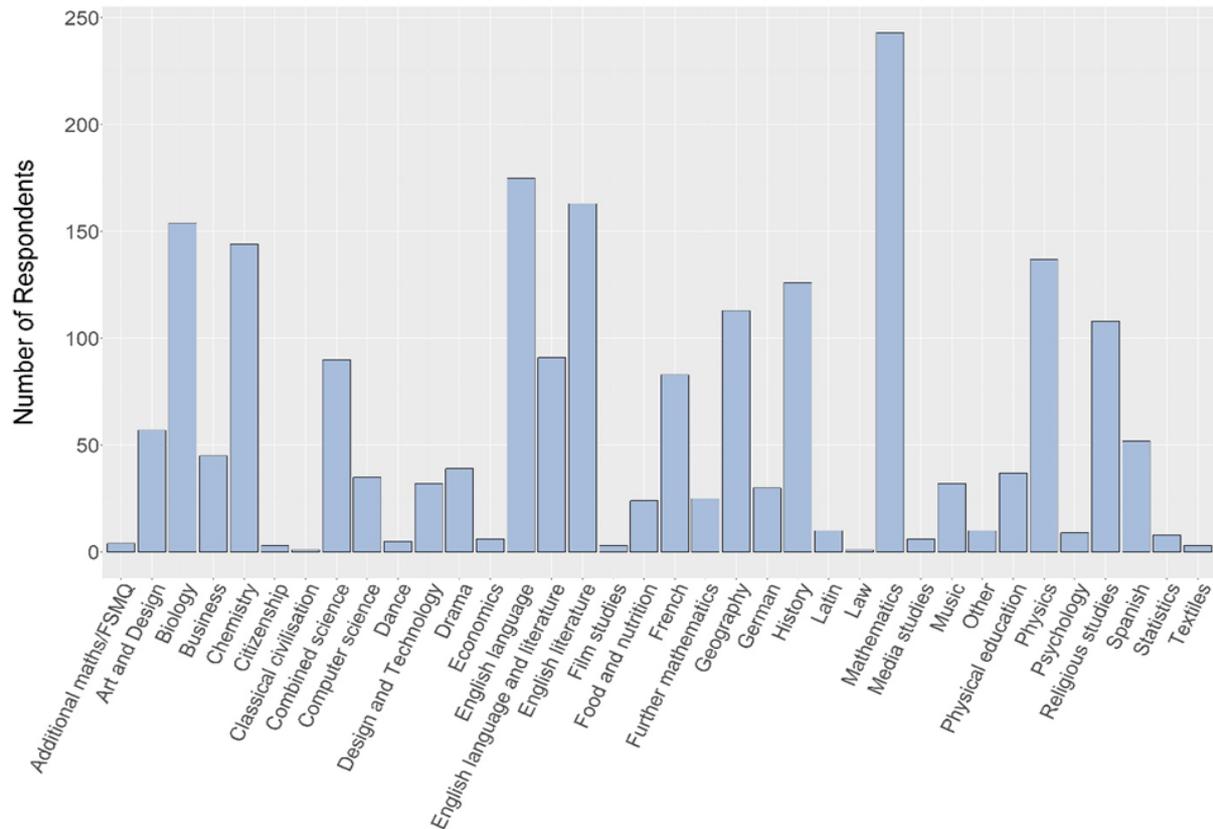


Figure 24: Number of respondents completing general qualifications in each GCSE subject (N = 271)

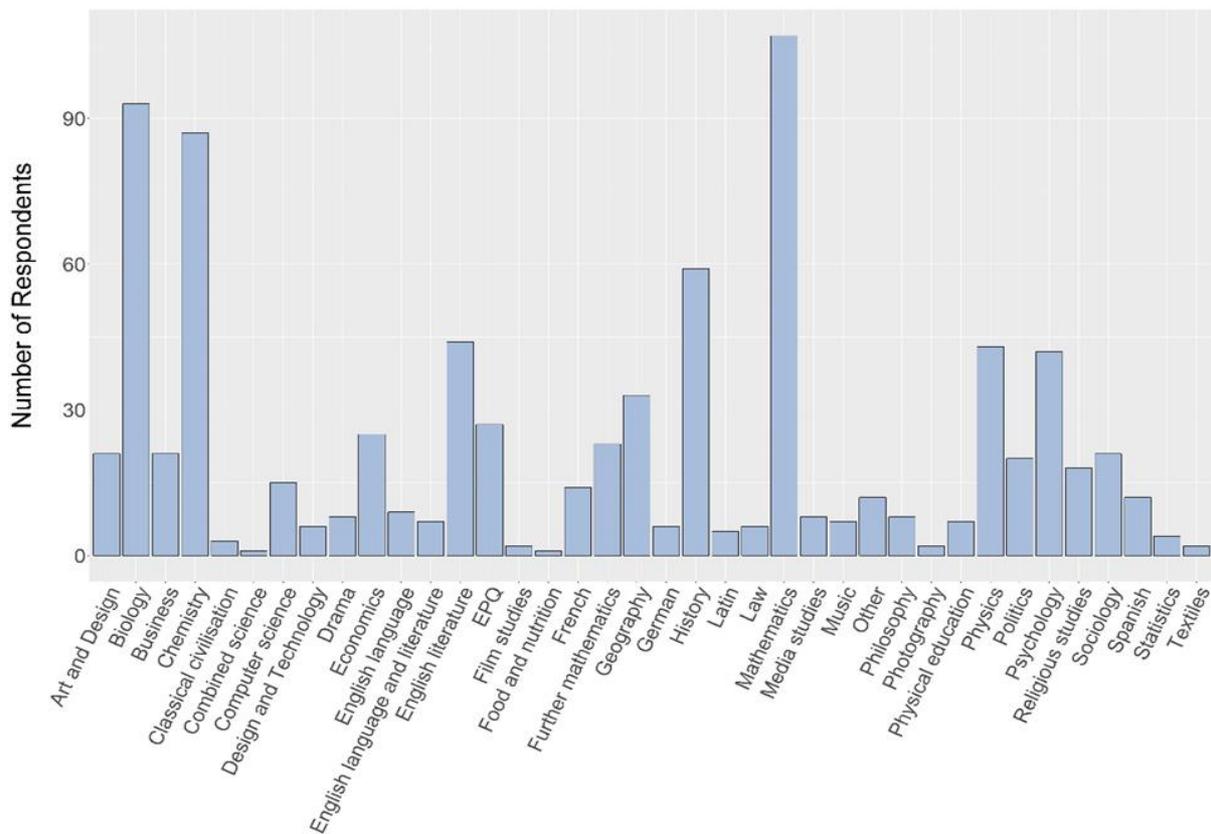


Figure 25: Number of respondents completing general qualifications in each AS or A level subject (N = 264)

The counts were fairly representative of national entries across different subjects at different levels. Mathematics, English subjects and sciences were most frequently reported for GCSE, and mathematics, biology, chemistry and history most frequent for AS or A level. Note that due to survey system constraints, ‘English language and literature’ was available as a GCSE option even though this is not a current qualification. The 91 students selecting this option were probably indicating that they took both English language and English literature GCSEs. Similarly, for combined science at AS or A level, respondents may have selected this option to reflect completing multiple science subjects as it is not an AS or A level qualification.

Q. For which types of vocational and technical qualifications did you receive TAGs this summer, if any?

Respondents indicated which VTQ types they received TAGs for. As shown in Figure 26, the most frequent VTQs for which respondents received TAGs were BTEC qualifications, with 41% (n = 28) of respondents receiving TAGs for BTEC

qualifications at Level 1/2 and 29% (n = 20) receiving TAGs for BTEC qualifications at Level 3.

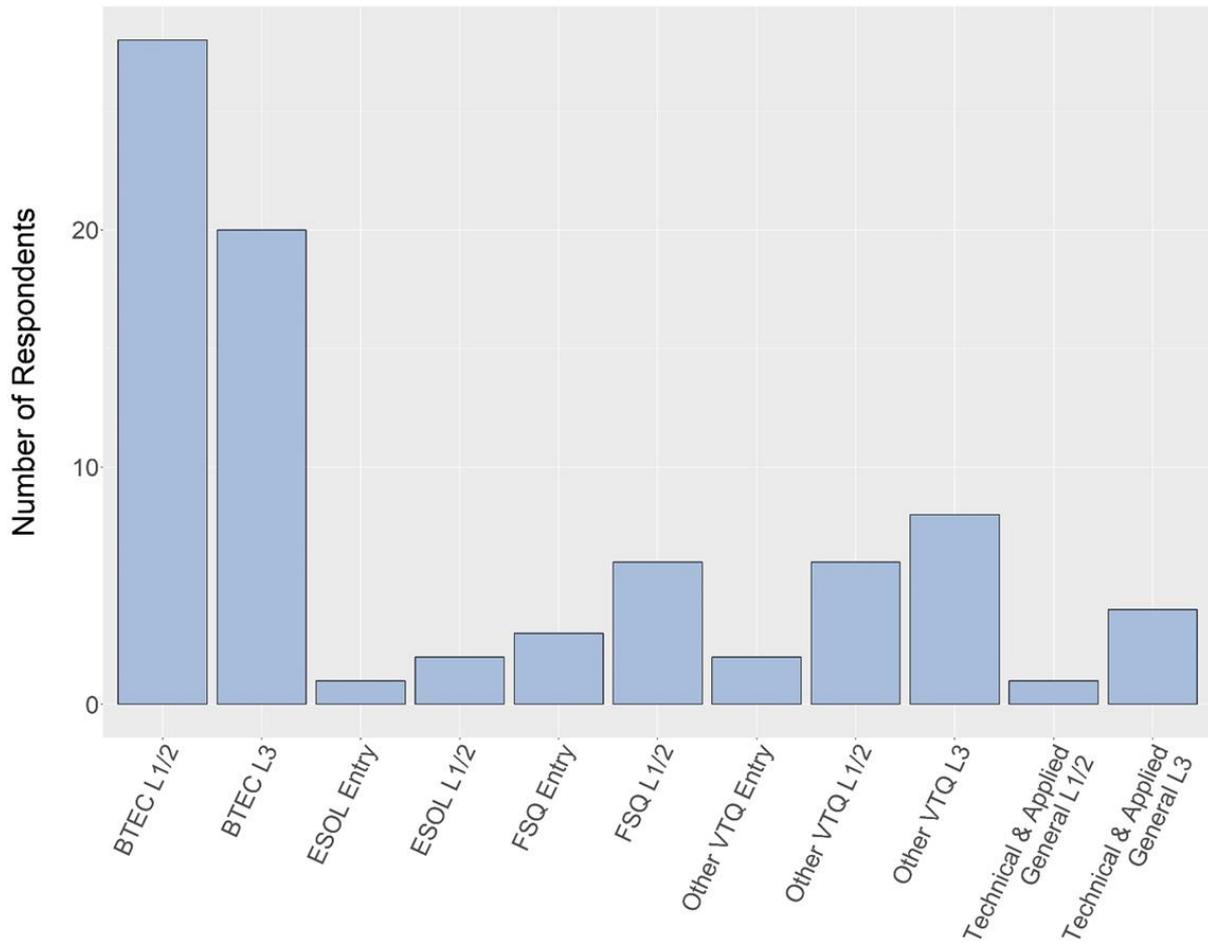


Figure 26: Number of respondents completing each type of vocational qualification (N = 68)

Initial thoughts and feelings

All respondents were asked the questions in this section, relating to their initial thoughts and feelings when assessments were cancelled in January.

