4

A Qualified Disaster: Allocating Student Grades During Covid-19

Hester Burn

Many students in England sit a set of national examinations at age sixteen and eighteen, called General Certificates of Secondary Education (GCSEs) and Advanced Level qualifications (A Levels) respectively. These examinations occur over a period of about four weeks toward the end of the academic year, and opportunities to re-sit after this time are rare. For each qualification, students usually sit the same examination paper at the same time across the country. The examination scripts are then marked externally, and a national regulatory body sets grade boundaries to ensure that the distribution of grades is comparable across years. At GCSE, grades range from one (low) to nine (high), with grades four or five being considered a pass and seven an indication of higher-level proficiency. At A Level, the grades range from "E" (low) to "A" (high). The average student will take eight or nine GCSE and three A Level, or equivalent, qualifications.*[1] *The grades that students receive have significant long-term consequences. Many schools and colleges admit students for A Levels and for skilled vocational training based on their GCSE grades, making this benchmark highly determinant of future educational prospects.*[2] *GCSE and predicted A Level (or equivalent) grades are also the primary means by which universities offer places to prospective undergraduates, and are often requested in job applications long after students have left school.*[3] *The grades achieved by a school cohort also determine that school's place in national league tables and their likely enrollment—and therefore funding—for the subsequent year.*

--

GCSE and A Level examinations were canceled in both 2020 and 2021 due to the Covid-19 pandemic. Policymakers were tasked with deciding how to assign grades in their place. In 2020, the Department for Education opted to use teacher predictions alongside a statistical moderation process to ensure that the grades that students received—both overall and within each school—were not considerably higher than those in previous years. However, this approach led to a public outcry of such magnitude that the policy was withdrawn the day after students received their grades, with the Prime Minister blaming a "mutant algorithm" for the U-turn and publicly dismissing a top civil servant.[4] In 2021, schools were instead asked to run their own examinations and then submit grades with evidence to back up each of their decisions. However, this approach did no better at commanding public trust, with some parents preparing formal appeals against their children's schools even before the 2021 grades had been released.[5] The challenge facing policymakers was that Covid-19 highlighted the fragility of the existing examinations, bringing to light both new and old issues embedded in long-established processes. If 2020 and 2021 grading needed to compensate for problems exacerbated by the pandemic then which, if any, should take precedence?

Fairness Between Cohorts

For some, the most important consideration for grading processes in 2020 and 2021 was consistency. For a decade, the grade boundaries for these national examinations had been specifically set so that the proportion of students achieving each grade was similar to that in previous years.[6] This was primarily intended to allow future employers and advanced education providers, such as universities, to fairly compare applicants from different cohorts who apply for the same opportunity.[7] When examinations were canceled in 2020 and 2021, and teachers were tasked with assigning grades to their own students, it was considered likely that the grades would be higher than those in previous years. This was, in the words of one school leader, because "schools may, understandably, [give] some students the benefit of the doubt when they are on the borderline" between two grades.[8] However higher grades for the 2020 and 2021 cohorts would make life harder for many students in non-disrupted years who would end up competing with them for university places and entry-level jobs, especially given that a considerable number of the 2020 and 2021 cohorts were likely to defer

applications until after the height of the pandemic had passed.[9] In addition, at least some students from these more and less generously graded cohorts were likely to apply for the same employment opportunities in the future, for which GCSE, A Level, and equivalent grades are often still requested. As another school leader commented, "when we're thinking about fairness and equality, of course, first and foremost, we think about this year's cohort. But we have to think about fairness and equality with previous cohorts and with future cohorts as well."[10]

In 2020, the Department for Education addressed concerns about consistency and fairness by deciding that students' examination grades would not be determined by teacher assessment alone. Instead, the Office of Qualifications and Examination Regulation (Ofqual) would determine the final grades received by students by applying an algorithm to those given by teachers. This would ensure that the proportions of each grade received at both the school level and the national level were consistent with previous years, informed by data on the historical results of each school and a ranking of students within them.[11] It also guaranteed that no school was raising their grades significantly more than another. The policy was welcomed by Universities UK, representing the UK higher education sector, who believed that the qualifications would therefore "hold their value" and that the processes would enable progression to further education or jobs without being at "the expense of academic standards."[12]

When the grades were released in August 2020, however, the public response was overwhelmingly negative. Many students felt as though they'd had little agency in determining the grades that they received. One reason for this was that much of the existing data that schools had on student attainment were through "mock" (practice) examinations collected months before schools had closed. Did these data not provide an outdated representation of students' full potential? As one young person noted, "when we took [the mock examinations], we were under the impression that they weren't really important, and now they are."[13]

