



A structured discussion of the fairness of GCSE and A level grades in England in summer 2020 and 2021

Victoria Crisp, Gill Elliott, Emma Walland & Lucy Chambers

To cite this article: Victoria Crisp, Gill Elliott, Emma Walland & Lucy Chambers (2025) A structured discussion of the fairness of GCSE and A level grades in England in summer 2020 and 2021, *Research Papers in Education*, 40:1, 44-71, DOI: [10.1080/02671522.2024.2318046](https://doi.org/10.1080/02671522.2024.2318046)

To link to this article: <https://doi.org/10.1080/02671522.2024.2318046>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 18 Feb 2024.



Submit your article to this journal



Article views: 1484



View related articles



View Crossmark data

A structured discussion of the fairness of GCSE and A level grades in England in summer 2020 and 2021

Victoria Crisp, Gill Elliott, Emma Walland and Lucy Chambers

Research Division, Cambridge University Press & Assessment, Cambridge, UK

ABSTRACT

In England, examinations for general qualifications (GCSE, AS and A level) were cancelled in summer 2020 and summer 2021 in response to the COVID-19 pandemic and replaced with contingency measures involving teacher judgements. For summer 2020, the intention was to calculate grades using rankings provided by centres, prior attainment data for the current learners and performance data for centres' candidates in previous years. This approach was later abandoned, and 'centre assessment grades' were issued instead (except where calculated grades were higher). For summer 2021, 'teacher-assessed grades' were awarded to learners. These drastically different to usual approaches to the assessment of high-stakes qualifications raised debate about fairness. This article discusses evidence that can inform judgements about the fairness of these grades, using an existing framework of different 'senses' of fairness to provide a structure. In this way, we judged how fair grades were in each sense compared to prior years for each of the approaches used (or nearly used) in summer 2020 and 2021. The analysis highlights tensions in the prioritisation of different aspects of fairness in the unusual circumstances of the pandemic.

ARTICLE HISTORY

Received 30 January 2023

Accepted 24 January 2024

KEYWORDS

Fairness; assessment; GCSE; A level; grades; Covid-19 pandemic

Introduction

In response to the COVID-19 pandemic, decisions were taken in England to cancel examinations for GCSEs, AS and A levels¹ in summer 2020 and summer 2021 and grades were instead derived from teacher judgements (DfE 2020a, 2021a). Reflecting on the fairness of these grades is important given that the assessment methods were different to usual, but the uses of results were similar. This article explores the fairness of GCSE, AS and A level grades in summer 2020 and summer 2021, in England, for the students receiving them, using different 'senses' of fairness (Nisbet 2020; Nisbet and Shaw 2019) to structure the discussion. The fairness of results from vocational and technical qualifications in summer 2020 and summer 2021 was also affected, arguably in more complex ways since the varied nature and purposes of these qualifications resulted in a need for varying approaches (DfE 2021a; Mattey 2021; Ofqual 2020b, 2021d). However, to contain the scope, this is not explored.

CONTACT Victoria Crisp  vicki.crisp@cambridge.org  Research Division, Cambridge University Press & Assessment, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA, UK

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Summer 2020 context

For students completing qualifications in summer 2020, learning was unaffected until 20 March 2020 when schools and colleges in England closed to most learners (DfE 2020b). Secondary school students were taught online and studied independently for the rest of the school year (Elliott 2021). Summer 2020 examinations were cancelled for many qualifications and an approach was developed to provide students with calculated GCSE, AS and A level grades based on assessment by schools and colleges (henceforth, ‘centres’) (DfE 2020a). For each subject, centres submitted centre assessment grades (CAGs) and a rank order of students within each grade (Ofqual 2020d). A statistical moderation method was developed by Ofqual (Office of Qualifications and Examinations Regulation) at the request of the Department for Education (DfE 2020a) to standardise grades across centres and, thus, maintain qualification standards as far as possible (DfE 2020a). The approach chosen, after some trialling, used data on the historical distribution of grades at each centre from recent years and value-added relationships between prior attainment and grades in the subject, to adjust (if needed) the distribution of grades for a subject within each centre whilst maintaining rank orders (Lee, Stringer, and Zanini 2020). Where numbers of learners at a centre studying a particular subject were small, greater weight was placed on CAGs (Ofqual 2020g) as there was more risk of error when applying the adjustment (Ofqual 2020a). Given that several awarding bodies in England offer GCSEs and A levels, a national adjustment was applied for each subject, to align results with expected national outcomes (Ofqual 2020h).

Calculated grades, derived from centres’ rank orders, were issued to AS and A level students on 13 August 2020 (JCQ 2020c). However, low public confidence in the statistical approach resulted in the government instructing awarding bodies to re-issue grades, giving students the higher of their CAG and calculated grade (Lee, Stringer, and Zanini 2020). GCSE grades were issued the following week based on the revised approach.

An autumn exam series was provided for any learners wishing to take exams (Ofqual 2020a).

Summer 2021 context

For students completing qualifications in summer 2021, learning experiences were affected for a longer period of their two-year courses. Between March and July 2020 and between January and March 2021 most learners were taught remotely. In late 2020, it was hoped that summer 2021 exams could go ahead on an adjusted basis (e.g. reduced content, certain requirements removed) (Ofqual 2020c). However, with the escalation of a new COVID-19 variant, the government decided in early January 2021 that most exams would not go ahead and that GCSE, AS and A level grades would be based on teacher judgements with no calculated adjustment (DfE 2021a). For these teacher-assessed grades (TAGs), students were assessed using a range of evidence from throughout their course (e.g. centre-devised tasks, work for non-examined assessments, responses to example exam questions) but only on content they had been taught (Ofqual 2021e).

Again, an autumn exam series was provided (DfE 2021a).

Defining fairness

Fairness, and the related concept of justice, have been subject to much theorisation in various social science disciplines such as philosophy, psychology and law (e.g. Lind 2019; Moore 1997; Rawls 1958). In the context of educational assessment, the term ‘fairness’ is commonly used but the implicit meaning behind its use varies (Nisbet and Shaw 2019). Over time, literature on educational measurement has gradually given the concept of fairness more attention, with fairness viewed as the absence of construct-irrelevant variance² in assessment results and the avoidance of bias against specific groups of learners or individuals (AERA/APA/NCME 2014; Karami and Mok 2013).

Drawing on conceptualisations from Rawls 1958, Leventhal (1980) and others, one common structure for evaluating fairness considers procedural fairness (relating to rules being transparent and impartial such that different individuals or groups have equal opportunity for a satisfactory outcome) and distributive fairness (relating to resources being allocated in a socially just manner). This has been applied in contexts as broad as predicting game-playing behaviour (Krawczyk 2011), explaining value dimensions affecting views on political scenarios (Rasinski 1987) and understanding community reactions to energy industry developments (Witt, Whitton, and Rifkin 2018). Whilst this structure captures key fairness concerns, users have sometimes needed to apply a more detailed sub-structure to their evaluations (e.g. Witt, Whitton, and Rifkin 2018). In considering how to evaluate the fairness of pandemic-era grades, we felt that a more detailed framework, designed for educational assessment, would be valuable. Existing frameworks in the context of assessment often focus on whether various specific factors are addressed by test procedures, such as ensuring that accommodations are provided for relevant learners (e.g. Jonson, Trantham, and Usher-Tate 2019). Some theorists have also incorporated social impacts of assessment (Kunنان 2004). Nisbet and Shaw (2019) take a wider view than most other assessment theorists and identify four ‘senses’ of fairness:

- **Formal** – whether the formal rules are applied accurately and appropriately (regardless of the quality of the rules themselves). This relates to the guidance and regulation around how assessment should be conducted.
- **Implied contractual** – whether the ‘legitimate expectations of those affected’ (p.613) are met. This can be understood as about what the student can reasonably expect about how they will be assessed.
- **Relational** – whether ‘(relevantly) like cases’ (p.613) are treated alike. This concept sees discrimination between learners as fair when based on relevant considerations (e.g. different performance on a test) and not on irrelevant considerations (e.g. ethnicity, school). This sense of fairness includes comparability between results awarded to learners and avoiding construct-irrelevant variance in results.
- **Retributive** – whether an outcome is ‘an appropriate reward (or penalty) for what has gone before’ (p.614). This relates to whether the outcome (in this case, the exam grade) is deserved and justified, and whether learners meet requirements (for a particular grade).

This framework was selected for the current analysis because the senses of fairness echo key themes from the literature in other fields and are highly applicable to

assessment. Nisbet and Shaw (2020) identify two further senses of fairness: retrospective, the fairness of past actions that affect results; and consequential, the fairness of actions based on results. Both are excluded from the current analysis because any unfair actions prior to the start of the relevant learners' two-year courses occurred in usual (pre-pandemic) circumstances and because an evaluation of consequential fairness would be more complete if conducted later when further evidence on actions and effects has emerged.

Stakeholder perceptions are important, and often explored, in the fairness literature (Hilliard, Guenole, and Leutner 2022; Krawczyk 2011). This notion of perceived or 'felt' fairness, that a situation can be perceived as more or less fair by those affected (Nisbet and Shaw 2020), may be particularly relevant to the implied contractual and retributive senses. A perception of unfairness does not necessarily mean that something is of low fairness, as the perception could be based on false assumptions (Nisbet and Shaw 2020) or misunderstandings.

In this article, we treat fairness as a relative concept that exists along a continuum (Nisbet and Shaw 2020; Rasinski 1987). Arguably, no assessment can be absolutely fair, but there may be a cut-off point on the continuum at which an assessment can be judged sufficiently fair, with the location of that point potentially needing adjustment if surrounding circumstances change (Nisbet and Shaw 2020).

This links to a view that assessment should be seen as an event taking place within a particular context, and that surrounding circumstances should be considered. In other words, a 'situated' (rather than 'isolated') view should be taken when evaluating fairness (Nisbet and Shaw 2020). Relatedly, Mislevy et al. (2013) propose a 'conditional' sense of fairness, arguing that it may be appropriate to vary how learners are assessed based on known learner needs (in contrast to general tendencies to assume that fairness is best achieved by testing all learners in the same way). These arguments provide space to argue that it is reasonable for assessments to be conducted differently across learners and to argue that unusual circumstances should influence what is considered sufficiently fair.