Q. Thinking back to when summer assessments were cancelled in January, how did you feel?

Respondents were asked to enter 3 separate words into free text entry boxes to indicate how they felt. Using the same methodology as for the teacher survey (see Figure 17), we analysed the words given by their frequency and plotted them as a word cloud in Figure 27.

These words seem to reflect quite mixed feelings of students when summer assessments were cancelled in January. Some students were glad to have assessments cancelled and some disappointed, and some held a mixture of views. These views are explored further in the following section.

Q. Thinking back to when summer assessments were cancelled in January, how fair or unfair did you believe this cancellation of exams and assessments was for you?

Respondents selected one choice on a 5-point scale from very fair to very unfair plus don't know. Reflecting the words in the previous question (see Figure 27), the respondents were mixed in their views on the cancellation of exams and assessments (see Figure 28). Slightly more respondents (46%) believed the cancellation to be either fair or very fair, compared to 33% who believed the cancellation was unfair or very unfair. In addition to this, 19% of respondents felt the cancellation was neither fair or unfair.

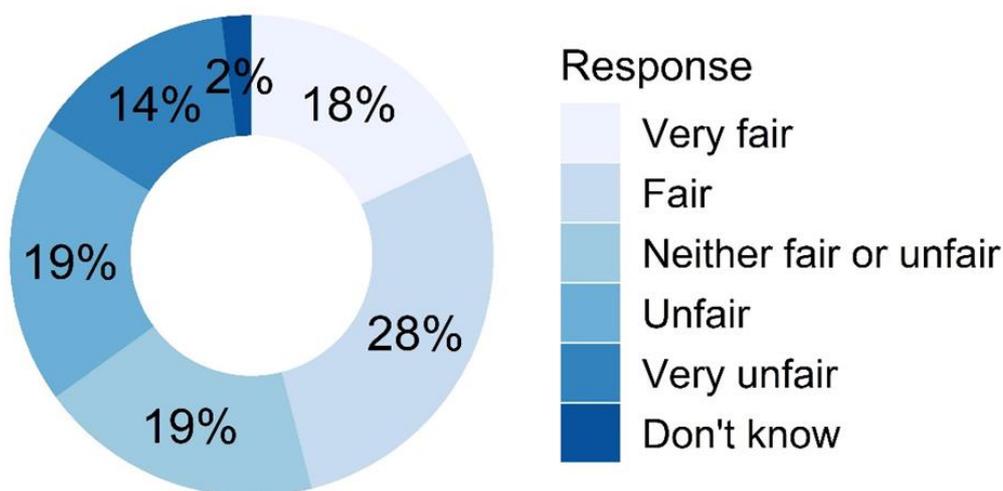


Figure 28: Fairness when summer assessments were cancelled in January (N = 536)

We also considered how fair respondents believed the cancellation of summer assessments in January to be, based on their centre type, but found that views across different centre types were fairly comparable, with the highest proportion of respondents in each centre type considering the cancellation of summer exams as 'fair'.

Q. Thinking back to when summer assessments were cancelled in January, would you have preferred to take exams and assessments as planned?

Respondents selected 'yes', 'no' or 'don't know' (N = 537). Respondents had mixed views on whether they would have preferred to take exams and assessments as planned, when the announcement was made in January that they would be cancelled. Slightly more respondents (44%) indicated that they would have preferred to not take exams and assessments than those who would (38%). These mixed feelings around the cancellation of summer assessments are reinforced in their responses throughout this section.

We explored this further by analysing responses by centre type. For those attending secondary or high schools, colleges or independent schools, more respondents preferred not to take exams and assessments as planned (44-49%) than would have (35-42%). However, for selective schools a higher proportion of respondents would have preferred to take exams (46%) compared to those who would not (38%).

Disruption

All respondents to the survey were presented with the questions in this section, relating to the disruption to their learning that they experienced due to the pandemic.

Q. Across all of your qualifications, approximately how much of the content do you think you were NOT taught?

Respondents (N = 538) selected one choice from a series of options in steps of 10%. The median value was 21-30% of content not taught. The most frequently indicated amounts were 0-10% (25%), 21-30% (21%) and 11-20% (20%). Most of the students therefore felt that they were taught the majority of content for their courses. However, 12% of students reported being taught less than half the content for their course, and 1% indicated less than 20%. We note that this might be quite a difficult question for students to answer, given that they might not have full awareness of the teaching time required for any un-taught content. We consider this further in the Discussion.

We also considered this question in terms of the centre type respondents attended. By removing the 'don't know' responses and using the midpoint of each bin, we were able to generate a mean amount of untaught content (as reported by students) for each centre type. This mean was highest for respondents attending secondary and high schools (mean of 29%), followed by those attending colleges (mean of 25%), those attending selective schools (mean of 24%) and those attending independent schools reported the lowest amount of content not taught (mean of 21%).

Q. Were you able to effectively learn all of the content you WERE taught for your qualifications?

Respondents selected either 'yes', 'no' or 'don't know' (N = 540). A slight majority, 52%, of respondents indicated that they were not able to effectively learn all of the content they were taught. Forty-one per cent of respondents felt that they were able to learn effectively for all the content they were taught, with the remainder (7%) responding that they didn't know.

There were no substantial differences between centre types, but a year group analysis showed that 47% for year 10 or 11 reported that they were not able to effectively learn all of their content compared to 58% for year 12 or 13. Inability to learn all taught content was a bigger problem in year 12 or 13 for our survey sample.

Q. What factors limited how well you were able to learn the content you were taught?

Those respondents stating that they were not able to learn all taught content effectively were asked this question (N = 279). They selected from multiple options listed below, with the percentage selecting each shown in brackets:

- Less effective remote teaching compared to face-to-face teaching in class (77%)
- Difficult to engage with remote learning (was boring) (73%)
- Other disruptions or distractions whilst learning remotely (61%)
- Unplanned school or college closure or being sent home due to Covid cases on site (59%)
- Inability to complete certain learning tasks whilst school or college shut (59%)
- Less remote teaching compared to face-to-face teaching in class (52%)
- Inability to complete certain learning tasks whilst school or college open due to, for example, social distancing (40%)

- Less teaching or contact time while centre open (39%)
- Illness (yours or family) (25%)
- Limited or no access to technology required for remote learning (15%)

Those students who had struggled to effectively learn all taught content perceived that the effectiveness of, and engagement with, remote teaching were the biggest obstacles, together with distractions occurring while learning at home. No substantive differences were seen across centre types or school years.

Details on TAG evidence

All respondents to the survey were presented with the questions in this section, which focused on the evidence, such as work and assessments, that teachers or tutors used to judge TAGs.

Q. Across all of your qualifications, how much awareness did you have of the evidence that was selected by your teachers/tutors to support your TAGs?

Respondents (N = 539) indicated their level of awareness of the evidence used to support their TAGs. Most respondents (89%) had at least some awareness of the evidence that was being used to support their TAGs, although this was predominantly those who had only 'a little' awareness (58%) with only 31% indicating 'a lot' of awareness. A further 11% had no awareness at all. JCQ guidance stated that centres should ensure students were aware of the evidence used to determine grades. It is possible that, for those who had little or no awareness, the information was communicated but not received.

There were some differences between centre types. Secondary and high schools had a higher proportion of students suggesting they had no awareness (13%) and a smaller proportion of students reporting they had 'a lot' of awareness (25%) compared to the other centre types. There were no substantial differences in awareness of the evidence used between year groups.

Q. What kind of evidence types were used to support your TAGs? (For GCSE, AS/A levels and other general qualifications, such as Pre-U and EPQ)

Respondents were presented with a list of evidence types and indicated whether each had been used in support of their TAGs in GQs (N = 539). The percentages selecting each of the options are given in brackets.

- Mock or practice exams – part or whole past papers (88%)
- Class tests (60%)
- Non-exam assessment or coursework (completed) (42%)
- Assessment tasks provided by exam boards (40%)
- Assignments (34%)
- Class work (26%)
- Homework (21%)
- Participation in performances in practical or performing arts subjects (14%)
- Non-exam assessment or coursework (not completed) (8%)

Therefore, TAG judgements were often based on similar forms of evidence to normal summative assessments. Interestingly, the assessment tasks provided by exam boards (40%) was less frequently reported than class tests. It is possible that students were unaware that the tests they sat were constructed from these materials. Five per cent of respondents reported that they didn't know.

Q. What kind of evidence types were used to support your TAGs? (For Vocational and Technical Qualifications)

Respondents were presented with a list of evidence types and indicated whether each had been used in support of their TAGs in VTQs (N = 86). The percentages for each of the options are given in brackets.