A second reason was that, as the grades that students received were informed by the prior performance of their school, some students in previously low-attaining schools were unable to be assigned top grades even if they would have performed considerably better than their older peers, resulting in many missing the grades that they needed to meet their university offers.[14] Indeed students from disadvantaged backgrounds were found to have been more likely to be downgraded at A Level, impacting—

in the words of the president of the National Union of Students—"who's had to pick a different university, who's been forced to defer … [and] on the demographics of our universities" in a manner which was "completely unjust."[15] One student who was initially downgraded and therefore missed a conditional offer at her chosen university stated: "I just think it is very unfair that because of past performance in this area and particularly in my school that I haven't got in … that computer doesn't know that I got to school every single day at 7am revising … that computer doesn't even know who I am and that I exist."[16]

Was it defensible that these students' futures had been upended due to the decisions of a statistical model? Such outcomes were near-impossible for the government to justify.

In the end, the fiasco was instead blamed on a "mutant algorithm" and the higher, purely teacher-assessed grades were restored within a matter of days.[17]

Fairness Between Schools

By 2021, although students had more or less returned to classrooms, new issues regarding the fairness of national examinations had risen to the surface. During school closures, the education that schools were able to provide at a distance had varied greatly, with some able to rely on their students each having a personal laptop while others had relied on packets of printed resources.[18] Moreover, many of the poorest areas of the country were the most heavily impacted by school absences and partial closures after the height of the pandemic, meaning that, even once schools reopened, their students missed out on a greater amount of face-to-face learning than those in more affluent regions.[19] The injustice that these circumstances unveiled was not new, but it was more blatant. The potential negative consequences for students in more economically marginalized schools if examinations went ahead "as normal" were palpable.

For policymakers deciding how grades should be assigned in 2021, one solution was to still allow students to sit examinations but to put contingencies in place which accounted for differences in learning opportunity during periods of school closure. Teachers were therefore instructed to disregard the full extent of the curriculum and to instead create examination papers which were customized to test their students only on topics about which

they had actually been taught. Schools would then combine these results with other data at their disposal and use this evidence—which could be checked by an external adjudicator—to assign each student a grade. OCR, an exam-setting organization, stated that such a "strongly evidence-based approach" would "promote public confidence" in the examination results received.[20]

This compromise, however, was not wholly watertight. The flexibility granted to schools was incomparable to the strict processes of the pre-pandemic years, including allowing schools to distribute lists of the topics to be examined in advance and to hold the examinations in low-intensity classroom settings. Many felt that such flexibility was long overdue. However, when the proportion of top A Level grades in one fee-paying school soared from 34 percent in 2019 to 90 percent in 2021, and when this was found to be exemplary of a wider trend by which grade increases were a lot higher in fee-paying schools than those which were government-funded, concerns grew.[21] Was it fair to merely replace variation in how schools were impacted by Covid-19 with variation in how schools designed and marked their own assessments? Could it really be claimed that these results were as robust as those in non-disrupted years? And, even if this was possible in theory, did the importance of students' results to school league tables in England introduce incentives against which it was impossible to mitigate? Although, on release, the 2021 grades were not subjected to the overhaul experienced the previous year, students lodged four times as many appeals against their schools' decisions.[22] Students of teachers who had put faith into the system acting fairly were likely the ones who faced the greatest repercussions when it became apparent that it had not.

Fairness Between Students

Finally, despite the 2020 and 2021 grading policies attempting to achieve a semblance of fairness *between* schools, it was impossible to ignore the fact that schools would be facing dilemmas about fairness *within* their own student bodies. How should the examination performance of a student who, during school closures, had had consistent internet access and a quiet room in which to work be compared to that of a student who shared a single device with their siblings? How should teachers judge the attainment of a student who had to take on extra paid work during the pandemic because

others in their household were shielding? And what of students with sensory impairment, for whom there would have been additional learning loss through remote instruction? Research by the education think tank The Sutton Trust revealed that 40 percent of middle-class children had undertaken over five hours of schoolwork a day during school closures, compared to just 26 percent of those in working-class households.[23] Although some of this gap will have been driven by differences between the schools that these students attended, significant gaps were also likely to exist between middle- and working-class students within the same school, and indeed within the same classroom.