In the context of the pandemic and the public health driven decision that examinations could not go ahead, it is unlikely that any assessment approach could provide grades that were as fair as usual years in all senses. Adjusting what was considered sufficient was necessary and trade-offs between different senses of fairness were unavoidable (though the latter was perhaps not appreciated until later).

Aims and scope

This article explores the fairness of summer 2020 and summer 2021 GCSE, AS and A level grades in England. Fairness is evaluated in each of the four main senses proposed by Nisbet and Shaw (2019, 2020). To contain the scope of discussion, the focus is the fairness of *grades* for the learners receiving them and excludes wider fairness issues. Evaluations are made relative to prior years when assessments were not affected by the pandemic.

The discussion focuses on the three assessment approaches relevant to summer 2020 and 2021:

- calculated grades based on centre assessment, as planned for use in summer 2020.

- the better of centre assessment grades (CAGs) and calculated grades, as finally used in summer 2020.
- teacher-assessed grades (TAGs) as used for summer 2021.

Whilst the way that grades were established in these approaches were different, it is assumed that in each case grades were intended (as usual, see Shaw and Crisp 2012) to provide a measure of achievement, which should provide some indication of the learners' future potential.

Due to limited available evidence and because learner and public views are relevant to interpreting certain aspects of fairness, this article makes considerable use of grey literature and newspaper and other media articles. In particular, Ofqual and DfE (Department for Education) reports are important as they set out the assessment processes and related rationales. Some research from respected organisations was published in reports rather than journals, presumably for speed given the circumstances. Relevant experts provided media commentary and online analysis, for example, Sharon Witherspoon from the Royal Statistical Association wrote an article in Schools Week (Witherspoon 2021), and Jake Anders and colleagues from University College London wrote several blogs (e.g. Anders, Macmillan, and Wyness 2021). For newspaper articles, some caution is needed due to greater risk of inaccuracies and accounts being anecdotal, but for the current context we argue that they are sometimes a useful source of learners' and others' views.

When evaluating whether, and to what extent, the fairness of grades was different to prior years, the following factors were considered:

- Whether there were differences in approach that would conceptually and logically affect fairness in a particular sense.
- Whether there was evidence (e.g. from statistical results data, from interviews or surveys or other sources of stakeholder views) that fairness was likely to be affected in a particular sense for more than a small proportion of learners. Feelings of unfairness were included here for the fairness senses where this is pertinent (i.e. implied contractual and retributive).
- The likely scale of any changes to fairness conceptually, logically or based on the evidence.
- If the above did not indicate changes to fairness, then fairness was assumed to be the same as prior years (though further evidence could emerge and require this to be revisited).

Analysis of fairness

Formal

For summer 2020 and 2021, extraordinary rules were put in place for arriving at GCSE, AS and A level grades.

Formal: Summer 2020 intended and final approaches

The intended and final 2020 approaches are considered together in this section since both relate to whether centres followed the relevant assessment guidance (e.g. JCQ 2020f; Ofqual 2020d), which in effect set out the rules for what was required. For each subject, centres were required to use existing records and evidence (e.g. complete or incomplete non-examined assessment, performance on classwork, homework and/or mock exams), balancing these evidence sources appropriately, to make ‘a holistic professional judgement’ of the student’s performance (Ofqual 2020d, 5). Grades should be realistic and objective and reflect what the student was most likely to have attained if learning and assessment had proceeded as usual (Ofqual 2020d). Teachers were not required to incorporate student work conducted after schools closed (to most) in March 2020 and were asked to consider any work after this time with caution if it showed a change in performance (JCQ 2020f). In terms of establishing a rank order for students at each grade, guidance stated that this should be for the whole subject cohort at a school and that teachers within a department should discuss and come to a shared view of the standard (Ofqual 2020d). Viewing examples of student work to aid decisions was recommended. The guidance noted that several methods for determining ranks and standardising across different teachers were possible and described one example, but did not mandate a particular method (Ofqual 2020h). Additional guidance materials detailed use of the centre’s results in prior years and the current students’ prior attainment to inform judgements (ASCL 2020a, 2020b).

In survey and interview research, many teachers reported using mock or practice exams as the main basis for centre assessment grades and ranks (Johnson and Coleman 2021; Howard, Holmes, and Case 2021). Using marks from past classwork and class tests was also common and work for non-examination assessments was used in relevant subjects. Smaller numbers of teachers attempted additional remote assessments but found these problematic in terms of work not being submitted and/or suspicions of inappropriate parental support or use of notes (Johnson and Coleman 2021). Some teachers reported using an element of ‘teacher intuition’, which was sometimes linked to an apparent lack of evidence (Johnson and Coleman 2021). Whilst the research suggests variation in evidence used by different schools, the reported approaches fall within the (arguably flexible) guidance. However, less appropriate approaches may have been used but not reported. ‘Teacher intuition’ feels least in keeping with the focus on evidence of learner performance, but since teachers were effectively asked to predict how learners would have performed in exams, this may not be unreasonable and is likely to have been informed by years of observing learners’ performances in schoolwork and their final exam results.

Teachers often reported that their centre followed the guidance in terms of considering the profile of student results in previous years, but some expressed concern that other centres might not have (Johnson and Coleman 2021). At some centres, the profile of historical results in a subject was used as the starting point when determining CAGs, whilst others used these later to check and potentially adjust teacher judgements (Ofqual 2020a). Both strategies fall within the guidance. In an Ofqual survey, some teacher respondents (around 100³ of 762) reported pressure from senior management to match appropriate grade distributions via a perceived threat of grades being moderated

downwards (Holmes et al. 2021). Smaller numbers (around 20 of 762 respondents) reported pressure from senior management to be optimistic, sometimes in relation to specific students. Small numbers (around 20) also reported pressure from parents and/or students. It is not known whether such pressures influenced judgements.

Based on the available evidence, it seems likely that the vast majority of centres followed the guidance for determining grades and ranks. However, as Ofqual acknowledged at the time, it was likely that some learner evidence was incomplete (Ofqual 2020h) and research with teachers indicates that the evidence types used varied (Johnson and Coleman 2021; Howard, Holmes, and Case 2021). Nonetheless, as there is no evidence that rules were broken, we conclude that 2020 calculated grades and final grades can be considered as fair as previous grades in the formal sense.

Formal: Summer 2021 approach

For TAGs in 2021, different guidance was provided for determining grades. Centres were required to prepare a Centre Policy detailing roles and responsibilities, internal training for teachers, learner evidence to be used and internal quality assurance processes (JCQ 2021e; Ofqual 2021c). Centre Policies were reviewed by exam boards and centres were contacted for virtual visits if the exam board had questions or concerns (JCQ 2021e). Centres were asked to assess the standard at which learners were performing based on a broad range of evidence from throughout their course (e.g. responses to centre-devised tasks, work for non-examined assessments, responses to past paper exam questions) but only for content they had been taught (Ofqual 2021e). Exam boards provided various materials to support decision making, including training materials, example assessment questions and grade descriptors (Ofqual 2021c). Where centres chose to conduct tests using exam board assessment materials, examination conditions were not compulsory (Ofqual 2021c). Centres were required to conduct internal quality assurance, with student grades being agreed by at least two teachers including the head of department or subject lead (Ofqual 2021j), and results being considered in relation to the profile of student results in previous years (Ofqual 2021e). Research suggests that teachers used the expected types of student evidence, such as mock exams, unit/topic tests (sometimes using exam board support materials) and non-examined assessments (Johnson and Coleman 2021; Vitello and Leech 2022). Internal standardisation processes aligned with guidance, with teachers reporting review of learner evidence and grades by senior colleagues (Johnson and Coleman 2021). However, some teachers felt the guidance lacked clarity regarding how much evidence was sufficient and that grade descriptors were difficult to use (Johnson and Coleman 2021). Whilst Johnson and Coleman's research used self-reports from teachers, Vitello and Leech (2022) analysed student evidence samples submitted to an exam board, adding robustness to the evidence.

The exam board quality assurance process reviewed samples of learner evidence from centres. Further scrutiny was considered necessary for 15% of the sampled centres (Ofqual 2021k) but only 195 TAGs were changed as a result (Ofqual 2021h). This suggests that the vast majority of centres' approaches aligned with guidance.

There were reports of parents (and sometimes learners) pressurising teachers regarding TAGs. According to BBC news (Coughlan 2021), the Association of School and College Leaders had heard of teachers being pressured. Further evidence comes from

Holt-White and Cullinane (2021) who report a survey of 3221 teachers in which 15% responded that they had been approached or pressured about learner grades by at least one parent. In a Times Educational Supplement survey of 2800 teachers, 2% admitted being influenced by such pressure (Lough 2021b), suggesting that a small proportion of teachers may not have acted within the rules. Indeed, this could be an underestimate as teachers might have been unwilling to admit being influenced. Whilst it is difficult to confirm the quality of these surveys as little information on participants and recruitment methods is provided, given the large respondent numbers there seems to be enough evidence to suggest that some teachers were approached about grades and that a small proportion may have been influenced by this.

Compared to 2020, guidance for teachers about determining grades was more detailed for summer 2021 with more support materials available. Research on the learner evidence used suggests that guidance was followed. That said, there was considerable flexibility in the exact evidence and approach, and perhaps some lack of clarity over sufficiency of evidence and grading standards. Given that the majority of centres appear to have worked within the guidance, but a small proportion of teachers may have been influenced by pressure when determining certain learners' grades, we surmise that summer 2021 grades can be considered of similar (but not the same) fairness to usual in the formal sense.

Implied contractual

In legal terms, an implied contract is an obligation assumed to exist between parties, based on actions, behaviour or circumstances.⁴ For assessments, this implied contract relates to the relationship between student and assessment provider and what the student can reasonably expect about how they will be assessed. If the learner's legitimate expectations are not met, the learner may experience a 'nasty shock' that violates the implied contract, and which may 'feel' unfair (Nisbet and Shaw 2019).

Implied contractual: Summer 2020 intended approach

When learners began their two-year GCSE and A level courses, they expected to be assessed via traditional examinations marked by external examiners, perhaps (depending on the subject) accompanied by non-examined assessment marked by teachers and externally moderated. Learners would also have expected to receive a result entirely based on how they performed. The intention to use calculated grades instead was made public in late March 2020 (DfE 2020a), several months before results were issued in August. Arguably, this information should have adjusted student expectations such that the different approach and some differences between calculated grades and CAGs were expected. Indeed, students reportedly understood the need for statistical moderation (Taylor 2021). However, learners may have assumed that only small differences were likely given public messaging: 'we expect ... almost all grades students receive will be the same as the CAGs or within one grade' (Ofqual 2020f). That said, even one grade difference can affect learners' next steps.