- Mock or practice exams – part or whole past papers (43%)
- Other coursework or internal assessment (completed) (41%)

- Assignments (29%)
- Class work (24%)
- Class tests (23%)
- Assessment tasks provided by awarding organisations (21%)
- Participation in performances in practical or performing arts subjects (10%)
- Other coursework or internal assessment (not completed) (10%)
- Homework (9%)
- Group work (6%)

While the most frequent source of evidence reported was mock or practice exams, similar to that for GQs, this had a lower reported percentage (43% versus 88% in GQ) and there were a broader range of sources of evidence used, reflecting the broad range of assessments normally used in VTQs. Nineteen per cent of respondents reported that they didn't know.

Q. What proportion of the evidence selected to support your TAGs was completed BEFORE the announcement in January that exams and other assessments were cancelled?

Respondents selected one choice from the following options:

- Most of it
- A little more than half
- About half
- A little less than half
- Hardly any of it
- Don't know

The most common response was that 'hardly any' of the evidence used to support TAGs was completed before the announcement in January (39%, see Figure 29). Furthermore, the majority of respondents (77%) reported that half or more of the evidence used to support their TAGs was completed after January. This may be a result of JCQ guidance which suggested that more recent evidence is likely to be more representative of student performance, influencing teachers and centres to use evidence gathered before January sparingly.

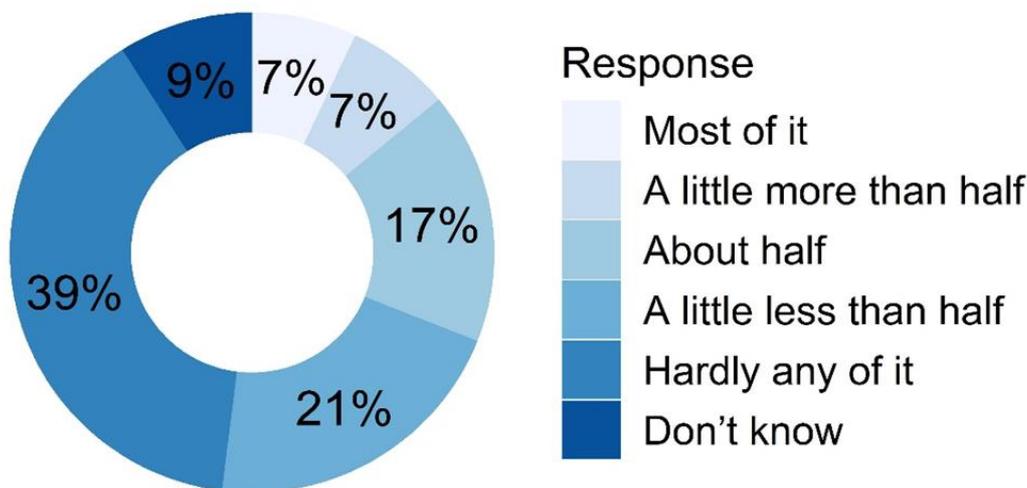


Figure 29: Proportion of evidence completed BEFORE exams and assessments were cancelled in January (N = 538)

For all 4 centre types, respondents most frequently reported that ‘hardly any’ of the evidence was completed before January. Considering respondents that reported half or more of the evidence completed before January (combining response options ‘about half’, ‘a little more than half’ and ‘most of it’), this accounted for 35% of those attending college, 32% of those attending secondary or high schools and 29% of those attending selective schools. Independent school respondents had a smaller proportion, 22%, so were least likely to use older evidence to support TAGs.

Q. Did you sit any tests under exam conditions, such as mocks, to produce evidence for your TAGs?

Of those respondents who answered this question (N = 540), 93% said they sat tests under exam conditions to produce evidence for their TAGs. A very small number did not sit any (6%), perhaps representing those who were taking qualifications largely assessed by coursework or NEA. This reflects the nature of the normal assessment arrangements for most of the qualifications for which TAGs were required, particularly GQs.

Q. For what proportion of these timed tests under exam conditions did your teachers/tutors tell you in advance (at least the day before) what would be on the test?

Respondents selected from the following options, with the percentage selecting each option in brackets:

- All (18%)
- A lot (19%)
- A few (36%)
- None (23%)
- Not sure (4%)

The most common response was that respondents were given advance notice for a few timed tests (36%). There was a great deal of variation across respondents though, with 18% being told about the content in advance for all of these tests and 23% being told about the content for none of their tests. We did not ask for further detail, so it is not clear how specific the information given to students was. For some this may have simply been the general topic areas that would be covered while for others this could potentially have been specific questions. There were no substantive differences across centre types. The most frequent response for all of the centre types was 'a few'.

Q. On average across all of these timed exams, how thoroughly were you able to revise/prepare?

Respondents (N = 504) were asked to indicate how thoroughly they had been able to revise or prepare for exams.

- As thoroughly as if they were actual summer assessments (16%)
- Fairly thoroughly (47%)
- Not very thoroughly (38%)

Respondents did not in general think they had been able to prepare as thoroughly as they would for normal summer assessments. There were differences between centre types for those who reported they could not prepare very thoroughly. Those

attending selective schools had the largest proportion (44%) followed by secondary or high school (40%), colleges (34%) and independent schools (30%).

Q. Do you think you sat more tests/assessments to support your TAGs than you would have if the normal exams and assessments had not been cancelled?

Respondents selected either 'yes', 'no' or 'don't know' (N = 540). The majority of respondents, 68%, stated that they sat more tests and assessments to support their TAGs than if they had done normal exams, while 24% disagreed and the remainder (8%) did not know. This may explain the number of respondents reporting that they were not able to properly prepare for their tests and assessments, due to the number that they were required to sit.

For all centre types, respondents most frequently reported that they sat more tests than they would have if exams had not been cancelled, but there was variation across centre types. The proportion reporting they sat more assessments to support TAGs in each centre type was 85% for independent schools, 70% for secondary or high schools, 56% for colleges and 55% for selective schools.

Q. To what extent do you agree or disagree with each of the following statements about the evidence collected to support TAGs?

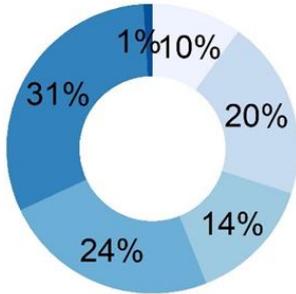
Respondents were presented with several statements, each requiring a response on a 5-point Likert scale from 'strongly agree' to 'strongly disagree' plus 'don't know':

- Because there were other opportunities to perform better, I did not always try as hard as I could for every test or assessment.
- I was pleased to have the opportunity to show how well I could do.
- I was stressed about the amount of tests or assessments I was asked to complete.
- Overall, across all my qualifications, I believe the evidence used to support my TAGs fairly reflected my knowledge and skills.

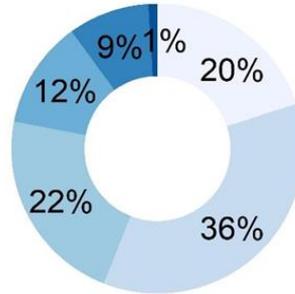
Whilst not intended to be comprehensive, these statements were devised to represent some of the common views that students may have had about the tests and assessments they sat.

The most consistently-held view was that 79% of respondents strongly agreed or agreed that they were stressed about the amount of tests or assessments they were asked to complete, with 59% strongly agreeing (see Figure 30). This is consistent with the previous question showing that generally respondents reported that they completed more tests than they would have, had normal exams and assessments gone ahead.

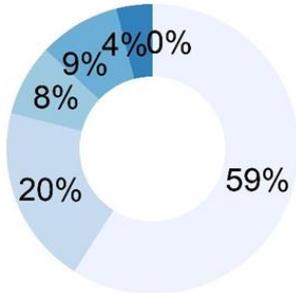
Because there were other opportunities to perform better, I did not always try as hard as I could for every test/assessment.



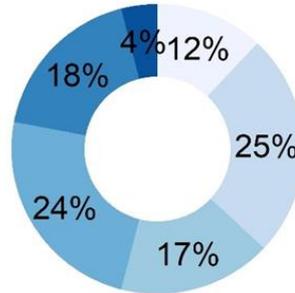
I was pleased to have the opportunity to show how well I could do.



I was stressed about the amount of tests/assessments I was asked to complete.



Overall, across all my qualifications, I believe the evidence used to support my TAGs fairly reflected my knowledge and skills.



Response

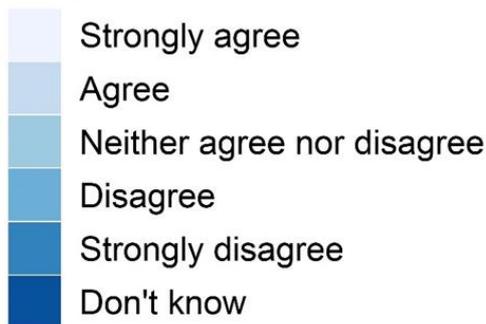


Figure 30: Agreement with statements about the evidence collected to support TAGs (N = 540)

Considering the question of whether having multiple assessments meant that students didn't always try as hard as they could, the majority disagreed or strongly disagreed (55%) compared to 30% who strongly agreed or agreed. Despite often having to sit more tests than usual, many students tried as hard as they could for all of their assessments. Fifty-six per cent of respondents agreed or strongly agreed that they were pleased to have the opportunity to show how well they could do. There were mixed responses to whether respondents believed the evidence used to

support their TAGs fairly reflected their skills and knowledge. Thirty-seven per cent agreed or strongly agreed, compared to 42% who disagreed or strongly disagreed.

Between centre types there were some differences. For 'I was stressed about the amount of tests or assessments I was asked to complete' the sum of those who agreed or strongly agreed, independent schools had the highest percentage at 86%, then secondary schools at 83%, and selective schools at 79%. Those who attended colleges reported slightly lower stress, with 72% of respondents agreeing or strongly agreeing. Perhaps this again reflects the types of qualifications taken in colleges where more ongoing assessment may have been used as evidence for TAGs and so fewer tests may have been sat.

For 'overall, across all my qualifications, I believe the evidence used to support my TAGs fairly reflected my knowledge and skills', respondents attending selective schools less frequently reported that they agreed or strongly agreed with this statement (26%) compared to respondents attending secondary schools (38%), colleges (42%) and independent schools (43%).