The difficulty for policymakers in considering grading fairness between individual students, however, was that it invited questions about examination grades that extended far beyond Covid-19. The attainment gap between students who are eligible for free lunches and their more advantaged peers had been shown to be growing rather than shrinking *prior* to Covid-19.[24] This situation had only worsened since.[25]

Policymakers may have considered trying to account for such systematic disadvantage by allowing teachers to raise grades beyond that for which they had evidence, based on extenuating circumstances. But how would these circumstances have been categorized, and what precedent would it have set for future years? Moreover, if sorting students into grades is merely a mechanism by which existing disadvantages in society are quantified and reproduced, then should it even be done at all? Students with graduate parents were found to be unfairly advantaged by teacher-assigned grades in 2020, illustrating how less standardized assessment systems may be more easily gamed by those with greater resources.[26] Nonetheless more standardized assessments, even if they are less easy to manipulate, are arguably still merely measuring the years of advantage that those with greater resources already have.[27]

Looking Back and Looking Ahead

Deciding how to assign grades during Covid-19 in England required those with decision-making power to engage with significant practical and ethical dilemmas. One of the biggest of these was whether it was more important to prioritize fairness between cohorts, fairness between schools, or fairness between groups of pupils who experienced hardships to different extents during the pandemic. However, a second question about England's national

examinations also rose to the surface of public debate during this time: Were they ever fair in the first place? Should exams be canceled again, how should the Department for Education in England respond? Do the issues with high-intensity, high-stakes examinations in England that were exposed by Covid-19 mean that changes need to be made whether another crisis occurs or not? And finally, if so, what might an ethical assessment and accountability system in England look like in the future?

Conversation

Hester Burn: Welcome to our case study discussion on grading in a time of Covid. It's actually A Level results day in England today [August 18, 2022], which means we're about a full year after the time period discussed in the case study. But Covid still has a strong legacy, even this year, with students still concerned about the results, and policymakers having to grapple yet again with trade-offs about what will be included in examination papers, and how to set grade boundaries. This is very much going to be on our minds today. So thank you all so much for being here.

I was previously a maths teacher and I recall feeling very emotionally turbulent about assigning grades to my own students in 2020 and 2021. So I was interested to know: did reading the case study and returning to that period bring back any uncomfortable emotional memories for you?

Josiah Isles: Oh, yeah, lots. I moved schools in the interim between 2020 and 2021, and the approaches from school to school, top down from leadership, were so different. One of the biggest problems in the first year was trying to offer some parity to previous cohorts.

Hester Burn: Diana, do you want to talk a little bit about the dilemmas that universities were facing?

Diana Beech: Absolutely. When you asked the first question, I felt, first of all, relief that I wasn't in the front and center of things as a teacher. At the time, I was firmly part of the university sector, working for a large, research-intensive, high tariff [academically selective, elite] UK institution. And when the A Level results came out, there were more students who were predicted to have higher grades, because of the computer-generated results. But it wasn't a boon for universities, as the press were describing. Actually, there was genuine concern on the ground. How do we take all these students? How do we provide them a top-rate student experience? Have we got the accommodation? Have we got the staff? It wasn't

just, "we've got more students, we've got more money," as I think some of the press were portraying it. Universities were really concerned about maintaining that high-quality student experience that they're known for, and to do justice to that cohort of students, which of course needed extra support, having come through the pandemic and adjusted to online learning as well. Can we make hybrid work? Can we cater for all students? And of course with international students, and travel restrictions that still exist to this day—how can we make sure we're providing an inclusive education experience, so they can join in and participate fully as well? Those were the questions I remember facing.

Hester Burn: Was there much of a concern about whether admissions procedures were going to be as accurate as they normally are?

Diana Beech: I think that was secondary. From the universities' perspective on the ground that year, they wanted to do what's right for those students who'd just overcome absolute turmoil. They just wanted to make sure that they had something secure for the future and that those who deserved it and wanted to go to university could. So I think the first year was just a mad scramble to do what's right.

Tom Richmond: It was always going to be very challenging for teachers. As a former A Level teacher, myself, I'm well aware of the pressures that teachers are under. But of course, the lockdown came in March 2020, when students were at the very end of their courses. They actually only missed probably about six to eight weeks of teaching at the end of a two-year course. So when the government said to teachers, can you tell us what you think those students would have got in the exam, which was basically the official instruction they were given, then that was a fairly reasonable request, given that teachers had access to two years' worth of assessment data. But of course, the following year, that changed very dramatically. The question that teachers got asked in 2021 was very different, which was, can you tell us what level these students are performing at? And the fact that these were two different questions escaped public attention and journalists' attention inevitably, because they were buried in guidance from the exam regulator to schools and colleges. At least in the first year, although there clearly was a political meltdown as much as a grading meltdown, teachers had a pretty strong basis to make decisions. But when it came to 2021, we'd had two academic years badly disrupted, and no exams, and students were told a long time in advance that they wouldn't have exams, and every school in the country just had to figure out its own answer. Universities were saying, hold on a second. How do we know which of these students is actually performing at a higher level compared to other students, when every school has come up with its own answer to "Should we use homework to judge students?," "Should we use mock exams?," "How many mock exams should we use?," "Should we include

long term coursework assignments?" Everyone was coming up with different answers, and that just put teachers in a really difficult position, because they were, ultimately, the arbiters of their students' future. And I think one of the things we forget is that the anonymity of marking that we have with the public exam system is an incredibly important feature that we lost. We just said to teachers, you have to take on all the responsibility now. And I think that was incredibly unfair on teachers, and indeed, the school leaders and college leaders who then had to sign off on those grades as well.