Receiving a grade that differed from expectations based on teacher feedback would have been a shock, even though the assessment approach should not have been a surprise

by the time results were released. The somewhat ‘black box’ nature of the calculations, with no individual agency, felt unfair according to some learners even before results were released (Montacute and Holt-White 2020). Once A level results were released, there were media reports of discontent and shock, and public protests (BBC news 2020; Busby 2020). Reflecting back on events, Roger Taylor, former Chair of Ofqual, described it as a ‘colossal error of judgement’ on the part of government ministers to think that calculated grades would be accepted by the public (Taylor 2021).

Inaccurate assumptions may have partly underpinned perceptions of low fairness. Media coverage commented that many A level grades submitted by teachers were ‘downgraded’ by calculations (e.g. Adams 2020). Such comments were based on a misconception that CAGs, rather than ranks, were the starting point for calculated grades and ignored that CAGs were higher than grades in previous years. The more pertinent issue was whether an algorithm could correctly identify which grades were ‘wrong’ (i.e. different to what they would have been if learning and exams had continued as normal) and appropriately adjust them. Some media coverage referred to the calculation method as a ‘rogue’ algorithm (Lamont 2021), a misleading phrase since a set of code simply conducts the programmed calculations or modelling and cannot ‘go rogue’. Such terminology potentially encouraged views that results were incorrect. In conclusion, calculated 2020 grades did not meet the legitimate expectations of many students. Whilst the appropriateness of assumptions underpinning perceptions of fairness could be debated, fairness in this sense appears to have been lower than in prior years since results were often lower than learners expected based on teacher feedback and as results were not entirely based on learners’ own performances.

Implied contractual: Summer 2020 final approach

The decision to use CAGs (except where calculated grades were higher) was publicly welcomed and appeared to satisfy students. CAGs may have been better accepted due to better alignment of grades with expectations based on teacher feedback, greater transparency, results being based directly on students’ own work, and/or simply because learners receiving a higher grade than expected are unlikely to complain.

Grade appeals are one potential indicator of satisfaction with grades. Rates were higher for summer 2020 than for summer 2019 (e.g. 2,220 GCSE appeals in summer 2020 compared to 745 in summer 2019) (Ofqual 2021b, 2022). This increase could suggest that final grades did not meet expectations for some learners. However, there are other potential explanations, including that (unlike usual years) unsuccessful appeals in summer 2020 were free.

Non-trivial numbers of learners chose to take autumn examinations (JCQ 2020a, 2020d), for example 15,881 A level grades were issued (JCQ 2020a), relative to 718,857 grades issued in summer 2020 (JCQ 2020b). This could include students whose grades did not meet their expectations. Alternatively, students may have felt they could perform better with further study and/or wanted to take exams to attempt to meet course entry requirements.⁵

Overall, the final summer 2020 approach seemed to meet most students’ expectations about grades and, therefore, grades can be considered of similar fairness in the implied contractual sense to usual years.

Implied contractual: Summer 2021 approach

For summer 2021, it was announced earlier in the year (January 2021) that teacher assessment would be used instead of exams (gov.uk 2021; House of Commons 2021), and the approach had an approximate precedent given the 2020 approach. Schools were not permitted to reveal TAGs before results days (Ofqual 2021f), nonetheless, learners may have had feedback on some elements of work that contributed to TAGs, informing their expectations about grades. This might plausibly have reduced the risk of results mis-aligning with student expectations.

Learners surveyed in April 2021 were more positive about likely fairness of results than those surveyed in April 2020 (Holt-White and Cullinane 2021). Once results arrived, there was no outcry of dissatisfaction, though there were still some isolated reports in the media of unexpected results (Fazackerley 2021b).

Numbers of appeals were considerably higher in summer 2021 than in summer 2020 or prior years (Ofqual 2022). These could have resulted from instances of unexpected grades or, again, there may be other explanations, such as no fees for unsuccessful appeals or fewer universities having capacity to accept learners who did not meet offer requirements (something that several commentators predicted, see Fazackerley 2021a). The latter raises an additional issue; reduced scope to gain course entry is arguably in conflict with legitimate expectations about the value of grades. UCAS (Universities and Colleges Admissions Service) data shows that numbers of applications to higher education increased in 2021 and whilst the overall numbers of offers also increased, the proportion of applications resulting in an offer reduced from 74% in 2020 to 69% for higher tariff institutions (Corver 2021).

The numbers of students taking examinations in the autumn were lower in 2021 than 2020 for both GCSEs and A levels (JCQ 2020a, 2020d; 2021a, 2021b, 2021d; 2022). In contrast to the appeals data, this could suggest that grades better matched expectations, though again there could be other reasons for autumn entries.

In general, summer 2021 TAGs seemed to meet expectations for the majority of learners. Thus, grades can be considered of similar fairness to usual years in the implied contractual sense.

Relational

Treating learners differently based on irrelevant factors would be unfair in the relational sense (Nisbet and Shaw 2020). This may be in terms of learning experiences and assessment approaches, and links to notions of comparability, in terms of applying the same standard across different learners and assessments (Newton 2007). This section considers whether learners were treated differently based on several irrelevant factors: year of assessment; learning opportunities; school attended; and personal characteristics. These factors are not mutually exclusive but provide a means to marshal discussion of somewhat different but related themes.

Relational – year of assessment

Comparability over time is a much-discussed theme in assessment, the evaluation of which is complex (for discussion, see Goldstein 1983; Newton 1997). Considerable effort

is made in usual circumstances to achieve comparability of standards between years (Newton 2007) or, to put it another way, to ensure that learners assessed in different years are treated the same. Once the government decided that exams would not go ahead in summer 2020 and 2021 but that learners would still be given grades, attempting to provide results that were comparable to previous years was fundamentally going to be more challenging than usual.

Relational (year): Summer 2020 intended approach. A key aim of using calculated grades was to achieve relational fairness between years, with standards maintained and a similar grade profile (DfE 2020a). At the time, more stakeholders agreed than disagreed (54% versus 33%) that the model should emphasise evidence of centres' past performance (Ofqual 2020d), suggesting that many understood the justification for this.

For A levels, the percentages of learners receiving grades A and above, and C and above, when summer 2020 grades were initially issued were only a little higher than in 2019 (27.6% compared to 25.2% and 78.0% compared to 75.5% respectively, JCQ 2020c). However, changes were larger for some subjects. For example, percentages receiving an A or above increased from 36.4% to 46.0% for A level French, from 18.0% to 27.9% for Drama, and from 53.1% to 61.8% for Further Mathematics (JCQ 2020c). As has been discussed elsewhere (Nye and Thomson 2020), larger differences in some subjects could have been an artefact of greater emphasis being placed on CAGs where only small numbers of candidates at a centre entered a particular subject. Alternatively, differences could partly reflect changes in the nature of the cohorts for some subjects and/or continued 'bedding in' of relatively new reformed qualifications.

There is also a question of how relationally fair grades were between years for individuals even if they were comparable for the cohort as a whole. Using forecast A level grades provided by teachers in 2014 close to the time of examination, Benton 2021 estimated reliability to be 0.72, whilst A level results based on exams have an approximate reliability of 0.93. This suggests that CAGs and rankings submitted by teachers in 2020 are likely to have lower reliability than grades from exams, and thus contain more error in individual learners' results.

Whilst the year-on-year relational fairness of calculated grades appears to have been similar to usual for the cohort as a whole, it appears to have been lower for individual subjects and learners.

Relational (year): Summer 2020 final approach. Abandoning the use of calculated grades (except where higher than CAGs) deprioritised year-on-year relational fairness. Final grades were considerably more generous than in 2019. For example, percentages of A levels in England awarded at A or above were 38.1% (compared to 25.2% in 2019) and for GCSEs 25.9% of grades were a 7 or above (compared to 20.7% in 2019) (JCQ 2020b, 2020e). Differences were larger for some subjects than others, for example, there was a large percentage increase in candidates gaining an A level grade A or above for Computing (17.6% to 36.2%) but a smaller increase for English Language (11.3% to 21.8%). Evaluating changes in assessment standards over time is more complex than simply interpreting an increase in the percentage of good grades as evidence that grading standards were more generous (Goldstein 1983; Newton 1997), as grades can also be affected by factors such as changes in the overall general ability of the cohort or changes in curriculum or teaching methods or in the kinds of learners studying certain

subjects. However, the large differences in results between 2019 and 2020 and the very different assessment method would suggest that in this case grades were more generous, on average.

One potential driver may have been a tendency for teachers to be generous in borderline cases (Benton 2021). For such learners, it may be difficult for teachers to predict which of them would and would not have achieved the grade in examinations. Therefore, it was perhaps reasonable to assign the higher grade to all who had a good chance of achieving it. Teachers reported pressure from senior management not to be generous compared to prior years (Holmes et al. 2021), which probably reduced risks of intentional generosity.

Using the higher of CAGs and calculated grades led to an easing of grading standards and, at the level of the whole cohort, fairness was lower than normal in the year-on-year relational sense. The effect was larger in some subjects than others, and, given that reliability is likely to have been lower (Benton 2021), some individual learners were probably affected more than others.

Relational (year): Summer 2021 approach. Using TAGs for summer 2021 again gave lower priority than usual to year-on-year relational fairness. Exam boards provided centres with support materials (including assessment materials and grade descriptors). These materials and the requirement for Centre Policies that set out grading approaches (JCQ 2021e) could potentially have better supported accuracy and consistency compared to summer 2020. However, reduced face-to-face time with the 2021 cohort was potentially challenging.

Grades in summer 2021 were more generous than 2019 and 2020 grades. The percentage of A level grades A or A* in England increased to 44.3% in 2021 (compared to 25.2% in 2019 and 38.1% in 2020) (JCQ 2021b) and for GCSEs 28.5% of grades awarded were at 7 or above in 2021 (compared to 20.7% in 2019 and 25.9% in 2020) (JCQ 2021c). Differences were greater in some subjects than others.