Q. Please add any comments you have about how well the evidence selected to support your TAGs reflected your knowledge and skills.

Respondents were asked to give additional comments regarding whether they felt the evidence selected to support the TAGs was a fair reflection of their knowledge and skills. In total, 190 respondents gave additional comments. Many of these responses were comments around ways in which the evidence used was not fair for them. The use of evidence that was collected before the announcement that exams were cancelled was mentioned by 15% of respondents (n = 28). These respondents suggested that, where they had undertaken assessment with no awareness that it would count towards their final grades, they had expended less effort, and this was considered very unfair to include. Included in this count were also a few comments reflecting on the use of remote assessments sat under less well controlled conditions.

Only eight respondents (4%) made comment about being assessed on content they had either not been taught, largely where full past papers were used, or had not been able to effectively learn due to remote teaching. However, a substantial number of respondents (24%, n = 45) referred to the sheer number of assessments they completed in a short space of time and the impact this had on them and their performance. Many reflected on their inability to revise adequately due to either short-notice given for tests, or lack of time between each test. Others commented on

the stress that this large number of assessments had caused and how this may have caused them to under-perform.

Many of the comments reflected that students thought that they would have performed better in live assessments since they would have occurred later in the year than when the TAG evidence was collected. A small number of comments were received about sitting tests under exam conditions, specifically that it had been difficult to concentrate when doing so during normal classes.

A few students expecting high grades (3%, n = 5) were frustrated by only sitting tests that were either quite short, restricted to only basic content, or where questions were known (and possibly practiced) by students in their class where the AO-materials had been open to all. They felt they had been limited in their ability to demonstrate exceptional performance. The degree of access to test materials fed into a number of comments (7%, n = 14) regarding what students described as 'cheating' in other centres. Concerns focussed on the tests used by centres and whether students had already practiced or learnt the mark schemes for those questions, and the way centres may have let their students know what was on the tests. There were also a variety of general observations over the lack of comparability and standardisation across centres (11%, n = 21).

There were quite a few opposing views on specific aspects of evidence selection. Criticism of their centre basing TAGs on final exams alone was made by 18 respondents (9%), as they felt that this ignored their ongoing performance and over-weighted any under-performance in those particular exams. In contrast, 10 respondents commented on how some evidence from assessments completed during the post-announcement period was included but had not been completed to the same standard since it had been undertaken remotely or under disrupted conditions. Some students commented they knew that evidence had been used that was simply not reflective of their abilities (4%, n = 7) and there were also more general comments about the inappropriateness of using the same evidence across all students, rather than picking specific evidence for each student (2%, n = 4).

Specific concerns around circumstances the students had experienced during assessments, or the use, or lack thereof, of special considerations and access arrangements were reported by 6% of students (n = 12). Specific unfairness for students that might expect to show significant improvement towards the end of the year was noted, given that TAG evidence was collected earlier than the final exams would have been sat. Issues with the undue difficulty of the bespoke tests created by teachers was also raised, as was the perceived unfairness around the use of incomplete NEA or NEA that was undertaken under difficult working conditions (this was specific to creative subjects).

Six general comments (3%) around fears of biased or unfair teacher judgements were received, and 3 students (2%) commented on the unfairness of limiting grades

through the use of past centre performance. Fifteen students detailed how they had not been informed by their centre of what evidence had been used, and several perceived that this was not consistent with the guidance centres should have been following.

Finally, there were some positive comments about the TAGs, with 10% of students (n = 19) reflecting that the evidence used was a fair reflection of their ability, and 7 (4%) specifically mentioning the greater number of opportunities to show what they could do, compared to normal assessments. Five students (3%) also reflected that overall, the TAG process had been easier and less stressful than normal exams for them.

Q. To what extent do you agree or disagree that the judgements made by your teachers/tutors on which grade to award you, were made only on content that you had been taught?

A 5-point Likert scale from 'strongly agree' to 'strongly disagree' plus don't know was used. Respondents (N = 539) usually agreed that the judgements made by their teachers or tutors were made only on content they had been taught, with 63% reporting that they agreed (40%) or strongly agreed (23%). The JCQ guidance on grading for teachers focused on considering only what had been taught, so this may explain why the majority of students experienced this. Note that 13% disagreed and 6% strongly disagreed with this statement. Some students may have been considering content that was taught but that they missed due to, for example, illness, or teaching that they did not feel adequately delivered the content. Six per cent were not sure and 12% did not agree or disagree.

Across centre types, the percentage agreeing or strongly agreeing varied from 72% of those attending independent schools, 68% of those attending colleges, 67% of those attending selective schools to 58% of those attending secondary or high schools.

Your views on fairness of TAGs

This section asked all respondents about their views on the fairness of TAG judgements made by their teachers and tutors. We note that these perceptions were reported before results were issued, so students did not yet know their TAGs.

Q. Overall, do you think the TAGs will be fair or unfair for you?

A 5-point Likert-type scale was used from 'very fair' to 'very unfair' plus 'don't know'. Fifty-three per cent of respondents felt the judgements made by their teachers or tutors would be fair or very fair (see Figure 31), compared to 24% who felt they would be unfair or very unfair. This suggests students were reasonably confident that the judgements teachers made would be fair for them. This was similar to the views of teachers, who also believed their judgements were fair (Figure 21).

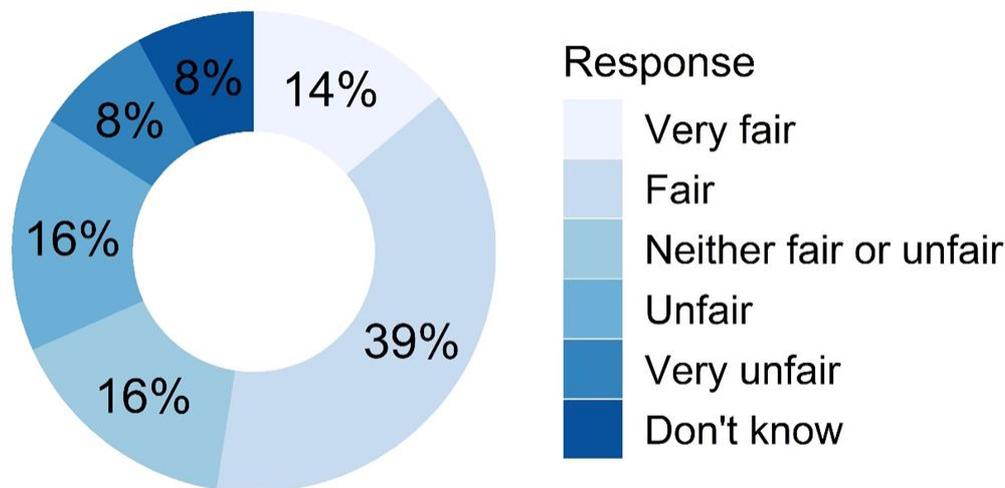


Figure 31: Overall fairness of TAGs (N = 539)

For all centre types, 'fair' was the most common response. Combining the responses of 'very fair' and 'fair', independent school students had the highest percentage (61%), then colleges (54%), secondary schools (51%) and finally selective schools (43%).

Q. To what extent do you agree or disagree with each of the following statements?

The following statements were presented and rated by students using a 5-point Likert scale from 'strongly agree' to 'strongly disagree' plus 'don't know':

- My relationships with my teachers or tutors may have affected the grades they judged for me.

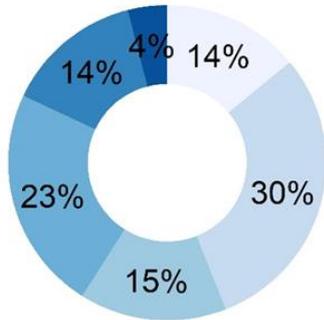
- I believe I will have been judged fairly compared to other students in MY school or college.
- I believe I will have been judged fairly compared to students in OTHER schools or colleges.
- The work and assessments I completed will have had a real impact on the grades I receive.

These statements were devised to represent some of the main concerns that students may have had around fairness.

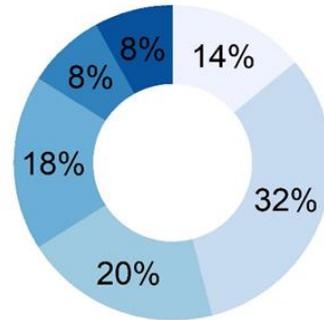
In terms of whether students felt their relationships with teachers or tutors would affect their grades, 43% of respondents agreed or strongly agreed that this would affect the grades judged for them (see Figure 32). However, 37% of respondents disagreed or strongly disagreed with this. Therefore, a substantial number of students did think this factor may have made a difference, suggesting they feared a degree of bias in TAG judgements. Given the lower number of respondents reporting that they thought they had been judged unfairly in the previous question, this may reflect students who thought they may have been judged both harshly and generously.

Forty-six per cent of respondents agreed or strongly agreed that they had been judged fairly compared to other students in their school or college, compared to 26% who disagreed or strongly disagreed. Considering whether students felt they were judged fairly compared to students in other schools or college, 54% of the respondents disagreed or strongly disagreed. Therefore, it seems that while more students felt their own grades are likely to be fair, there was greater concern around comparability of grades across the country. Students agreed that the work and assessments they completed had a real impact on the grades they would receive, with 80% agreeing or strongly agreeing with this statement.

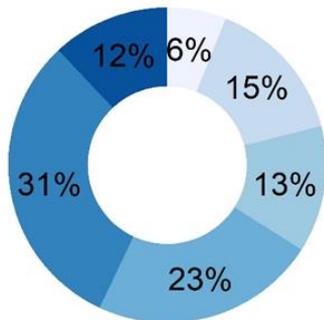
My relationships with my teachers/tutors may have affected the grades they judged for me.



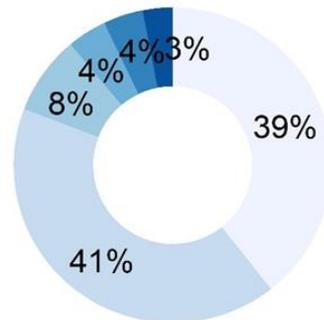
I believe I will have been judged fairly compared to other students in MY school/college.



I believe I will have been judged fairly compared to students in OTHER schools/colleges.