Josiah Isles: Picking up on your points, Tom, about the process with regards to the fairness to school leaders and teachers, I would say that it brings into context the actual material that was used to assign the grades as well. When you go for an exam process, that material is so vigorous in terms of the content. I'm not convinced that in all subjects, across all our schools, across the country, that all teachers adhered to that level of detail. Because if you've got, for example, a really new and inexperienced faculty who has not been an examiner themselves, to then ask them to write enough material and to be quite vigorous with those grades is a difficult job. And that puts more emphasis on teacher judgment. And it's really quite evident that in some schools they've used materials supplied by exam boards and in some schools they have made their own, and it caused quite a lot of issues with regards to consistency across schools. It was quite a major issue with regards to whether students have been given materials in advance or given the list of materials to revise [review]. Even that small aspect still is different in different schools. So it was really quite a difficult position with not a lot of guidance about how to keep that additional scrutiny across schools really tight and consistent. Where's the parity between my school in Bolton compared to a school in Surrey, for example: we haven't discussed anything, there's no communication across maths departments, across English departments, between school leaders. So where's the parity coming from? Well, we're relying purely on that consistency or teacher approach. But that can't happen unless you've communicated in some way.

Hester Burn: What are the dilemmas that emerged about the UK education assessment system in general during this time?

Dylan Wiliam: Education researchers, particularly those focusing on assessment have moved away from the idea of validity as being a property of an assessment towards the idea that validity is a property of the conclusions that you draw. So there's no such thing as a valid test—a test is valid for some purposes and not others. And so validation needs to focus on: what are we entitled to conclude about a student from seeing how well they do on an assessment? I think if people had actually focused on that, we'd have had a more sensible debate. With A Levels in particular, there's an unspoken issue about whether they

are meant to be measuring achievement or aptitude. If you want to apply to a British university, you have to apply to a particular department—that's not the case in the United States, where you typically select your major in your second or third year. In the UK, if you want to apply to do mathematics, you apply to the mathematics department, and they will want to see a high standard of mathematics achievement. But if you want to study law, they don't expect you to have studied law in school. So there's actually a massive variability in the kinds of requirements the universities have for specific achievement in subjects, even subjects that will be the focus of your studies. Often the focus is more on some measure of aptitude: is this person somebody who can learn?

The other thing that I think that is really quite important to get right, is that we've actually had a post-qualification system for many years in the UK and at many universities, because we tend to rely on predicted grades. When I was chair of the admissions committee at King's College London, we had a real problem that sometimes when we gave students offers, if they didn't eventually make the grades, we ended up rejecting them. And then we found we were short of students. So we went back to fill the places and we often ended up actually accepting students who were at a lower level than those who we'd rejected. So we actually started saying, "Look, if you've given them an offer, take them whether they make the grades or not," because the examination grade isn't as reliable as the kinds of processes that led to the offer in the first place. So I think there's several issues that need to be unpacked here, but the real problem comes down to this aptitude versus achievement issue.

I think that in Britain, we have a tendency at the universities to wait for talent to knock on the door, whereas in the United States, there's much more of a tradition of universities seeing it as part of their job to identify talent. If you score 1600 on the SAT you may actually get a contact from a top school before you've actually even got the results from the College Board. Here's the challenge for me. A Level is called the gold standard. But for me, the question should be: are A Level grades the only way of identifying talent? I think we should be highlighting that the A Levels are not perfect. The differences between subjects are really strange. We now know that chemistry, science and physics and mathematics A Levels are much harder than other subjects. What sense does it make to compare an economics A Level which is based on just two years of study and an English A Level where kids have been learning this subject for eighteen years? The other thing people don't want to talk about is the fact that some students don't get the same opportunities to learn as others. If you want to assess students then, fine, assess their opportunity to learn. But I don't want to certify doctors and airline pilots taking into account their opportunity to learn; I want them to be good at the things that we want them to be able to do.

All these tensions basically come down to defining the constructs more carefully: what it is we want students to be able to do. When people designing

curriculum fail to specify the constructs, the people designing the assessments have too much power. So in terms of the ethics, with clear construct definition, assessment development should be a largely technical exercise, in that different people should be able to agree about whether a particular assessment is adequate or not. When the construct is not defined, reasonable people can disagree about the adequacy of the assessment. They think it's an argument about the assessment when it's actually an argument about the underlying construct. These debates are really about the fact that we don't actually even agree about what it is that A Levels should be assessing.