Again, one mechanism driving this generosity may have been teachers giving benefit of the doubt to borderline learners (Benton 2021). Additionally, the assessment approach for TAGs allowed teachers to use a series of tests and other evidence, meaning students could revise for each in turn and had multiple opportunities to perform well (Stannard 2021). Furthermore, teachers knew that their grades would be the final grades (unless changed through the quality assurance process or appeal) which could, theoretically at least, have tempted teachers towards generosity. Pressure from parents and students (Holt-White and Cullinane 2021) could have increased this risk.

The evidence indicates that year of assessment affected results and, therefore, year-on-year relational fairness was lower than usual at the cohort level, for different subjects, and for individual learners.

Relational – learning opportunities

In any educational system, there is potential for inherent differences in learning opportunities, for example, due to socio-economic status (Shavit and Blossfeld 1993). This can affect exam results, and therefore affect grade fairness in the relational sense. Whilst this issue arises in ‘normal’ circumstances, here we focus on variation in learning opportunities created or exacerbated by the pandemic.

The closure of schools to most learners from March to July 2020 and January to March 2021 led to remote learning for many, and even for those attending school the structure and nature of teaching was different. Schools varied in their preparedness for remote learning (Johnson and Coleman 2021). There were differences in the quality of remote learning between state and private schools, and according to socio-economic status and ethnicity (Coleman 2021; Holt-White and Cullinane 2021; Montacute and Cullinane 2021; The Sutton Trust 2021). Differences in home circumstances, such as internet access, computer or digital device availability, appropriate study space and parental support, affected learning (Coleman 2021). Even when schools were open, attendance levels were affected by individuals or groups needing to self-isolate or whole schools needing to temporarily close (Holt-White and Cullinane 2021). Analysis shows that school closures and attendance varied by region, with schools in the North and London being more affected (Anders, Macmillan, and Wyness 2021; Holt-White and Cullinane 2021).

Evidence suggests that independent learning was more successful among higher achievers, and that less affluent students disengaged more (Johnson and Coleman 2021; Redmond, Golding, and Grima 2021). The impact on learning for each individual will have been a complex interaction of school and home setting, personality, motivation, social confidence and the practical and emotional disruption suffered.

Relational (learning opportunities): Summer 2020 intended and final approaches. The intended and final assessment approaches are considered together since the same learners and their learning opportunities are relevant. For these learners, teaching and learning were unaffected by the pandemic until towards the end of their two-year courses. Most course content is likely to have been covered before schools closed, though students may have missed out on face-to-face teaching of some final topics and on revision and exam preparation (which is common at this time of year, Thorn and Vincent-Lancrin 2021). Differences in internet and device access, availability of a quiet study space, and parental support, affected learning opportunities from March 2020 onwards (Anders, Macmillan, and Wyness 2021; Coleman 2021; Montacute and Cullinane 2021; The Sutton Trust 2021).

Affluent schools more often had access to appropriate software platforms, and therefore used more online lessons during the first lockdown (Coleman 2021). There was greater concern amongst teachers for students with special educational needs, mental health issues, caring responsibilities and less parental support (Johnson and Coleman 2021). Lack of devices was more frequently a concern in lower income families (Thorn and Vincent-Lancrin 2021). Furthermore, parental education level influenced parents' confidence regarding home schooling (Thorn and Vincent-Lancrin 2021).

Teachers could perhaps partly compensate for differences in learning opportunities when making assessment judgements for summer 2020, given the aim to provide learners with the grades they would have achieved if teaching and assessment had been unaffected (DfE 2020c; Jadhav 2021). This potentially allowed teachers to evaluate students' work only up to March 2020 or to take known challenges experienced by learners into account. Nonetheless, teachers might have taken different approaches. Using calculated grades could potentially have mitigated for differences in how teachers dealt with lost learning.

In conclusion, calculated grades and final summer 2020 grades may have been slightly less fair than usual in terms of differences in learning opportunities.

Relational (learning opportunities): Summer 2021 approach. Learners completing their GCSEs or A levels in summer 2021 experienced longer periods of remote learning and disruption during their two-year courses. Again, learning opportunities may have varied between schools and between individual learners based on differences in the quality and quantity of lessons and other resources provided by schools, and differences in technology, space and support at home (Holt-White and Cullinane 2021; Montacute and Cullinane 2021). In 2021, learning opportunities were reportedly affected by the need for schools to focus on assessment judgements, with increased time spent on revision at the expense of new content (Johnson and Coleman 2021).

Evidence suggests differences in the quality of remote learning between state and private schools, and along lines of socio-economic status, region, special educational needs status and ethnicity (Holt-White and Cullinane 2021; Montacute and Cullinane 2021; The Sutton Trust 2021; Coleman 2021, Johnson and Coleman 2021). There were higher absences for students from lower socio-economic status backgrounds, those with special educational needs, those eligible for free school meals, and for certain ethnic groups (DfE 2021b; Nelson, Lynch, and Sharp 2021).

Assessing learners only on the content taught may have partly compensated for differences in learning loss. However, this may not have addressed individual-level learning loss (Holt-White and Cullinane 2021). Learners could inform teachers of mitigating circumstances (DfE 2021c), however, it is unclear whether this occurred in practice. Additionally, whilst illness could be taken into account when making judgements, lost learning due to the circumstances of the pandemic was not seen as a case for special consideration (JCQ 2021e).

The evidence indicates that 2021 grades were less fair than usual in the relational sense as learning opportunities varied.

Relational – centre

Ideally, students should not be assessed differently because of the centre, or type of centre, they attend as this would reduce comparability and fairness. As already discussed, centre attended potentially affected learning opportunities during school closures; this section discusses other ways that centre or centre type affected grade fairness.

Relational (centre): Summer 2020 intended approach. By its very nature, assessing learner performance involves qualitative judgement, requiring the assessor to have a concept of the quality of work necessary to a task and the ability to evaluate learner performance relative to that concept (Sadler 1989). Such judgements often require consideration of multiple criteria, selected as appropriate from a pool of possible criteria (Sadler 1989). Because such judgements are complex, particularly in a context where teachers were attempting to make a holistic judgement of learner performance on a variety of tasks over a period of time, there will have been a risk that different teachers made judgements differently and applied different standards. A key aim of calculated grades was to standardise results across centres (DfE 2020a; Ofqual 2020h).

Teachers reported differences in their approaches to making judgements (Ofqual 2020a), with some schools relying heavily on learner data and others using some intuition (Johnson and Coleman 2021). Performance in mock exams, and in other tests or tasks, were key for many schools (Ofqual 2020a). As mentioned in the section on formal

fairness, centres in summer 2020 varied in whether they used the profile of historical results in a subject as the starting point when determining CAGs or as a later check (Ofqual 2020a). Ofqual's early analysis of CAGs suggested that centres might have applied different levels of optimism (2020a).

The procedure for calculating grades attempted to predict results based on a centre's results in that subject in recent years and considering prior attainment of the current students (Key Stage 2 tests⁶ for GCSE, and GCSE for A level). The rank orders provided by centres were maintained, but grade boundaries were effectively shifted, if necessary (Ofqual 2020a). This process, by its very nature, treated students differently depending on their centre, but with the aim of adjusting for differences between centres in how students were evaluated. In public discourse at the time, concern was expressed that learners could be negatively affected by weak performance of their centre's prior cohorts (Ofqual 2020d).

The calculation process treated some centres differently where the performance data or prior attainment data did not provide a sound basis for calculation (Ofqual 2020a, d). Additionally, because the performance of a centre's learners in small entry subjects was unlikely to be stable over time, greater emphasis was placed on CAGs in these cases (Ofqual 2020a). There was concern that this favoured learners at smaller schools and/or those taking more unusual subjects and thus might favour independent schools (Ofqual 2020a).

With regard to relational fairness across centres, potential differences in standards applied by centres and differences in whether or not standardisation could be applied, mean that calculated grades were less fair than usual grades.

Relational (centre): Summer 2020 final approach. As already described, centres varied in 2020 in the types of learner evidence used, how prior centre results were used and the standard applied (Johnson and Coleman 2021; Holmes et al. 2021; Ofqual 2020a). The late decision to primarily use CAGs in summer 2020 removed issues around the standardisation process treating different centres differently, but, resulted in little mitigation for differences in the standards applied by centres. Retaining calculated grades where higher than CAGs may have addressed cases where centres made particularly severe judgements.

Final grades appeared to benefit independent school students. For A levels, percentages of learners receiving grades A or A* compared to 2019 increased more for independent schools than for almost all other centre types (Ofqual 2020i). For GCSEs, percentages of learners at grade 7 or above increased most compared to 2019 for independent and other selective schools (Ofqual 2020i). Ofqual argued that this apparent advantage was probably due to more learners at such schools being on the boundary of high grades and being given benefit of the doubt (Ofqual 2020i).

With the likelihood that centres used different approaches and applied slightly different standards, and the lack of standardisation across centres, final summer 2020 grades can be considered much less fair than pre-pandemic grades for relational fairness across centres.

Relational (centre): Summer 2021 approach. For summer 2021, Centre Policies (JCQ 2021e) and support materials provided by awarding bodies (including grade descriptors and past paper exam questions) offered some potential to facilitate greater consistency in assessment approach and standards. However, teachers and students

expressed concern that different centres could use different approaches (DfE and Ofqual 2021). Indeed, as well as some similarities in learner evidence used and in procedures for marking and grading learners, considerable differences have been observed (Vitello and Leech 2022). Most teachers reported using at least one assessment conducted under exam conditions but centres varied in their use of past papers, mock exams, teacher-written tests, classwork and homework and in the number of assessments used per subject (Holt-White and Cullinane 2021). Given the guidance to assess learners only on the content taught (Ofqual 2021e), there was probably variation in the proportion of course content assessed at different centres. Research analysing student evidence samples submitted to an awarding body suggests that there was variation in the volume of evidence considered (Vitello and Leech 2022). There was no stated minimum with teachers left to decide how much they considered ‘sufficient content to form the basis of a grade’ (JCQ 2021e, 18).

Differences have been found in whether centres converted from assessment marks to grades by using grade descriptors, by using grade boundary marks from the past papers taken by learners, whether they adapted grade boundaries in some way, or (in one instance) used grade boundaries from a prior year (Vitello and Leech 2022). It seems unlikely that this level of variety in approaches to evidence and grade assignment, however diligently conducted, could result in all centres applying the same standard.

More teachers at independent schools reported pressure from parents about learners’ grades than teachers at other school types (Holt-White and Cullinane 2021). Any differences between schools in whether teachers succumbed to such pressure would be a concern for relational fairness.