The work and assessments I completed will have had a real impact on the grades I receive.



Response

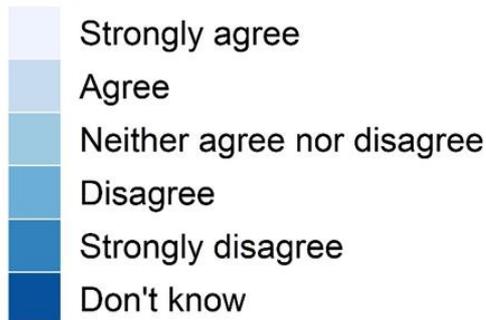


Figure 32: Agreement with statements around fairness (N = 539)

We found no substantial difference between centre types except for responses to 'I believe I will have been judged fairly compared to students in OTHER schools or colleges'. Here, more respondents attending selective schools disagreed or strongly disagreed (80%). By way of comparison, this figure was 48% for those attending colleges, 52% for those attending secondary or high schools and 59% for those attending independent schools.

Q. Do you think you would have achieved better or worse grades if you had been able to take exams and assessments normally?

Respondents selected one option from a 5-point Likert-type scale from 'much better' to 'much worse' plus don't know'. Thirty per cent of respondents felt they would have achieved better and 21% felt they would have achieved much better if they had been able to take exams or assessments normally, making up 51% of responses (see Figure 33). Twenty-six per cent felt they would have achieved about the same and only 16% felt they would have achieved worse or much worse if they had sat exams. Students generally believed their TAGs would lead to lower grades than if they had sat exams.

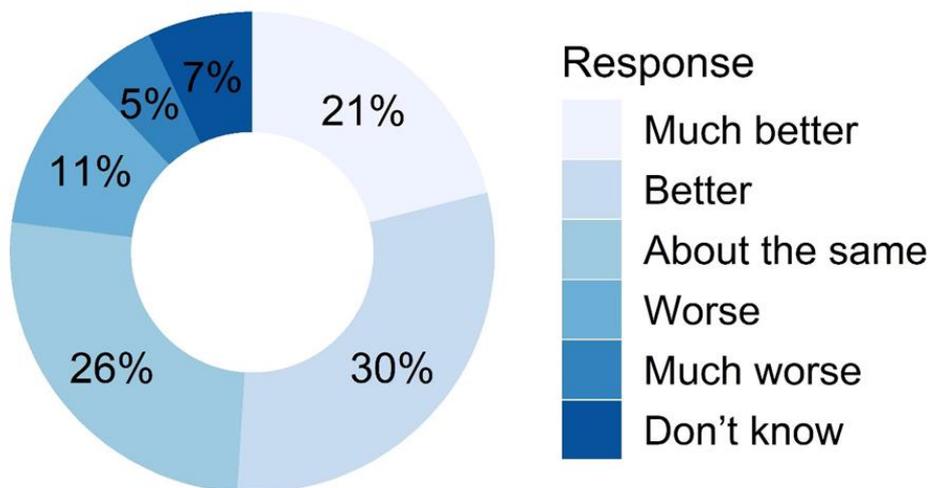


Figure 33: Achievement of better or worse grades if exams had been sat as normal (N = 540)

Reasonable adjustments

All respondents completed this section, relating to whether or not they received reasonable adjustments.

Q. Would you normally receive reasonable adjustments (or access arrangements) such as extra time or assistive technology?

Respondents were asked to respond 'yes', 'no' or 'don't know' to this question (N = 539). In most cases (73%) they indicated they did not normally receive reasonable adjustments while 3% did not know. The remaining 24% of respondents who reported that they did normally receive reasonable adjustments were shown the other questions in this section.

Q. How was the reasonable adjustment applied for the evidence collected to support your TAGs?

Those respondents (N = 130) who indicated that they normally received a reasonable adjustment were then asked about how this reasonable adjustment was applied. They were asked to choose from the following 3 options and the percentages selecting each are given in brackets:

- Available for all assessments (45%)
- Taken into account when judging TAGs (5%)
- Both, depending on the subject or assessment type (19%)
- Don't know (31%)

For the most part reasonable adjustments were available during assessments. This is consistent with responses to the similar question for teachers. However, it is worth noting the high proportion of respondents who selected 'don't know' (31%). Given that a student would know whether the adjustment was available in an assessment these may be cases where they were taken into account when judging TAGs but students may not have been made aware of this by their teachers.

Q. Do you agree or disagree that your reasonable adjustment will have been fully and fairly taken into account when your grades were judged?

Respondents selected one option from a 5-point Likert scale from 'strongly agree' to 'strongly disagree' plus 'don't know'. There were fairly mixed views around whether reasonable adjustments were fully and fairly taken into account when the grades were judged (see Figure 34). The most common response was that students agreed they had been, with 42% agreeing or strongly agreeing, although 31% disagreed or strongly disagreed with this. Some students were therefore not satisfied that they would be treated fairly with regards to their reasonable adjustment.

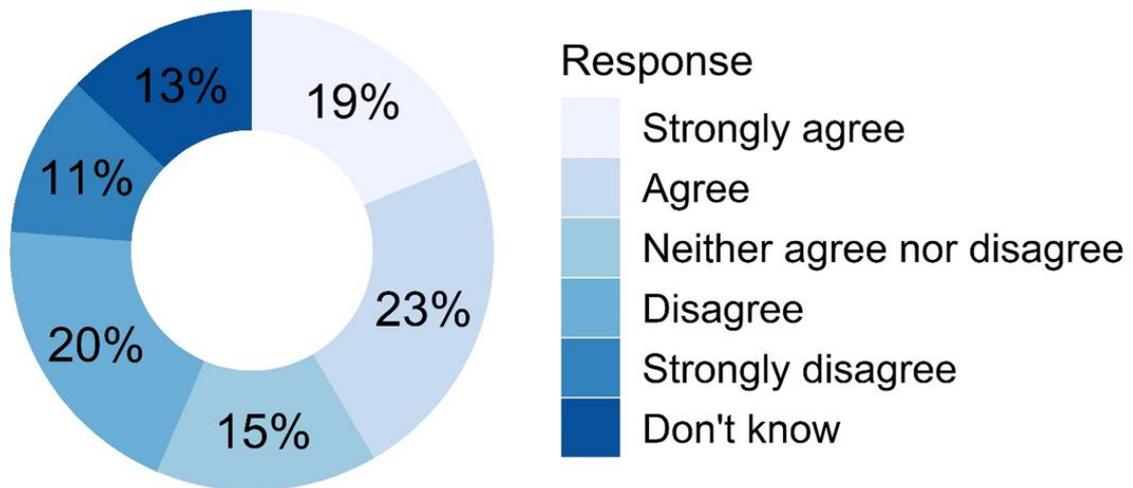


Figure 34: Agreement that reasonable adjustment will have been fully and fairly taken into account (N = 131)

Balance between carrying out assessments to support TAGs and further learning of new content

All respondents to the survey were asked the questions in this section, which focused on the balance between time spent on assessments that were completed to support TAG judgements, and the learning of new content.

Q. Between returning to school after Easter to mid-June, approximately what proportion of your time in school/college did you spend completing assessments or work to support TAGs?

Respondents selected one choice from a series of options in steps of 10%. Excluding 'don't know' responses, the median percentage time spent completing assessments and work to support TAGs was 71-80%. The most commonly selected percentages were 91-100% (19%) and 81-90% (17%) (see Figure 35). Many students therefore perceived that their centre had concentrated on assessment over teaching new material during this period. There was no significant difference between centre types.

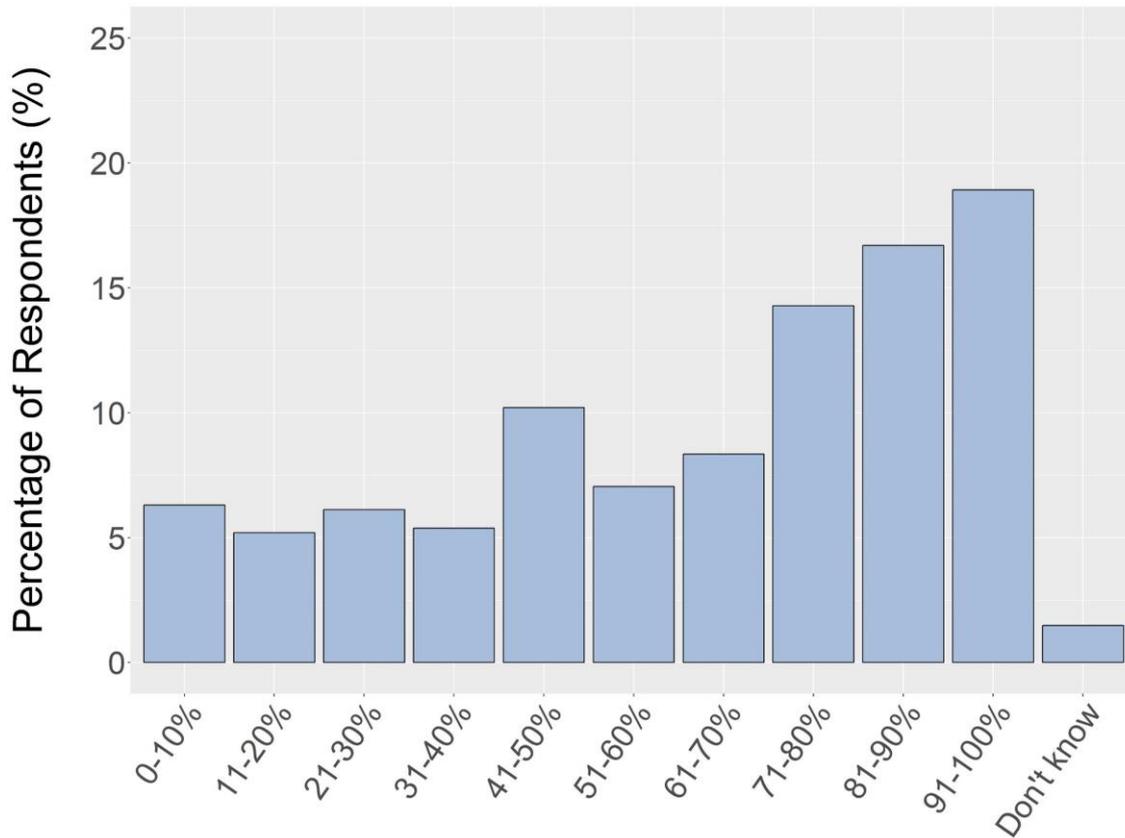


Figure 35: Proportion of time in school/college spent completing assessments or work to support TAGs after Easter (N = 539)

Q. Do you think the balance between time spent on assessments to provide evidence for TAGs and learning new content/materials was appropriate?