Hester Burn: We've got differences in the values, aims, and beliefs about what assessment of any kind is there to do. Does anyone want to talk more about the differences in values and aims for assessment in this country?

Dylan Wiliam: The case explains that we use assessments to certify the achievements and maybe the potential of individuals, but we also use them to hold schools to account. What's bizarre about this is that data from the Department of Education, and the Organization for Economic Cooperation Development, show that schools account for about 8 percent of the variation in student achievement. Ninety-two percent of the variation in student achievement is not caused by the quality of the school. So we use these results in ways that are completely illegitimate, in that the primary determinant of our kids' grades when they leave school is how much they knew when they started at that school. And that, I think, is the real problem. Average school grades are a really bad indication of how good that school is.

Tom Richmond: When it comes to the aim of assessment, the more aims you attach to a single assessment, the less it does any of them well. If you look at the assessments you have at age eleven, for primary school pupils, they are used for an eye-watering number of different purposes. And that places an enormous strain on them. This is sometimes used as a trigger for talking about university entrance exams in the UK and changing to a system, like they have in the United States, where they separate out the two roles: you'd have a test to get you into University, and a separate test to determine how students are performing in their school.

When you look at all the different actors who were in the system in 2020 and 2021, we had the government thinking that it was their job to hold schools and colleges accountable, and the exam regulator feeling the weight of their legal duty to uphold public confidence in the exam system, and also to make sure that standards are consistent over time. Then you had parents and students who were looking to get that highest possible grade, knowing that it might unlock doors at university or for employment. You have teachers and schools and colleges trying to make sure they were behaving with integrity and fairness in an extremely

uncertain and unstable environment. And ultimately, with the greatest respect to everyone who was involved, you just can't achieve all of those things. I don't think politicians were honest enough with the trade-offs that were going to have to be made. They tried to go down one route in 2020, only to have to slam on the brakes in the face of enormous public opposition, because they weren't sure they understood the system very well themselves. That had a huge impact on 2021. That's why I think the two years ended up looking so different and posing such different ethical and moral challenges for teachers, and technical and technocratic challenges for the regulators.

Dylan Wiliam: It might be useful to compare the UK system to other systems like the ones in Sweden and Texas. In those systems, when you apply to university, you can actually take an aptitude test, the SAT in Texas or the SweSAT in Sweden. But you also apply on the basis of your grade point average, the grades your teachers give you. I think in Texas, if you're in the top 10 percent of your graduating high school class for GPA, you automatically gain admission to the state University of Texas system. In Sweden, you are automatically considered for a university place on whatever type of assessment gives you the best chance of getting in. The idea is to allow students to show us what they can do in a range of different ways. So maybe in the UK we should have multiple routes. And I think that would have been something that we could have done quite intelligently in 2020 and 2021, and let people let different users weight these components differently.

Josiah Isles: In the last couple years we've had some students who have had the opportunity to really show what they can do in the classroom, rather than in an exam, which previously they wouldn't have been able to. I'm a science teacher, and my students often find the chemistry, physics, and biology individual components of the A Levels really, really difficult. They are solid subjects. You often have students who are really good at science, but can't access the exam papers in a conventional manner. And I think over the last couple of years, these students have been able to realize that they're good at science in an unconventional way. The problem is that, in the UK system, the BTECS [Business and Technology Education Council vocational qualifications] that are offered for science are actually A Level in terms of content, just crammed into a format that is meant to be more accessible. And when it isn't, it's just as difficult. It would be helpful to get the opportunity to stop and think about how we can get some of these students through without writing them off and saying, "Because you can't access that A Level paper, science is not right for you." At the moment, either you can access the exam papers and do really well or you can't, and we let those students down. We allow them to do apprenticeships, which is wonderful, but there are quite a lot of opportunities that we could allow access to if there were alternative pathways.

Diana Beech: I wonder if we shot ourselves in the foot a little bit by the way we talk about A Levels as the golden ticket to higher education. Because, to pick up on something that Professor Wiliam said earlier, university admissions professionals in the UK do actually consider A Level results and BTEC results alongside a bigger basket of measures as well, to measure things that are not necessarily captured in testing achievement. And this does include the aptitude of our students, in the learning gains that they've taken to get their three A Stars, which could be dependent on their own personal circumstances: other commitments they're carrying, their school background, and other factors.

Hester Burn: Let's rewind, back to 2020, or 2021. Does anyone have any firm idea about which purpose of the A Levels should have been prioritized?