As with 2020 grades, the performance gap between independent and state schools widened in summer 2021 (Ofqual 2021k). Ofqual analysis showed a small relative increase in mean A level grades at independent schools compared to academies once prior attainment was controlled for, but a relative decrease in grades C or above (Lee 2021). Small relative decreases in A level performance for learners at secondary selective schools, sixth forms and tertiary colleges compared to learners at academies (matched for prior attainment) were also found (Lee 2021). At GCSE, a decrease was seen for secondary selective schools compared to academies (learners matched by prior attainment). Additionally, there was a relative increase in GCSE outcomes for independent schools compared to academies in terms of mean grades and numbers achieving grade 7 or above, but not for grade 4 or above (Lee 2021). Differences could suggest that more generous standards were used at independent schools or, as described earlier, that such centres have more learners around the borderline of high grades who benefited from a tendency to give the benefit of the doubt in borderline cases. Alternatively, changes in relative performance between centre types could result at least partly from differences in learning opportunities rather than representing a difference in the standards applied by teachers at different centre types.

In summer 2021, learners at all centres were treated the same in that TAGs were used. However, the types and volumes of evidence used and approaches to grading varied between centres with little means for differences in grading standards to be corrected. Therefore, summer 2021 grades seem likely to be much less relationally fair between centres than pre-pandemic grades.

Relational – other irrelevant factors

There is potential for relational unfairness if some groups or individuals are treated differently for irrelevant reasons. This is possible in normal circumstances, hence the avoidance of bias being a key focus in the assessment literature on fairness (AERA/APA/NCME 2014). Evidence regarding bias in teacher assessments suggests that teachers may over-estimate grades for learners from lower socio-economic backgrounds (Wyness 2017) and ethnic minorities (Murphy and Wyness 2020) or favour learners based on personality (Papageorgiou et al. 2020) (For a literature review on bias in teacher assessment see Lee and Newton 2021; Lee and Walter 2020).

Relational (other): Summer 2020 intended approach. Ofqual stated that teacher judgements should be objective and that learner characteristics should not be allowed to affect judgements (Ofqual 2020h). Teacher guidance included advice on avoiding unconscious bias and encouraged reflection on whether preconceptions might be influencing evaluations (Ofqual 2020h). Nevertheless, the degree to which people can effectively adjust for unconscious bias in their thinking is debated (see Noon 2018). Most teachers were confident that judgements were bias-free (Ofqual 2021g).

When determining the calculation model, the option least likely to worsen existing inequalities for learners with protected characteristics was selected (Ofqual 2020a). Nonetheless, the calculation method could not adjust for any bias in teachers' judgements of how individual learners compared to each other as it maintained the rank orders provided by centres. Stakeholders expressed concern around there being no mechanism to address biases for groups (House of Commons Education Committee 2020a), with this seen as potentially capping outcomes (Holt-White and Cullinane 2021). Additionally, concerns were raised about the suitability of using prior data in calculations for certain groups, for example, the performance of cohorts with special educational needs may fluctuate more from year to year than that of other cohorts (Ofqual 2020d).

The appeals process could potentially have corrected for biases but high volumes could have made this unmanageable (Nisbet 2020). Whilst Ofqual provided guidance and helplines were available (e.g. National Careers Service) (Ofqual 2020g), some commentators argued that grounds for appeal were narrow and that families from more privileged backgrounds (as in usual years) would find the appeals system easier to navigate (Sherwood and Cuthbert 2020).

Ofqual analysis found that performance gaps between groups (based on gender, ethnicity, free school meal status, first language, socio-economic background and special education needs) were very similar to prior years, with minor exceptions (Ofqual 2020a). They concluded that there was no evidence of systematic bias (Lee, Stringer, and Zanini 2020; Ofqual 2020a, 2021g).

The vast majority of teachers will have done their utmost to treat all learners the same when making their judgements, and Ofqual's guidance should have encouraged reflection regarding potential bias. However, unconscious bias could have affected teacher judgements, and the calculation process would not have been able to address this. Therefore, we argue that the relational fairness to groups and individuals may have been slightly lower than usual.

Relational (other): Summer 2020 final approach. Many of the themes discussed in the previous section apply here. Again, there was some risk of unconscious bias, though

Ofqual guidance may have encouraged reflection to reduce this (Ofqual 2020h), and teachers were confident that judgements were bias-free (Ofqual 2021g).

A category for ‘bias or discrimination’ was added to the possible types of assessment malpractice in summer 2020. Some suspected cases were investigated but no penalties were issued (Ofqual 2020e, 2021h). However, this does not rule out that instances of bias occurred but went unreported or could not be evidenced.

Interestingly, Ofqual’s equalities analysis found that the relationships between student characteristics and results were less closely maintained compared to prior years with final 2020 grades than with calculated grades (Lee, Stringer, and Zanini 2020; Ofqual 2021g). This suggests relational fairness issues with regard to how different groups were evaluated, though changes in patterns were small and Ofqual again concluded that there was no systematic bias (Lee, Stringer, and Zanini 2020; Ofqual 2021g).

For the final 2020 grades, there was again some risk that particular groups or individuals were inadvertently treated differently. There was no evidence of systematic bias, but some changes in patterns of performance mean that bias cannot be ruled out. Therefore, the relational fairness of grades in terms of irrelevant group or individual characteristics may have been slightly lower than usual.

Relational (other): Summer 2021 approach. Again, there was potential for unconscious bias in judgements based on learner characteristics, which Ofqual tried to counter through guidance (Ofqual 2021i, 2021f). Stakeholders noted that achieving objectivity can be difficult even for trained examiners when marking exams, so TAGs could not be expected to be perfect (Ofqual 2021a).

Periods of remote learning may have reduced teachers’ familiarity with students’ work, however, increased opportunities to gather evidence for assessment compared to 2020 may have improved consistency between learners in the evidence available. Centres were advised to provide reasonable adjustments and access arrangements when gathering assessment evidence from eligible learners, or to take into account learner needs when making judgements (Ofqual 2021c, 2021j).

Teachers were advised to take temporary learner injury, illness and other personal circumstances (but not lost learning due to Covid-19) into account when making judgements (JCQ 2021e; Ofqual 2021j). This echoed ‘special consideration’ arrangements available in usual circumstances to adjust marks post hoc in such cases. Lost learning due to the circumstances of the pandemic was not seen as a case for such consideration as this was addressed through flexibility in evidence types and the assessment only of taught topics (JCQ 2021e).

No malpractice penalties were issued in summer 2021 for ‘bias or discrimination’ (Ofqual 2021h).

Ofqual analyses found some small changes in the relationships between learner characteristics and results in relation to region, ethnicity, special educational needs, gender, and free school meal eligibility (Lee 2021). These changes may at least partly reflect differences in learning loss rather than biased judgements, but it is difficult to know (Lee 2021; Witherspoon 2021).

Overall, the vast majority of teachers will have done their utmost to assign grades based on relevant factors. However, it is difficult to rule out bias and there is evidence of some shifts in performance gaps. On balance, we judge relational fairness of grades with regard to other irrelevant factors as lower than usual.

Retributive

The notion of retributive fairness draws from the concept of retributive justice used in theorisation about punishment. Within the concept of retributive justice, those committing wrongful acts are seen to morally deserve an appropriate and proportionate punishment (Moore 1997). In educational assessment, retributive fairness relates to whether candidate results appropriately reward learner performance, thus providing grades that are deserved (Nisbet and Shaw 2020).

Considering what is deserved is complex in the context of pandemic-era grades. Given reduced learning opportunities, it is less likely that learners could meet the usual requirements met by prior learners of similar ability. Strictly, this means fewer 2020 and 2021 students ‘deserve’ high grades. Nevertheless, lost learning was not the students’ fault and many adapted and reportedly worked hard despite disruption (House of Commons Education Committee 2020b). Therefore, there may be a moral case for taking a supportive stance (Shaw and Nisbet 2021).

Retributive: Summer 2020 intended approach

For the 2020 cohort, the fact that teaching was not disrupted until late in their courses should have aided teachers’ judgements of learner achievement in relation to specification criteria. However, challenges in assessing what learners deserved may have included limits in available student evidence, and how to predict which students might, in usual circumstances, have performed better or worse than expected.

The generosity of 2020 CAGs compared to exam grades in previous years means that some students were probably given higher CAGs by teachers than they deserved in terms of criteria met. In theory, calculated grades could have improved retributive fairness by adjusting standards and getting closer to awarding students what they deserved (on average) given their performances in relation to specification requirements. However, given the probable lower reliability of teacher assessment compared to exam-based grades (Benton 2021), calculated grades were probably lower than deserved for non-trivial numbers of students. Cases where grades were perceived as lower than deserved were problematic to public perceptions (Lamont 2021; Nisbet 2020) and such views were probably justified given the lower reliability. Therefore, we argue that calculated 2020 grades were less fair than usual in the retributive sense.

Retributive: Summer 2020 final approach

Again, the undisrupted face-to-face learning for much of learners’ two-year courses should have aided teachers in assessing what was deserved but accurately evaluating learners based on sometimes limited evidence and predicting which learners might have performed better or worse in exams than expected would have been difficult.

Analysis of summer 2020 CAGs found that the association between grades and prior attainment was slightly stronger in summer 2020 than in prior years (Stratton, Zanini, and Noden 2021). This could suggest that CAGs eliminated some of the unpredictable variations in grades that occur in usual circumstances and, thus, were a better reflection

of deserved grades, or that teachers relied too much on prior attainment (Stratton, Zanini, and Noden 2021).

The higher numbers of appeals than in prior years (Ofqual 2021b, 2022) may indicate that some students (and parents) felt that grades did not reflect what was deserved. However, other factors, such as lack of fees, could also explain this. Students choosing to sit examinations in the exceptional autumn 2020 series (JCQ 2020a, 2020d) may have felt that they did not receive the grades deserved, or may have sat exams for other reasons.

In the context of lower reliability assessment (Benton 2021) with a tendency for generosity, using the better of CAGs and calculated grades probably means that more students than usual received grades higher than deserved. This could be considered less fair on peers who received the ‘correct’ grade. However, perhaps more importantly, the final grading approach should mean that most learners received *at least* the grade deserved and no more students than usual were disadvantaged by measurement error (Benton 2021). This seems like the more just situation given the difficult circumstances and given we cannot know which students were advantaged or disadvantaged by lower reliability. Therefore, 2020 final grades can be considered of similar fairness to prior years in the retributive sense.