Respondents (N = 539) selected one choice from the following options, with the percentage selecting each given in brackets:

- Far too much time spent completing assessments (25%)
- Too much time spent completing assessments (32%)
- Balance between time spent on assessments and learning new content was appropriate (24%)
- Too much time spent on learning new content (9%)
- Far too much time spent learning new content (4%)
- Don't know (6%)

When considering whether the balance between time spent on assessments and learning new content was appropriate, the respondents felt there was generally too much time spent on assessments (57% 'too much' or 'far too much'). Students in independent schools felt this most strongly (66% 'too much' or 'far too much'), 59% for secondary or high schools, 55% for selective schools and college students the least (48%).

The future

This section was shown to all respondents. The focus of this section was views on how prepared respondents felt for their next step after completing these qualifications, while noting that these perceptions were reported before they received their TAGs.

Q. To what extent do you agree or disagree that you are well prepared for your next step in education or employment/training?

Respondents selected one option on a 5-point Likert scale from 'strongly agree' to 'strongly disagree' plus 'don't know'. More respondents thought that they were prepared for next year than were not (see Figure 36). Forty-six per cent agreed or strongly agreed that they were well prepared, but 33% disagreed or strongly disagreed.

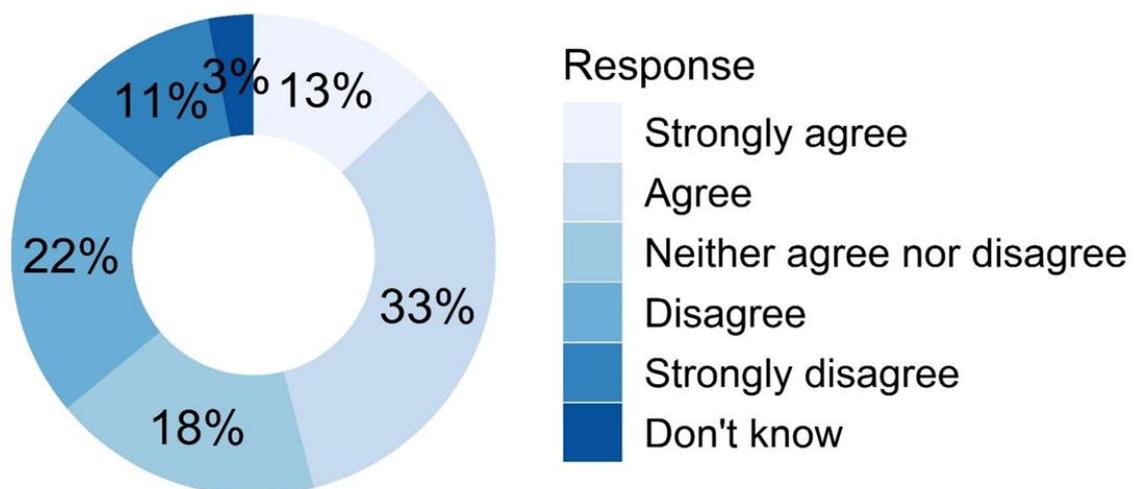


Figure 36: Preparation for next steps in education or employment/training (N = 538)

Respondents attending independent schools reported feeling most prepared for their next steps, with 66% agreeing or strongly agreeing they were well prepared. This was followed by those attending college (53%), secondary schools (42%) and selective schools (41%) who were less confident of their preparation.

Year 10 or 11 students had a slightly higher proportion of respondents agreeing and strongly agreeing that they were well prepared, with 50% compared to 44% of those in year 12 or 13. However, there were similar levels of disagreement across years because year 12 or 13 students had a slightly high proportion of respondents reporting they neither agreed nor disagreed. Year 13 students may not be certain about what either employment or university might demand of them.

Q. Do you have any additional comments you would like to make about this?

Respondents were asked whether they had any additional comments they would like to make about their preparedness for their next step. In total, 127 respondents gave additional comments. Most of these responses were more general comments with respondents giving their views of the TAG process in general, particularly comments around the unfairness or inequality of the TAGs. These are not detailed below.

Considering preparedness for next year, comments were predominantly (24%, n = 31) focussed on missed content and teaching, with specific circumstances being detailed. Some comments talked about the lack of much, or any, teaching following Easter, with a focus only on exams and other assessments. This included students suggesting that there was no further teaching once all of the evidence for TAGs had been collected, despite there being several weeks of term remaining. In addition, some disjoint between what was taught and what was assessed was noted by some respondents. Some concerns about the potential impact of the arrangements on specific skills were also highlighted, covering practical science skills, more hands-on activities in creative or practical subjects, field trip skills and work experience.

In contrast, six respondents (5%) talked about the efforts their centre had made to catch up with teaching and learning, including providing catch-up classes to cover missed content before the end of the summer term. These responders were content with their preparation for the future. Five respondents (4%) also recognised the effort of their teachers commenting on the stress they may have endured during the process. Some respondents described their own independent efforts to ensure they were ready, rather than anything their centre did (6%, n = 7).

Respondents also reflected on the fact that the TAG evidence process had been somewhat closer to the kind of assessment they expected in higher education (6%, n

= 8) and so considered this a good thing, while one respondent even noted that all of the TAG assessments had been good practice for taking exams. However, it was more common (6%, n = 8) for respondents to mention concerns around their lack of experience of taking formal exams. This was particularly so for those in year 11 that expected to progress on to A levels. Concerns around both the appeals process and the expected implementation of the autumn resit opportunity (2%, n = 3) were also noted.

Final thoughts

This section was answered by all respondents and sought respondents' overall views about the TAG process and the collection of evidence to support this.

Q. Having been assessed through the TAG process, would you have preferred to have taken exams and assessments as originally planned?

Respondents (N = 537) chose either 'yes', 'no' or 'don't know'. By a small majority, respondents indicated that they would have preferred to have taken exams and assessments as originally planned (46%) than to have been assessed through the TAG process (38%). This question can be compared to a similar one that was asked about preferences at the beginning of the process (see page 80). It seems that by the end of the process, some respondents who preferred not sitting normal exams when they were first cancelled in January, would have preferred to have by the end of the TAG process (38% responded yes to the first question, compared to 46% here). This may reflect the number of assessments students were required to complete which seems to have been one of the most difficult issues in the TAG process for students.

Q. Please enter up to THREE words below that sum up your experience and feelings about the TAG process now that the TAGs have been submitted.

Respondents were asked to provide 3 separate words to indicate how they felt. As before, we provide the frequency of these words and plot them as a word cloud (Figure 37).

The 10 most frequent words were:

- anxious (n = 106, 21%)
- stressful (n = 82, 16%)
- unfair (n = 80, 16%)
- worried (n = 67, 13%)
- relieved (n = 56, 11%)
- nervous (n = 47, 9%)
- stressed (n = 35, 7%)
- happy (n = 34, 7%)
- fair (n = 31, 6%)
- uncertain (n = 24, 5%)

These words seem to reflect that students were feeling fairly negative about their experience of the process, especially compared to how they reported their feelings when exams were cancelled, which was more mixed (see Figure 27). At the stage when the survey was completed students had not yet received their grades and, in some cases, may not have been fully aware of the evidence used. Anticipation of results might account for some of the anxieties.

A number of respondents also commented on perceived inconsistency within their centres, both between different subjects and between individual teachers, although this was less common. Considering practices within centres, several respondents suggested that they had observed cheating or malpractice from other students, or ways of gaming the system. There were a number of respondents who considered that the use of past papers that could be accessed prior to the assessments had led to some students being able to prepare in advance and obtain a better grade than was deserved.

Many respondents considered that the evidence used, and therefore the grades determined, were not representative. They suggested 3 factors causing this: when pieces of evidence were completed, the amount and areas of content that were taught, and the types of evidence used. Some respondents felt that they had been disadvantaged by their teachers not keeping them informed of what evidence was being used or what their latest results or grades were, meaning they could not improve and did not know which areas they needed to work on.

Some respondents also considered the impact that absence, school closures and COVID more generally had had on them, with several suggesting this should have been taken into account to a greater degree. There were some responses suggesting that the TAGs may not be completely fair due to teacher biases and discrimination and some respondents suggested that private candidates and candidates that could not afford tutors may be disadvantaged.

Forty-five per cent of the responses (n = 87) reported difficulties that respondents had experienced with the TAG process. This was most commonly a belief that there were too many assessments. Many respondents felt that the period where centres were gathering evidence for TAGs through tests and assessments, was more intense than a normal exam series. This, they suggested, was due to a higher number of assessments in a shorter time period, making testing feel constant. Many of the respondents suggested that there was not enough time to prepare and revise effectively for these assessments as a result of the intensity.

There were also a number of respondents who expressed issues with remote learning and learning in general throughout the academic year, as well as less support while revising. As a result of many of the issues discussed so far, many of the respondents seemed concerned that their grades could not fully reflect their ability or be fair in comparison to students in centres that had had different experiences. A number of respondents considered the stress, anxiety and mental health issues that they had experienced as a result of the TAG process, with some suggesting more should have been done to support students during this time.

Some respondents considered issues with the decision making around the TAG process. It seemed, similarly to the teachers, that students felt the decisions had taken too long to be made and were not clear enough. A few students also

considered the burden that had been placed on teachers and felt that this had been unreasonable.

Twelve respondents (6%) mentioned concerns for the future, with some anticipating disappointment on results day and others not feeling prepared for their next steps. A few respondents also expressed concern around disruption to their futures, with some suggesting employers would not respect their grades as they would not be considered comparable to those in previous years.