Dylan Wiliam: I would have actually administered traditional A Level exams and then put an asterisk next to the scores. So you keep the standard the same, and if the students fall short because of the disruption, you can take that into account. I would have made sure that universities understood this, and tried to give them some indication of the disruption. There's quite a lot of work about how much learning was lost as a result of the pandemic, including different estimates for different kids from different social groups. So you could actually tag it and say, if you're a middle-class kid, you probably lost two months of learning, if you're a kid from a much less affluent background, you probably lost four months. Then I think the admissions tutors could have taken that into account in deciding how to admit people. But once you change the standard, then it's really hard to figure out what these results mean. So I think even though it would be politically difficult, I wouldn't change the A Level standard, I would qualify and provide more support and contextualize it. You know, I wouldn't let people pass a driving test because they had a hard time learning to drive.

Diana Beech: That was going to be my initial thinking as well. However, I've got one niggling thought in my head and that is: what happens ten years down the line, when those students have graduated and are competing for jobs with people from two years previously? Are employers going to take that asterisk into consideration if they are looking at A Level grades as a differentiating factor?

Hester Burn: Can you talk a bit more about that? Where's the inequality or social justice issue that you might be seeing?

Diana Beech: Well, let's assume all these graduates come out with nice 2:1s and firsts. And let's take a five-year cohort. Some of the big employers do still look back to A Level grades as a differentiating factor for who they hire. And I just wonder, will it be in employers' consciousness that some of that cohort took their A Levels in the pandemic year? Do the students even want to be known as the pandemic generation? You would think there might be an issue of personal

pride at play as well. People want to be in the position on their own merit. I don't think these students will want to feel that they're sort of dragged down by that asterisk, although, as Professor Wiliam said, it is probably the fairest thing to do in the short term.

Hester Burn: Do you agree that it should have been the priority to signal achievement, comparable to other years, even if it had to come with an asterisk?

Josiah Isles: Absolutely. I think that the students should have had an opportunity to actually just sit the exams. I don't see why that wasn't something that wasn't considered. It could have been rolled out quite easily and would have been fairer. Even if it was online, that would have allowed for a comparison to previous years and to their compatriots around the country. It would have been a wonderful thing to have done.

Tom Richmond: I'd very much echo that. I mean, the political considerations became overwhelmingly, and I would say disproportionately, important in the summer of 2020. But when it came to 2021, there was no reason why GCSE students couldn't at least have sat their English and maths GCSEs [taken at age 16]. Because when you're talking about the purpose of the assessments, there was an argument for saying, we need to get some core skills across these students and really test them because they're going to be going on to further study or employment or an apprenticeship. And that wouldn't have been impossible at all! Schools and colleges were open all around the country, and we could have offered students the opportunity to maybe sit two or three other subjects as an option rather than moving to teacher assessment. And for A Level students, maybe we could have let them sit one of their subjects out of their three or four, and had the teacher assess the grade for the other two. That would have given some grounding to the system, and maybe would have approached a system that people could trust and have confidence in. I think the biggest concern about what happened in 2021 is that nobody had confidence in the grades by the end of the process. And although as I was saying earlier, having an assessment with too many purposes attached to it is a massive problem, having an assessment that serves no purpose at all is probably even worse. If you'd go back and look at what was happening in the run up to the use of that so-called "mutant" algorithm in 2020, there were very few concerns being aired before exam results day because everyone knew it was a half decent solution to an unexpected and difficult situation. It was only when politics got involved that we moved away from that in a very dramatic snap and all of a sudden, we just couldn't really get back to any sensible conversations anymore. The results that have come out today are showing that we are in the process of returning back to normal, whatever normal now means. And by 2023, everything really should be back to pretty close as it was in 2019. Whether that means we've missed the opportunity

for some quite important conversations is at the front of a lot of people's minds at the moment, because we have learned a lot about what the options are and what the benefits and drawbacks are of different models over the past couple of years.

Hester Burn: The one niggle that I've still got involves what influence schools have in this process. Dylan said at the beginning that they account for 8 percent variation in student outcomes, which is mostly determined already by where pupils are when they begin school. And yet there was quite big variation during Covid in what schools were able to deliver in terms of home learning. Is that something we should be taking into consideration?

Dylan Wiliam: The trouble for me is that, yes, we should, but allowing teachers to take those things into account produces a huge additional variability, because you have no idea just what to take into account. As a libertarian, my first question is, why is the government involved in examinations at all? Why is the government controlling this? Why isn't it a decision for individual schools and academies? I think that the debate I would want to have is at a system design level. What information do universities want? And the other thing I'd want to put into the equation is: Why is our qualification of kids at eighteen completely driven by universities? Is admission to university the most important reason to describe student achievement at the age of eighteen? I would like to see us deciding what it is we'd like to know about students at the age of eighteen, where the exam boards are just one of the partners involved and employers are others, but then, society as a whole needs to have some stake in how we're certifying student achievement when they leave school.