Retributive: Summer 2021 approach

Reduced face-to-face time with learners over their two-year courses (resulting in less learning evidence, Johnson and Coleman 2021), potentially hindered teachers’ ability to award the grades deserved. Positives for 2021 were greater notice that teachers would assess students and schools being open to most students from March 2021 onwards, providing more opportunity for evidence gathering and greater learner agency. Nonetheless, reduced learning opportunities (Montacute and Cullinane 2021) meant that learners may have had reduced knowledge, understanding and skills compared to previous cohorts, making it harder to meet assessment criteria through no fault of their own. Assessing only taught content may have partly addressed this. However, how teachers made judgements for learners with less learning access than others could have varied. Some teachers reported organising additional assessment opportunities for students who were thought to be under-performing due to the circumstances (Johnson and Coleman 2021). Only 52% of teachers surveyed by the Times Educational Supplement felt that the grading process at their school would give all students the grades deserved (Lough 2021a). Although it is difficult to verify the robustness of this survey conducted by a press organisation, even if this was an underestimate this finding may be concerning. On the other hand, further detail, not provided in reporting on the survey, would be useful, such as whether these teachers felt that a small or large number of learners would not get the grade deserved. The survey finding could simply reflect teacher awareness of the potential for greater measurement error.

There were increased numbers of appeals to summer 2021 grades compared to summer 2020 and prior years (Ofqual 2022) but numbers of students taking examinations in autumn 2021 were lower than in autumn 2020 (JCQ 2020a, 2020d; 2021a, 2021b, 2021d; 2022). Appeals and autumn entry numbers do not provide a clear picture of whether perceived retributive fairness was similar to that in the previous year (and as noted earlier, appeals and autumn exam numbers could reflect a number of factors).

As with the 2020 final approach, the tendency for generous grades should mean that, despite the probable lower reliability of grades compared to exam grades, most students received *at least* the grade they deserved. It is possible (perhaps likely) that some students received grades higher than they deserved based on the criteria met, particularly given lost opportunities to develop relevant knowledge, understanding and skills. However, it seems morally appropriate to factor in the circumstances when considering what learners deserve. Therefore, 2021 grades can be considered of similar fairness to usual in the retributive sense.

Summary

Having discussed the issues affecting the fairness of summer 2020 and 2021 GCSE and A level grades in England in relation to Nisbet and Shaw's (2020) senses of fairness drawing on available evidence, this section summarises our judgements and provides an overall perspective on the relative fairness of grades in these years.

Table 1 summarises the judgements about grade fairness. It should be noted that the robustness of some of the currently available evidence varies, that these judgements potentially involved an element of subjectivity and that, consequently, refinement of these evaluations may be appropriate as further evidence emerges.

A key overall pattern is an apparent dilemma between prioritising relational fairness (as in the planned approach for summer 2020, in line with Ofqual's usual remit, DfE

Table 1. Judgements regarding the fairness of summer 2020 and 2021 grades for GCSE and A level, for the learners receiving them, compared to usual years.

	2020 intended: Calculated grades	2020 final: Centre assessment grades or calculated grades if better	2021: Teacher-assessed grades
Formal	Same as usual	Same as usual	Similar to usual
Implied contractual	Lower than usual	Similar to usual	Similar to usual
Relational – year of assessment	Similar to usual at the level of the whole cohort, ...	Much lower than usual at the level of the whole cohort, ...	Much lower than usual at the level of the whole cohort, ...
	...but lower for individual subjects and individual learners	...and lower for individual subjects and individual learners	...and lower for individual subjects and individual learners
Relational – learning opportunities	Slightly lower than usual	Slightly lower than usual	Lower than usual
Relational – centre	Lower than usual	Much lower than usual	Much lower than usual
Relational – other	Slightly lower than usual	Slightly lower than usual	Lower than usual
Retributive	Lower than usual	Similar to usual	Similar to usual

Key:

Same
Similar
Slightly lower
Lower
Much lower

2020a), and prioritising implied contractual and retributive fairness (as in the final approaches in both years). The latter arguably reflects a demand for justice for learners faced with difficult circumstances. Nisbet and Shaw (2022) discuss this tension in terms of conflicting focuses on equality (relational fairness) and what is deserved (retributive fairness). Public opinion drove final decisions, with the kinds of unfairness that people were willing to tolerate and not tolerate becoming paramount to acceptance. Whilst the increases in percentages of learners achieving high grades would not normally be considered sufficiently fair in terms of year-on-year relational fairness, relaxing the usual threshold in order that the majority of learners received *at least* the grade they deserved was arguably appropriate given the likely reduction in reliability (as discussed by Benton 2021).

Taking a situated view and conceptualising fairness as a continuum (Nisbet and Shaw 2020), we would argue that the three assessment approaches evaluated in this article were probably as fair as was possible in the circumstances, with different senses of fairness prioritised in different options. Summer 2021 was perhaps slightly less fair overall than both 2020 approaches, but this was driven by lost learning opportunities rather than by the assessment approach. In the instance of any major future disruptions to assessment, it would be worthwhile for policy makers to take a situated view of fairness from the start and to explicitly consider different senses of fairness, keeping in mind that fairness should be treated as a continuum.

In evaluating fairness in this article, judgements were made relative to prior years when GCSEs and A levels were primarily assessed using examinations. However, achieving *absolute* fairness with any assessment is, arguably, unrealistic (Nisbet and Shaw 2022), and, therefore, GCSE and A level grades resulting primarily from examinations are unlikely to be *absolutely* fair. For example, learning opportunities may vary between learners even in ‘normal’ circumstances (Shavit and Blossfeld 1993) and some learners may not produce their best work during the pressurised situation of an exam due to anxiety (Hembree 1988). Evaluating the fairness of GCSE and A level grades in usual circumstances was beyond scope, but is, of course, an important issue where continued effort is needed to investigate and minimise issues threatening fairness. Nonetheless, using the fairness of GCSE and A level grades in prior years as the reference point for the current analysis felt appropriate, since these are long-established qualifications and, as discussed by others, the ‘disruption to the system in 2020 and 2021 was not caused by exams but by their absence’ (Christodoulou 2022).

Notes

1. GCSEs (General Certificate of Secondary Education), AS (Advanced Subsidiary) levels and A (Advanced) levels are subject-based general qualifications taken by many learners in England, usually at ages 16, 17 and 18 years respectively.
2. ‘Construct-irrelevant variance’ exists where there is excess variance in results that is not relevant to the knowledge, understanding and skills that the assessment was intended to measure (Messick 1989).
3. Exact number is not stated.
4. https://www.investopedia.com/terms/i/implied_contract.asp; https://en.wikipedia.org/wiki/Contract#Implied_terms

5. The use of grades from summer 2020 and autumn 2020 exams as alternatives raises issues of relational fairness between these two assessment opportunities. However, this is not discussed further here as it goes beyond the current article's focus.
6. Key stage 2 tests relate to the national curriculum in England and are taken by many learners at around age 11 years.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Victoria Crisp is a Senior Researcher in the Assessment Research and Development division at Cambridge University Press & Assessment. Her areas of interest include question design, validity and validation, comparability issues and methods, judgement processes in assessment, and digital assessment. Her doctoral research, conducted at the Institute of Education, London, explored the judgement processes underpinning marking by teachers.

Gill Elliott is the Deputy Director of the Research Division in the Assessment Research and Development division at Cambridge University Press & Assessment. Gill leads the research strand on comparability and has close involvement with the overall management of the Research Division. Her research interests include many aspects of comparability, particularly those related to educational change.

Emma Walland is a Researcher in the Assessment Research and Development division at Cambridge University Press & Assessment. She holds a master's degree in social and psychological research from the University of the Witwatersrand, Johannesburg and has recently completed a Postgraduate Advanced Certificate in Educational Studies: Educational Assessment from the University of Cambridge. Her current research interests include assessment and education reform, technological innovations in education and assessment, and marking strategies and practices.

Lucy Chambers is a Principal Researcher in the Assessment Research and Development division at Cambridge University Press & Assessment. She leads the research strand on qualifications and assessment. Her current interests include the moderation of school-based assessment, investigating the validity of comparative judgement and research data management.

References

- Adams, R. 2020. "Nearly 40% of A-Level Result Predictions to Be Downgraded in England." *The Guardian*, August 7.
- AERA/APA/NCME. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Anders, J., L. Macmillan, and G. Wyness. 2021. "How Should We Assess School Students Now That Exams Have Been Cancelled?" UCL Institute of Education Blog. Accessed January 8. <https://blogs.ucl.ac.uk/ioe/2021/01/08/how-should-we-assess-school-students-now-that-exams-have-been-cancelled/>.
- ASCL. 2020a. *Coronovirus (Covid-19): Guidance Regarding Centre-Assessed Grades for Summer 2020*. Leicester: Association of School and College Leaders. <https://www.ascl.org.uk/ASCL/media/ASCL/Help%20and%20advice/Leadership%20and%20governance/Coronavirus-Guidance-regarding-centre-assessed-grades-for-summer-2020.pdf>.
- ASCL. 2020b. *Step by Step Approach to Allocating Grades Using ASCL Approach*. Leicester: Association of School and College Leaders. <https://www.ascl.org.uk/ASCL/media/ASCL/Help%20and%20advice/Step-by-step-to-allocating-grades-using-ASCL-approach-v2c-D-Blow.pdf>.