Despite these comments, 16% of respondents (n = 31) expressed positive opinions of the TAG process. The most common benefit that students saw with the TAG process was that grades were not only awarded on the basis of one single exam. The TAG process had meant they were given more opportunity to demonstrate their abilities over time and reflect continued effort, which was especially beneficial for those that struggle with exam style assessments. Some respondents also suggested that TAGs had been less stressful than exams.

There were a number of respondents who felt that TAGs were necessary considering the disruption experienced and that it would not have been fair to conduct exams in the normal way. There were also some respondents who felt TAGs, or aspects of the TAG process, should be used in 2022 and could be the basis of a general change to the structure of assessments in future, to reduce the reliance on exams. This demonstrates the diversity of opinions about the TAG process.

Discussion

We highlight some of the main findings from the 2 surveys and explore some possible reasons for the findings.

Confidence in accuracy and freedom from bias

The teachers who completed our survey reported a high level of confidence in the TAGs they awarded in summer 2021, which was similar to reported confidence in the 2020 CAG survey. The median confidence in the accuracy of the submitted TAGs was 90, on a scale of 0 (no confidence) to 100 (high confidence), with nearly 1 in every 3 respondents (29%) rating their confidence at 100. Teachers were also highly confident that their judgements were free from bias, with a median confidence of 95 out of 100. Forty-three per cent of respondents rated their confidence in being free from bias at 100. Unlike the CAG survey, however, we observed no significant differences in confidence based on the respondent's job role. In fact, confidence in the accuracy of the TAGs, and in them being free from bias, was high across all roles, centre types, and subjects. As we will discuss in further detail, it is important to note that teachers' high confidence in their own centre TAGs does not necessarily translate to confidence in those submitted by other centres.

Those students who participated in the survey were much less confident in their TAGs being free from bias. When asked if their relationship with their teacher or tutor might affect their grade, 43% agreed or strongly agreed that it might (37% disagreed or strongly disagreed). Further, when asked if they felt they had been judged fairly compared with their peers at their centre, less than half agreed or strongly agreed (46%) whilst 26% disagreed or strongly disagreed. Although students were not asked directly to rate their confidence in their TAGs being free from bias, these questions demonstrate that a substantial proportion of students felt that there may have been some element of bias present in the judgement of their TAGs. It is worth remembering that these views were captured before any results were issued, so it is unclear if or how they might have changed post-results.

Perceived fairness

When we asked teaching staff how fair TAGs were, compared to the grades awarded following normal assessment in past years, the most common response was that TAGs were 'about the same fairness' (38%). However, where a difference in fairness was reported more respondents rated TAGs as slightly less fair (34% 'less' or 'much less' fair compared to 22% 'more' or 'much more' fair).

We observed that middle leaders (in other words heads and deputy or assistant heads of departments and key stage leaders) tended to rate the fairness of TAGs more highly than others, though the average for this group still fell in the 'less fair' category. At least in our survey sample, it was these middle leaders, particularly heads of departments, who were most heavily involved in the TAG process and so it is perhaps unsurprising that they felt relatively more positive about the fairness of TAGs. There were some centre-type differences observed around fairness, with respondents from colleges being more likely to rate TAGs as being fairer (30% versus 22% across all centre types) and respondents from independent schools being more likely to rate TAGs as less fair (39% versus 34%).

When asked if they felt TAGs would be fair or unfair on them, students were more positive, although note that teachers were asked about comparative fairness of TAGs and exams, and students about absolute fairness of TAGs. Over half (53%) felt that, overall, their TAGs would be fair or very fair. Sixteen per cent felt they would be 'neither fair or unfair' whilst 24% felt they would be unfair or very unfair. When broken down by centre type, we found that students from independent schools were the most like to believe their TAGs would be fair or very fair (61%) whilst those from selective schools were least likely (43%). This is somewhat at odds to the response from teachers – where respondents from independent schools were the least likely to rate TAGs as being fair. It is unclear why this discrepancy exists.

When students were asked if they thought they would have achieved better or worse grades had they been able to take exams and assessments normally, over half (51%) felt they would have achieved better or much better. This does seem to present a conflicted view from students: most felt the TAGs were fair but most also felt as though they would have performed better had they sat exams. Perhaps, therefore, students were indicating that they felt as though TAGs were as fair as they could be, given the circumstances. We also note that students had not yet seen their TAGs when they responded to this survey and so they could not have been certain of what their final grades were.

Perceptions of consistency across centres

One of the issues surrounding fairness that came up from both teachers and students alike was a potential lack of comparability and standardisation between centres. When asked if they agreed that they will have been judged fairly compared to students in other centres, most students (54%) either disagreed or strongly disagreed. Only 21% agreed or strongly agreed. Combining this with the relatively high agreement in the fairness of their own TAGs, indicates that significant numbers of students felt as though it was the other centres who were not judging TAGs fairly – presumably to the advantage of the students in those centres.

Teachers too, in some of the open response questions, noted that they had concerns about how other centres were approaching TAGs, such as the evidence they used, and some blamed a perceived lack of consistency on the guidance centres were provided. While the vast majority of the teachers responding to the survey reported thorough processes in their centres and, as noted, had very high confidence in the accuracy of the TAGs they produced, they were a self-selecting sample actively choosing to complete this survey. They may not be entirely representative of the national population (we discuss this issue further below).

Guidance

In the teacher survey, respondents were asked if they were aware of the Ofqual and AO guidance on making objective judgements and whether these were useful. For both sources, respondents indicated a high level of awareness (96% for Ofqual and 78% for AO guidance) which were increases relative to the same question in the CAG survey (90% and 66% respectively). However, only around half of respondents considered this guidance to be useful (50% for Ofqual and 47% for AO guidance) which is substantially lower than in 2020 (84% and 85% respectively).

There are several possibilities for this disparity. Fundamentally, the TAG process involved making an evaluation of the level students were working at on content they had been taught by considering completed and marked and/or graded work. This contrasts with CAGs, which involved a prediction of how a student would have been expected to perform if they had sat normal assessments.

While determining TAGs would often have involved an element of judgement (see [the interview report](#) for more detail on how grade decisions were made), this is perhaps a more evidence-based judgemental process than CAGs, where progression trajectories had to be estimated in addition to the level the student was working at before the cancellation of assessments. Less subjective judgement on the part of teachers was required and perhaps, therefore, respondents felt the issue of bias was less important.

Additionally, the lower number of respondents indicating that the guidance on making objective judgements was useful may have been because respondents had familiarised themselves with the guidance the previous year or because they felt that sufficient steps had been taken by their centre to mitigate bias. Indeed, 93% of respondents felt their centre had put in place at least some partially effective steps, up from 82% in the CAG survey, and reports of formal bias training taking place was twice as common this year than last.

Finally, we should note from answers to some of the open response questions that many respondents considered the assessment related guidance provided by

government (Ofqual and DfE) and the awarding organisations as being both relatively late and not as helpful or specific as they had hoped or expected.

Evidence used

The Ofqual guidance on TAGs made it clear that teachers could use a range of evidence to assess students and that students should only be assessed on content which they had been taught. Many of the specific evidence-collection decisions, such as what types of evidence to use, how much to collect, and when to collect it, were left to individual centres.

We asked both teachers and students what sort of evidence was used to determine TAGs. Tests taken under exam-like conditions (for example, mock exams and class tests) were reported as the main source of evidence for most GQs by both teachers and students alike. Nearly three quarters of students (73%) indicated that they had been told what would be on these tests at least on a few occasions. When conducting the follow-up interviews it was noted that this normally meant what topics would be tested, rather than the specific questions. Respondents from VTQ courses and the more creative GQ subjects indicated that class work and assignments were more commonly used.

We note that whilst carrying out the follow-up interviews, it became clear that we had not chosen the best labels for the categories for exam-type evidence. The materials provided by exam boards had not been used in quite the way we anticipated, and the terminology used in different centres had varied. The distinction between mocks or practice exams, class tests, and AO-sourced tests ended up being quite blurred – in fact many assessments were effectively all 3, being close to mocks in content, but run in the classroom (though usually under exam conditions), and partially making use of the AO sourced materials, but not always being entirely based on them.

Compounding the potential problem with our categories was the varied terminology centres themselves used. Partly, this stemmed from the announcement that exams had been cancelled. Some centres believed that it would be inappropriate to refer to any assessments as an examination since that would represent bringing back exams 'by the back door'. Therefore, a variety of terms were used in addition to exams, such as 'mini-tests'. In contrast, some centres called their assessments 'examinations' or 'mocks' as they believed that this was terminology familiar to students and made clear what they were about to sit.

Respondents may therefore have been unclear which options to choose, and this is likely to have spread the reported weightings across different exam-like options. The weighting reported for mock or practice exams in GQs was substantially higher last year (82 out of 100) than it was this year (67 out of 100). It may be that if we had offered a single option of 'written tests taken under exam conditions' it would have

received a weighting closer to that of mocks last year. Follow-up interviews showed that tests taken under exam conditions were by far the most important sources of evidence for most GQ subjects.

Teachers indicated that, typically, they used 4 to 6 pieces of evidence to judge TAGs for GQs, although there were some large outliers. There was little variation by centre type or subject (with art-related subjects being the main exception). Over three quarters of GQ respondents indicated that most of the evidence was collected after the January announcement that exams and assessments were cancelled. Students tended to agree with this sentiment, with 77% indicating that 'no more than half' of the evidence was collected before the announcement. We note, however, that teachers of VTQ courses tended to collect more evidence from before the announcement – likely reflecting the fact that continuous assessment is more widely used in those qualifications.

When asked if they were able to use the same evidence for all students, the majority of GQ teachers indicated that they were (54%). However, there was some variation by centre-type, with over half of respondents from academies (51%) indicating that they had to make exceptions for some students whilst only just over a third of respondents from secondary selective and independent schools had to do the same (37% and 38% respectively).

Content assessed for TAGs must have been taught to students, and most teaching respondents had not delivered 100% of the course content. Those from secondary selective and independent schools were able to teach more of their respective courses than average (90% versus 85%). Interestingly, students reported that less of their courses had been taught (a median of 70-80%) than teachers did. It is not clear whether this is due to students counting content they had missed due to their own absence or lack of engagement, or differing views on what had been properly taught. This may also have been a difficult question for students to answer accurately as they might not have been fully aware of what remained untaught, or what proportion of the course that represented.