Josiah Isles: Is that fair? I think that there's quite a disparity coming to the fore with regards to learners who are from disadvantaged backgrounds. The lack of technology, the lack of a private work space, all of these factors have been exacerbated over the last two years. And just to add to that, these households might have experienced Covid deaths, or they might also be in situations where parents have lost work. That gap between parents who have graduate jobs, for example, and those that are coming from disadvantaged backgrounds is now increasingly more apparent. And I think that the "return to normal" doesn't take into account that for the learners who are going to be taking GCSEs, most of their time in secondary school has been disrupted. And their home life has been significantly disrupted as well. So they will be doing "normal" exams, and they will be having "normal" opportunities, but their emotional wellbeing and support networks are even more reduced than they would have been normal in a situation of disadvantage. What should we do for those students? And that's where the teachers come into play, but they can only do so much. Currently we are throwing money at this problem with tutoring programs in the hope

that's going to improve results. But that isn't really going to improve anything if students try to do tutoring at home online, and they've got five brothers and sisters in the background, and mum's trying to find work for the family to get food. That doesn't take into account any of that, that doesn't actually fix the problem that some students are going hungry. We need to re-look quite heavily at that, because the results that come through for those learners when they sit their GCSEs, are going to be quite far from where we want them to be. And when they get through to A Levels the gap is going to be vast. We need to help with that home situation, which is quite a massive problem and needs to be looked at in more detail than just what we're doing in schools and with the exam policy.

Diana Beech *is Chief Executive Officer of London Higher, the representative body for almost fifty universities and higher education colleges across the London region. She was previously Policy Adviser to three former UK universities and science ministers and, before that, the first Director of Policy and Advocacy at the Higher Education Policy Institute.*

Josiah Isles *is the Assistant Headteacher at Ladybridge High School in Bolton, UK. A London born Science Teacher, he has a lifelong passion for helping disadvantaged learners to achieve their full potential.*

Tom Richmond *is the founder and director of the EDSK think tank, which conducts research on schools, colleges, apprenticeships, and universities in England. He is also a former teacher and former advisor to ministers at the Department for Education in London.*

Dylan Wiliam *is Emeritus Professor of Educational Assessment at the UCL Institute of Education. In a varied career, he has taught in urban public schools, directed a large-scale testing program, served a number of roles in university administration, including Dean of a School of Education, and pursued a research program focused on supporting teachers to develop their use of assessment in support of learning.*

Discussion Questions

1 In 2020, policymakers used an algorithm to modify teacher-assigned grades, in an attempt to address concerns about consistency and fairness. What role, if any, do you think algorithms and/or other

technological solutions should play in the allocation of educational goods, like qualifications and university placements?

2 The pandemic highlighted the vastly different resources available to different schools and even to different students within the same schools. How much, if at all, should those disparities influence the ways that universities and employers use examination scores to distribute placements?

3 Ensuring that GSCE and A-level test scores are assessed consistently across grade-level cohorts is important partly because these scores directly inform hiring decisions beyond school. What role, if any, should standardized test scores play in hiring decisions?

4 Policymakers needed to find a way to balance being fair to the cohort impacted by the pandemic with being fair to both previous and future cohorts. How would you have balanced these different cohorts as you consider how to assign grades fairly in 2020 and 2021?

5 What inequities in the pre-pandemic testing system did the pandemic reveal? Would you describe the pre-pandemic testing system as more or less fair than the various pandemic-era approaches to testing? Why?

6 All school-age children were impacted by the pandemic. How, if at all, should policymakers think about accommodations or adjustments in future years for students whose education was also disrupted in the past?

7 How does this case help you think about the role that high-stakes standardized assessments play in your own national setting?

Notes

1. Students may also choose to study one or more Business and Technology Education Council (BTEC) qualifications, which are equivalent to GCSEs or A Levels and have a greater focus on practical, vocational skills.
2. Students who are just a few marks short of a pass in their English GCSE are about 9 percentage points less likely to enroll in A Levels and 4 percentage points more likely to drop out of education entirely by age eighteen than they would be had they just scraped through. In Machin, Stephen, McNally, Sandra, and Ruiz-Valenzuela, Jenifer (2020). "Entry through the

Narrow Door: The Costs of Just Failing High Stakes Exams," *Journal of Public Economics* 190. https://doi.org/10.1016/j.jpubeco.2020.104224