- BBC news. 2020. "A-Levels and GCSEs: Student Tells Minister 'You've Ruined My life'." *BBC news*, August 15.
- Benton, T. 2021. "On Using Generosity to Combat Unreliability." *Research Matters: A Cambridge Assessment Publication* (31): 22–41. doi:10.17863/CAM.100414.
- Busby, M. 2020. "A-Level Students Speak: 'I Always Dreamed of Going to Cambridge'." *The Guardian*, August 15.
- Christodoulou, D. 2022. "Is This the Future of Assessment?" *Times Educational Supplement Magazine*, December 14.
- Coleman, T. 2021. *Digital Divide in UK Education During COVID-19 Pandemic: Literature Review*. Cambridge Assessment Research Report. <https://www.cambridgeassessment.org.uk/Images/62843-digital-divide-in-uk-education-during-covid-19-pandemic-literature-review.pdf>.
- Corver, M. 2021. "Harder Times Ahead for Applicants and Access." DataHE. Accessed July 26. <https://datahe.uk/2021/07/26/harder-times-applicants-access/>.
- Coughlan, S. 2021. "'Pointy-elbow' Parents Emailing Teachers to Boost Exam Grades." *BBC news*, March 11.
- DfE. 2020a. *Direction Under S 129(6) of the Apprenticeships, Skills, Children and Learning Act 2009. Letter Dated 31 March 2020*. London: Department for Education.
- DfE. 2020b. *Schools, Colleges and Early Years Settings to Close*. London: Department for Education. <https://www.gov.uk/government/news/schools-colleges-and-early-years-settings-to-close>.
- DfE. 2020c. *Taking Exams During the Coronavirus (COVID-19) Outbreak*. London: Department for Education. <https://www.gov.uk/government/publications/coronavirus-covid-19-cancellation-of-gcses-as-and-a-levels-in-2020>.
- DfE. 2021a. *Direction Under S 129(6) of the Apprenticeships, Skills, Children and Learning Act 2009. Letter Dated 23 February 2021*. London: Department for Education.
- DfE. 2021b. *Pupil Absence in Schools in England: Autumn Term*. London: Department for Education. <https://explore-education-statistics.service.gov.uk/find-statistics/pupil-absence-in-schools-in-england-autumn-term>.
- DfE. 2021c. "What You Need to Know About Grades in 2021." *The Education Hub Blog*, 25 February 2021. <https://educationhub.blog.gov.uk/2021/02/25/what-you-need-to-know-about-grades-in-2021/>.
- DfE, and Ofqual. 2021. *Analysis of Consultation Responses: How GCSE, AS and A Level Grades Should Be Awarded in Summer 2021*. Coventry: Ofqual.
- Elliott, G. 2021. "Generation Covid and the Impact of Lockdown." *Research Matters: A Cambridge Assessment Publication* (31): 68–83. doi:10.17863/CAM.100413.
- Fazackerley, A. 2021a. "A-level Results: Top Universities Expected to Stick Rigidly to Offer Grades." *The Guardian*, July 31.
- Fazackerley, A. 2021b. "'I'm Lost': Poorer Pupils Lose University Places after A-level Grade Surge." *The Guardian*, August 25.
- Goldstein, H. 1983. "Measuring Changes in Educational Attainment Over Time: Problems and Possibilities." *Journal of Educational Measurement* 20 (4): 369–377.
- gov.uk. 2021. *Prime Minister's Address to the Nation*, 10. Downing Street, London: Prime Minister's Office.
- Hembree, R. 1988. "Correlates, Causes, Effects, and Treatment of Test Anxiety." *Review of Educational Research* 58 (1): 47–77.
- Hilliard, A., N. Guenole, and F. Leutner. 2022. "Robots are Judging Me: Perceived Fairness of Algorithmic Recruitment Tools." *Frontiers in Psychology* 13:940456. doi:10.3389/fpsyg.2022.940456.
- Holmes, S. E., D. Keys, D. T. Churchward, and B. Black. 2021. *Centre Judgements: Teaching Staff Survey, Summer 2020*. Coventry: Ofqual.
- Holt-White, E., and C. Cullinane. 2021. *A Levels and University Access 2021*. London: The Sutton Trust.
- House of Commons. 2021. *House of Commons Debate. Covid-19: Educational Settings. Vol 686 Col 763, January 6*. London: House of Commons. <https://hansard.parliament.uk/commons/2021-01-06/debates/D2A64DD9-6A6B-4EAB-8001-C3BACDE25B02/Covid-19EducationalSettings>.

- House of Commons Education Committee. 2020a. *Getting the Grades They've Earned. Covid-19: The Cancellation of Exams and 'Calculated' Grades. First Report of Session 2019-21. Report, Together with Formal Minutes Relating to the Report*. HC 617. London: House of Commons.
- House of Commons Education Committee. 2020b. *Getting the Grades They've Earned. Covid-19: The Cancellation of Exams and 'Calculated' Grades. First Report of Session 2019-21. Volume II: Oral and Written Evidence*. HC 617-Evidence. London: House of Commons.
- Howard, E., S. Holmes, and N. Case. 2021. "Qualification Results in England in Summer 2020: Teacher Judgement in Action." Annual conference of the Association for Educational Assessment in Europe; November 2021, Online.
- Jadhav, C. 2021. "Teacher Judgements in 2021: What We Can Learn from 2020." *The Ofqual Blog*. Accessed May 17. <https://ofqual.blog.gov.uk/2021/05/17/teacher-judgements-in-2021-what-we-can-learn-from-2020/>
- JCQ. 2020a. *GCE A Level & GCE AS Level: Results Autumn 2020*. London: Joint Council for Qualifications.
- JCQ. 2020b. *GCE A Level & GCE AS Level Results Summer 2020. Thursday 10 September 2020*. London: Joint Council for Qualifications.
- JCQ. 2020c. *GCE A Level & GCE AS Level Results Summer 2020. Thursday 13 August 2020*. London: Joint Council for Qualifications.
- JCQ. 2020d. *GCSE (Full Course). Outcomes for Key Grades for England Including Age Breakdowns. Results Autumn 2020*. London: Joint Council for Qualifications.
- JCQ. 2020e. *GCSE (Full Course). Outcomes for Key Grades for UK, England, Northern Ireland and Wales, Including UK Age breakdowns. Results Summer 2020. Thursday 10 September 2020*. London: Joint Council for Qualifications.
- JCQ. 2020f. *JCQ Update on the Summer Series 2020. 6 May 2020*. London: Joint Council for Qualifications.
- JCQ. 2021a. *GCE A Level & GCE AS Level Results Autumn 2021. Thursday 16 December 2021*. London: Joint Council for Qualifications.
- JCQ. 2021b. *GCE A Level & GCE AS Level: Results Summer 2021. Tuesday 10 August 2021*. London: Joint Council for Qualifications.
- JCQ. 2021c. *GCSE (Full Course). Outcomes for Key Grades for UK, England, Northern Ireland & Wales, Including UK Age Breakdowns. Results Summer 2021. Thursday 12 August*. London: Joint Council for Qualifications.
- JCQ. 2021d. *GCSE (Full Course). Outcomes for Main Grade Set for Each Jurisdiction. Results summer 2021. Thursday 12 August 2021*. London: Joint Council for Qualifications.
- JCQ. 2021e. *JCQ Guidance on the Determination of Grades for A/AS Levels and GCSEs for Summer 2021: Processes to Be Adopted by Exam Centres and Support Available from Awarding Organisations*. London: Joint Council for Qualifications.
- JCQ. 2022. *GCSE (Full Course). Outcomes for Main Grade Set by Jurisdiction. Results Autumn 2021*. London: Joint Council for Qualifications.
- Johnson, M., and T. Coleman. 2021. *Teachers in the Pandemic: Practices, Equity, and Wellbeing*. Cambridge University Press & Assessment Research Report. <https://www.cambridgeassessment.org.uk/Images/639354-teachers-in-the-pandemic-practices-equity-and-wellbeing.pdf>.
- Jonson, J. L., P. Trantham, and B. J. Usher-Tate. 2019. "An Evaluative Framework for Reviewing Fairness Standards and Practices in Educational Tests." *Educational Measurement: Issues & Practice* 38 (3): 6–19.
- Karami, H., and M. M. C. Mok. 2013. "Editorial: Fairness Issues in Educational Assessment." *Educational Research & Evaluation* 19 (2–3): 101–103.
- Krawczyk, M. W. 2011. "A Model of Procedural and Distributive Fairness." *Theory & Decision* 70 (1): 111–128. doi:10.1007/s11238-009-9189-4.
- Kunnan, A. J. 2004. "Test Fairness." In *European Language Testing in a Global Context: Proceedings of the ALTE Barcelona Conference*; edited by M. Milanovic and C. Weir, 27–48. Cambridge: Cambridge University Press.
- Lamont, T. 2021. "The Student and the Algorithm: How the Exam Results Fiasco Threatened One Pupil's Future." *The Guardian*, February 18.

- Lee, M. W. 2021. *Summer 2021 Student-Level Equalities Analysis*. Coventry: Ofqual.
- Lee, M. W., and P. Newton. 2021. *Systematic Divergence Between Teacher and Test-Based Assessment: Literature Review*. Coventry: Ofqual.
- Lee, M. W., N. Stringer, and N. Zanini. 2020. *Student-Level Equalities Analyses for GCSE and A Level: Summer 2020*. Coventry: Ofqual.
- Lee, M. W., and M. Walter. 2020. *Equality Impact Assessment: Literature Review*. Coventry: Ofqual.
- Leventhal, G. S. 1980. "What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships." In *Social Exchange*, edited by K. J. Gergen, M. S. Greenberg, and R. H. Willis, 27–55. New York: Plenum.
- Lind, E. A., ed. 2019. *Social Psychology and Justice*. London: Routledge.
- Lough, C. 2021a. "2021 GCSE Grades Not 'Fair', Say Teachers." *Times Educational Supplement Magazine*, June 18.
- Lough, C. 2021b. "Parents Push Teachers to Raise GCSE Grades." *Times Educational Supplement*, June 16.
- Mattey, S. 2021. "A Guide to What Happened with Vocational and Technical Qualifications in Summer 2020." *Research Matters: A Cambridge Assessment Publication* (31): 42–50. doi:10.17863/CAM.100409.
- Messick, S. 1989. "Validity." In *Educational Measurement*, edited by R. Linn, 13–103. New York: Macmillan.
- Mislevy, R. J., G. Haertel, B. H. Cheng, L. Ructtinger, A. DeBarger, E. Murray, D. Rose, et al. 2013. "A 'Conditional' Sense of Fairness in Assessment." *Educational Research & Evaluation* 19 (2–3): 121–140.
- Montacute, R., and C. Cullinane. 2021. *Learning in Lockdown*. London: The Sutton Trust.
- Montacute, R., and E. Holt-White. 2020. *Covid-19 and Social Mobility Impact Brief #2: University Access and Student Finance*. London: The Sutton Trust.
- Moore, M. S. 1997. *Placing Blame: A Theory of the Criminal Law*. Oxford: Oxford University Press.
- Murphy, R., and G. Wyness. 2020. "Minority Report: The Impact of Predicted Grades on University Admissions of Disadvantaged Groups." *Education Economics* 28 (4): 333–350.
- Nelson, L., S. Lynch, and C. Sharp. 2021. *Recovery During a Pandemic: The Ongoing Impacts of COVID-19 on Schools Serving Deprived Communities*. Slough: National Foundation for Educational Research.
- Newton, P. 1997. "Examining Standards Over Time." *Research Papers in Education* 12 (3): 227–247.
- Newton, P. 2007. "Contextualising the Comparability of Examination Standards." In *Techniques for Monitoring the Comparability of Examination Standards*, edited by P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms, 9–2. London: Qualifications and Curriculum Authority.
- Nisbet, I. 2020. *What is Fair Assessment? Was It Fair This Summer? What About the Future?*. Chartered Institute of Educational Assessors. <https://www.herts.ac.uk/ciea/articles/fairness>.
- Nisbet, I., and S. Shaw. 2022. "Fair High-Stakes Assessment in the Long Shadow of Covid-19." *Assessment in Education Principles, Policy & Practice* 29 (5): 518–533.
- Nisbet, I., and S. D. Shaw. 2019. "Fair Assessment Viewed Through the Lenses of Measurement Theory." *Assessment in Education Principles, Policy & Practice* 26 (5): 612–629.
- Nisbet, I., and S. D. Shaw. 2020. *Is Assessment Fair?*. London: Sage.
- Noon, M. 2018. "Pointless Diversity Training: Unconscious Bias, New Racism and Agency." *Work, Employment and Society* 32 (1): 198–209.
- Nye, P., and D. Thomson. 2020. "A-Level Results 2020: The Main Trends in Grades and Entries." *FFT Education Datalab Blog*, August 13. <https://ffteducationdatalab.org.uk/2020/08/a-level-results-2020-the-main-trends-in-grades-and-entries/>.
- Ofqual. 2020a. *Awarding GCSE, AS, A Level, Advanced Extension Awards and Extended Project Qualifications in summer 2020: Interim Report*. Coventry: Ofqual.