The use of previous years' outcome data in quality-assuring TAGs was widely reported by teachers from across all centre types and subjects. Upon further analysis of some of the open-response questions, we found that these data were often used by senior leadership to align grade distributions with previous years. Some respondents reported friction between teachers and senior leaders, with the teachers feeling it was unfair to alter individual grades but leaders feeling the need to ensure consistency between years. This was also seen in 2020. We note that the number or proportion of grades that were actually changed during internal quality assurance is not known.

Impact on teachers and students

Common throughout the free-text responses on the teacher survey, and in the follow-on interviews, were the issues of pressure and workload. Indeed, when we asked teachers to provide us with 3 words summarising their experience of judging TAGs, the most commonly provided were “stressful”, “time-consuming”, and “exhausting”. Teachers also reported that they had spent around twice as many working days completing the TAG process compared to the CAG process in 2020 – likely due to the requirements around actually assessing students, rather than making a prediction.

In several of the open-ended questions, the respondents noted the extraordinary amount of effort required to undertake the TAG process. Many stated how they felt that they had to “do the exam boards’ job for them” whilst still attempting to teach their students during an unprecedentedly stressful time. For example, some respondents noted, to their disappointment, that AOs did not always provide new, unseen, assessment materials to examine their students and how, therefore, the responsibility to create, mark, and moderate these assessments fell to them.

Students also noted feeling pressure, though this more commonly took the form of anxiety and stress, perhaps relating to the fact that they were not yet aware of their final grades. Common across both teachers and students was wanting to return to normal assessment as soon as possible. Some teachers, however, did indicate that at least some of the elements of TAGs, such as empowering teachers to make judgments, could be useful and some students appreciated that they were given the opportunity to demonstrate their abilities over time rather than being judged based on just one exam.

Preparedness for next steps

When asked how well-prepared students were for the next step (whether this was education, employment or training), teachers were a little more confident than students. Only 20% of teachers felt that students would be poorly prepared, but 33% of students reported this. This may reflect a potential disjoint between what teachers felt they had taught, and what students felt they had learnt. It may also be a measure of uncertainty of the unknown. Students might be expected to be a little less clear than teachers as to what their next steps may demand of them, and so rate their preparedness lower. Hopefully, teachers have been more correct than students in this instance.

Limitations, demographics and survey bias

Finally, we note that all surveys suffer an element of response bias, whether carried out online, by telephone or on paper. Respondents who choose (or decline) to answer a survey may have specific reasons for doing so. As the regulator responsible for the overall approach for TAGs, these biases may possibly be even stronger. Respondents may have had particular grievances they wanted to air, or may have been keen to show that they completed a thorough and careful process to determine TAGs. Therefore, the views and data analysed in this survey may not entirely represent the national population of both teachers and students. The survey demographics do suggest some centre types were over-represented in both surveys.

We also note that our teacher survey respondent demographics show a clear over-weighting of more senior roles, particularly heads of department, as well as an over-representation of certain centre types (this was also true of the student responders). While this may not necessarily skew the results, this cannot be ruled out. One certainty is that the voice of newly-qualified teachers (NQTs) and early career teachers (ECTs) is under-represented here as our survey respondents were predominantly very experienced. These less experienced staff are the ones that we might imagine could struggle most with aspects of making the TAG judgements, since they would have less experience of leading students through normal assessments.

Despite these limitations, this report provides valuable insight into the views and experiences of teaching staff and students in 2021, a year in which qualification grades were determined via a unique process of teacher assessment. It also provides a useful comparison with 2020, during which a different process was used.

Annex A – Information screen on teaching staff survey

When first entering the survey, potential respondents saw the following information on-screen to help decide whether to take part

Ofqual's teaching staff survey on Teacher Assessed Grades, summer 2021

What is this survey about?

In England, many staff in schools, colleges and training providers have been involved in making Teacher Assessed Grade (TAG) judgements as a result of the cancellation of examinations and assessments this year.

Now that the TAGs have been submitted to exam boards and awarding organisations, as regulator of qualifications in England, Ofqual wants to understand the perspectives of those who have been involved in this important process. This information will be invaluable for us to understand, from a research perspective, how the process worked, how expert judgement was exercised, and the experience of those involved, including any challenges faced. It will also inform future practice in the event that anything similar is required again.

We will publish our findings later in the year. This survey will form one part of the overall research project. The information you give will not influence any standardisation or quality assurance measures taken by awarding organisations, and we will not identify individuals or centres in any way - the survey does not require you to give this information.

Who is this survey for?

This survey is designed to capture the perspectives of staff in **England** at all levels of seniority, who were involved in the process of determining TAGs, for both General Qualifications (GCSE, AS, A level) and Vocational and Technical Qualifications (e.g. BTECs, Applied Generals). This includes:

- Staff who judged TAGs for individual students they taught
- Staff who were involved in planning and designing the process, including internal QA
- Staff who carried out any kind of internal QA or checking role
- Staff who signed centre declaration forms to support the TAGs submitted to awarding organisations

How long will this take?

Depending on which aspects of generating and quality assuring TAGs you were involved in, the survey will likely take 15-20 minutes. Certain sections will be skipped depending on your answer to earlier questions – this is intentional.

If you were involved in many aspects and want to give fuller answers to the free text questions, it may take up to 30 minutes.

The survey is mostly made up of closed response/multiple selection questions, although there are a few places to tell us more in your own words. Please do not provide any personal information or identify centres or other individuals in your responses.

How will my responses be treated?

Before you start, we would like to assure you that your answers will be treated in strict confidence. We do not require or collect personal information as part of the survey responses. At the end, you are able to volunteer your contact details to indicate a willingness to participate in our follow up research, in which case, we ask for your contact details to correspond with you should you be selected to participate in the follow up research. Any personal data collected is processed in accordance with the Data Protection Act and the General Data Protection Regulation (UK GDPR). The information that you provide will be used only for the purposes of this research study, to inform our findings and help us regulate more effectively. Ofqual is permitted to carry out research under Section 169 of the Apprenticeship, Skills, Children and Learning Act, 2009. As such, we rely on public task as our lawful basis under data protection law for processing any personal data. We will not identify any individual or school within any published report.

We use Citizen Space to run this survey who act as our data processor. Their [privacy details can be found on the Citizen Space website](#). Once we receive the survey responses, these will be securely held by Ofqual and not shared externally. We will only keep these responses for as long as is necessary and will retain them for no longer than 7 years.

Participation in this survey is entirely voluntary and you do not need to take part. You can also withdraw your participation from this survey at any point by closing your browser window. Responses from partially completed surveys will not be saved. However, once you have submitted your responses it will not be possible to withdraw them as they are not individually identifiable.

If you have any questions about this survey please contact us at
TAGsurvey@ofqual.gov.uk

The deadline for responses is midnight on Saturday 7 August. Thank you for your time.

To continue with the survey please confirm the statements below

My centre has completed the submission of Teacher Assessed Grades to Awarding Organisations or Exam Boards for the qualifications I will tell you about AND I have read the information on this page and wish to continue with the survey

Annex B – Information screen on student survey

When first entering the survey, potential respondents saw the following information on-screen to help decide whether to take part

Ofqual's student survey on Teacher Assessed Grades, summer 2021

What is this survey about?

In England, many teaching staff in schools, colleges and training providers have been making Teacher Assessed Grade (TAG) judgements as a basis for the results students will receive this year. This is because of the cancellation of examinations and assessments due to the Coronavirus pandemic.

Now that the TAGs have been submitted to exam boards and awarding organisations, as regulator of qualifications in England, Ofqual wants to understand the views of those who have been involved in this important process. This will help us better understand the experiences of the TAG process for students like yourself, and how fair you think the outcomes will be.

We will publish our findings later in the year. This survey will form one part of the overall research project. The information you give will be anonymous and we will not identify you or your school or college in any way - the survey responses do not require you to give this information.

Who is this survey for?

This survey is designed to capture the perspectives of students in **England** who will be awarded at least one qualification based on TAGs judged by teachers/tutors. We are interested in both General Qualifications (GCSE, AS, A level) and Vocational and Technical Qualifications (for example, BTECs, Applied Generals).

How long will this take?

The survey will likely take 10-15 minutes.

The survey is mostly made up of closed response/multiple selection questions, although there are a few places where you can tell us more in your own words. Please do not provide any personal information or identify centres or other individuals in your responses.

How will my responses be treated?

Before you start, we would like to assure you that your answers will be treated in strict confidence. We do not require or collect personal information as part of the survey responses. At the end, you are able to volunteer your contact details to

indicate a willingness to participate in our follow up research, in which case, we ask for your contact details to correspond with you should you be selected to participate in the follow up research. Any personal data collected is processed in accordance with the Data Protection Act and the General Data Protection Regulation (UK GDPR). The information that you provide will be used only for the purposes of this research study, to inform our findings and help us regulate more effectively.

Ofqual is permitted to carry out research under Section 169 of the Apprenticeship, Skills, Children and Learning Act, 2009. As such, we rely on public task as our lawful basis under data protection law for processing any personal data. We will not identify any individual or school within any published report.

We use Citizen Space to run this survey who act as our data processor. Their [privacy details can be found on the Citizen Space website](#). Once we receive the survey responses, these will be securely held by Ofqual and not shared externally. We will only keep these responses for as long as is necessary and will retain them for no longer than 7 years.

Participation in this survey is entirely voluntary and you do not need to take part. You can also withdraw your participation from this survey at any point by closing your browser window. Responses from partially completed surveys will not be saved. However, once you have submitted your responses it will not be possible to withdraw them as they are not individually identifiable.

If you have any questions about this survey please contact us at TAGstudentsurvey@ofqual.gov.uk

The deadline for responses is midnight on Saturday 7 August. Thank you for your time.

To continue with the survey please confirm the statements below

I will be receiving at least one qualification grade based on TAGs judged by my teachers or tutors this summer AND I have read the information on this page and wish to continue with the survey.



© 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:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344

public.enquiries@ofqual.gov.uk

www.gov.uk/ofqual