- Ofqual. 2020b. *Consultation Decisions. Exceptional Arrangements for Assessment and Grading in 2020: Consultation on the Assessment and Grading of Vocational, Technical and Other General Qualifications*. Coventry: Ofqual.
- Ofqual. 2020c. *Consultation on Proposed Changes to the Assessment of GCSEs, AS and A Levels in 2021*. Coventry: Ofqual.
- Ofqual. 2020d. *Exceptional Arrangements for Exam Grading and Assessment in 2020: Consultation on Specified General Qualifications – GCSEs, AS, A Levels, Extended Project Qualifications and the Advanced Extension Award in Maths*. Coventry: Ofqual.
- Ofqual. 2020e. *Malpractice in GCSE, AS and A Level: Summer 2020 Exam Series*. Coventry: Ofqual. <https://www.gov.uk/government/statistics/malpractice-in-gcse-as-and-a-level-summer-2020-exam-series>.
- Ofqual. 2020f. *Ofqual Summer Symposium 2020*. Coventry: Ofqual. <https://www.gov.uk/government/news/ofqual-summer-symposium-2020>.
- Ofqual. 2020g. *Summer 2020 Grades for GCSE, AS and A Level, Extended Project Qualification and Advanced Extension Award in Maths: Guidance for Teachers, Students, Parents and Carers*. Coventry: Ofqual.
- Ofqual. 2020h. *Summer 2020 Grades for GCSE, AS and A Level, Extended Project Qualification and Advanced Extension Award in Maths: Information for Heads of Centre, Heads of Department/ Subject Leads and Teachers on the Submission of Centre Assessment Grades. Updated 22 May 2020*. Coventry: Ofqual.
- Ofqual. 2020i. *Summer 2020 Results Analysis – GCSE, AS and A Level: Update to the Interim Report*. Coventry: Ofqual.
- Ofqual. 2021a. *Analysis of Consultation Responses: The General Qualifications Alternative Awarding Framework. Consultation on Conditions and Requirements for GCSE, GCE, AEA and Project Qualifications in Summer 2021*. Coventry: Ofqual.
- Ofqual. 2021b. *Appeals for GCSE, AS, A Level and Project Qualifications: 2019/20 Academic Year*. Coventry: Ofqual. [https://www.gov.uk/government/statistics/appeals-for-gcse-as-a-level-and-project-qualifications-201920-academic-year](https://www.gov.uk/government/statistics/appeals-for-gcse-as-a-level-and-project-2019-to-2020-academic-year/appeals-for-gcse-as-a-level-and-project-qualifications-201920-academic-year).
- Ofqual. 2021c. *Awarding Qualifications in Summer 2021. Updated 20 April 2021*. Coventry: Ofqual. <https://www.gov.uk/government/publications/awarding-qualifications-in-summer-2021/awarding-qualifications-in-summer-2021>.
- Ofqual. 2021d. *Consultation Decisions. Alternative Arrangements for the Award of VTQs and Other General Qualifications in 2021: Approach to Awarding Qualifications Where Exams Do Not Take Place*. Coventry: Ofqual.
- Ofqual. 2021e. *Consultation Decisions: Decisions on How GCSE, AS and A Level Grades Will Be Determined in Summer 2021*. Coventry: Ofqual.
- Ofqual. 2021f. *Consultation Decisions: Technical Consultation – the General Qualifications Alternative Awarding Framework. Conditions and Requirements for GCSE, GCE, AEA and Project Qualifications in Summer 2021*. Coventry: Ofqual.
- Ofqual. 2021g. *Equalities Report 2021: January to December 2020*. Coventry: Ofqual.
- Ofqual. 2021h. *GCSE, AS and A Level Summer Report 2021*. Coventry: Ofqual. <https://www.gov.uk/government/publications/gcse-as-and-a-level-summer-report-2021/gcse-as-and-a-level-summer-report-2021>.
- Ofqual. 2021i. *Information for Centres About Making Objective Judgements: In Relation to Awarding Qualifications in 2021*. Coventry: Ofqual.
- Ofqual. 2021j. *Information for Heads of Centre, Heads of Department and Teachers on the Submission of Teacher Assessed Grades: Summer 2021. Updated 8 June 2021*. Coventry: Ofqual. <https://www.gov.uk/government/publications/submission-of-teacher-assessed-grades-summer-2021-info-for-teachers/information-for-heads-of-centre-heads-of-department-and-teachers-on-the-submission-of-teacher-assessed-grades-summer-2021-html>.
- Ofqual. 2021k. *Summer 2021 Results Analysis and Quality Assurance: GCSE and A Level*. Coventry: Ofqual.

- Ofqual. 2022. *Appeals for GCSE, AS and A Level: Summer 2021*. Coventry: Ofqual. <https://www.gov.uk/government/statistics/appeals-for-gcse-as-and-a-level-summer-2021/appeals-for-gcse-as-and-a-level-summer-2021>.
- Papageorgiou, K. A., M. Likhanov, G. Costantini, E. Tsigeman, M. Zaleshin, A. Budakova, and Y. Kovas. 2020. "Personality, Behavioral Strengths and Difficulties and Performance of Adolescents with High Achievements in Science, Literature, Art and Sports." *Personality & Individual Differences* 160:109917. doi:10.1016/j.paid.2020.109917.
- Rasinski, K. A. 1987. "What's fair is fair—Or is it? Value Differences Underlying Public Views About Social Justice." *Journal of Personality & Social Psychology* 53 (1): 201–211.
- Rawls, J. 1958. "Justice as Fairness." *The Philosophical Review* 67 (2): 164–194.
- Redmond, B., J. Golding, and G. Grima. 2021. "What is Lost When Exams are Cancelled?" Annual Conference of the Association for Educational Assessment in Europe; November 2021, Online.
- Sadler, D. R. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18 (2): 119–144. doi:10.1007/BF00117714.
- Shavit, Y., and H. Blossfeld, eds. 1993. *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview Press.
- Shaw, S. D., and V. Crisp. 2012. "An Approach to Validation: Developing and Applying an Approach for the Validation of General Qualifications." *Research Matters: A Cambridge Assessment Publication* (Special Issue 3). doi:10.17863/CAM.100449.
- Shaw, S. D., and I. Nisbet. 2021. "Attitudes to Fair Assessment in the Light of COVID-19." *Research Matters: A Cambridge Assessment Publication* 31:6–21. doi:10.17863/CAM.100410.
- Sherwood, D., and R. Cuthbert. 2020. "Halfon Right: Ofqual Has More to Do." *Higher Education Policy Institute Blog*. July 14. <https://www.hepi.ac.uk/2020/07/14/halfon-is-right-ofqual-has-more-to-do/>.
- Stannard, K. 2021. "Will There Be a Mexican Standoff Over Grades in August?" *Times Educational Supplement*, June 3.
- Stratton, T., N. Zanini, and P. Noden. 2021. *An Evaluation of Centre Assessment Grades from Summer 2020*. Coventry: Ofqual.
- The Sutton Trust. 2021. *Remote Learning: The Digital Divide*. London: The Sutton Trust.
- Taylor, R. 2021. *Is the Algorithm Working for Us? Algorithms, Qualifications and Fairness: Reflections on the Qualifications and the Pandemic from the Former Chair of Ofqual, Roger Taylor*. London: Centre for Progressive Policy.
- Thorn, W., and S. Vincent-Lancrin. 2021. *Schooling During a Pandemic: The Experience and Outcomes of Schoolchildren During the First Round of COVID-19 Lockdowns*. Paris: OECD Publishing.
- Vitello, S., and T. Leech. 2022. *What Do We Know About the Evidence Sources Teachers Used to Determine 2021 Teacher Assessed Grades?*. Cambridge: Cambridge University Press & Assessment.
- Witherspoon, S. 2021. "Next Year's Exams Could Be Statistically Contentious if We Don't Act Now." *Schools Week*, August 11.
- Witt, K., J. Whitton, and W. Rifkin. 2018. "Is the Gas Industry a Good Neighbour? A Comparison of UK and Australia Experiences in Terms of Procedural Fairness and Distributive Justice." *The Extractive Industries and Society* 5 (4): 547–556. doi:10.1016/j.exis.2018.09.010.
- Wyness, G. 2017. *Rules of the Game: Disadvantaged Students and the University Admissions Process*. London: The Sutton Trust.