3. For more information. see Taylor, Nick (2022). "The Real Story behind University Entry Requirements," *The Uni Guide*, Accessed June 14, 2022. https://www.theuniguide.co.uk/advice/ucas-application/the-real-story-behind-entry-requirements
4. Coughlan, Sean (2020). "A-levels and GCSEs: Boris Johnson Blames 'Mutant Algorithm' for Exam Fiasco," *British Broadcasting Corporation*. https://www.bbc.co.uk/news/education-53923279
5. Fazackerley, Anna and Savage, Michael (2021) "Parents in UK Prepare for Legal Action over A-level Results," *The Guardian*. https://www.theguardian.com/education/2021/may/30/parents-in-uk-prepare-for-legal-action-over-a-level-results?CMP=Share_AndroidApp_Other
6. This process is called the Comparable Outcomes Framework. It allows some adjustment for differences in the difficulty of examination papers and for the prior attainment of specific cohorts. For more information, see Balaban, Corina, Lloyd, James, and Surridge, Phoebe (2021). "Comparable Outcomes: Setting the standard?," *AQI*, Accessed June 14, 2022. https://www.aqi.org.uk/briefings/comparable-outcomes-setting-the-standard/
7. Students in England are in the same cohort if they begin school in the same academic year, usually when they are age four. Moving between cohorts is very rare, so these students are usually taught together until they leave compulsory full-time education.
8. Richardson, Hannah and Sellgren, Katherine (2020). "Pupils Get GCSE Grades as BTEC Results Are Pulled," *British Broadcasting Corporation*. https://www.bbc.com/news/education-53833723. In addition, Murphy and Wyness (2020) provide some evidence to support this assumption. Analyzing data from years not affected by Covid-19, they find that 75 percent of students are predicted A Level grades—which determine offers of university places in England—which are higher than those they actually receive. Nonetheless it is important to bear in mind that the contexts of these A Level predictions are not the same as teacher assessments in 2020 and 2021. See Murphy, Richard and Wyness, Gill (2020). "Minority Report: The Impact of Predicted Grades on University Admissions of Disadvantaged Groups," *Education Economics* 28:4: 333–50. https://doi.org/10.1080/09645292.2020.1761945
9. See Weale, Sally, Hall, Rachel, and Adams, Richard (2022). "First Post-Covid School Leavers Face Fight for Fewer University Places," *The Guardian*. https://www.theguardian.com/education/2022/jun/11/first-post-covid-school-leavers-face-fight-for-fewer-university-places

10. Lough, Catherine and Gibbons, Amy (2021). "GCSE and A Level Results 2021: What Did Teachers Learn?," *Times Education Supplement Magazine*. London: The Times. https://www.tes.com/magazine/news/secondary/gcse-and-level-results-2021-what-did-teachers-learn
11. The algorithm was based on the Comparable Outcomes Framework as well as the historical results of the school. In practice, teacher assessments were automatically scaled up or down so that they replicated the average proportion of grades that the school's previous three cohorts had achieved.
12. Universities UK (2020). "Universities UK Response to NUS Letter Calling for 'No Detriment' Exam Policy," *Universities UK*. https://www.universitiesuk.ac.uk/topics/covid/universities-uk-response-nus-letter
13. Otte, Jedidajah (2021). "'I'd 100% Prefer to Sit Exams': Pupils on Teacher-Assessed Grades," *The Guardian*. https://www.theguardian.com/education/2021/feb/25/pupils-on-teacher-assessed-grades-a-levels-gcse-england
14. In reality, 40 percent of teacher-assessed grades were initially downgraded by the algorithm in 2020. See Office of Qualifications and Examinations Regulation (2020). "Awarding GCSE, AS, A level, Advanced Extension Awards and Extended Project Qualifications in Summer 2020: Interim Report." Great Britain: Office of Qualifications and Examinations Regulation. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/909368/6656-1_Awarding_GCSE__AS__A_level__advanced_extension_awards_and_extended_project_qualifications_in_summer_2020_-_interim_report.pdf. See also Coughlan, Sean (2020). "A-levels: Anger over 'Unfair' Results This Year," *British Broadcasting Corporation*. https://www.bbc.co.uk/news/education-53759832
15. Packham, Alfie (2020). "NUS President Larissa Kennedy: 'I Worry Universities Won't Put Student and Staff Safety First'," *The Guardian*. https://www.theguardian.com/education/2020/aug/31/nus-president-larissa-kennedy-interview
16. Bremner, Juliet (2020). "'They're Robust, They're Good': Boris Johnson Defends England A-Level Results as Thousands of Students Degraded," *Independent Television*. https://www.itv.com/news/2020-08-13/a-level-results-day-number-of-students-accepted-to-university-rises-amid-exam-results-controversy
17. Coughlan, Sean (2021). "A-levels and GCSEs: Boris Johnson Blames 'Mutant Algorithm' for Exam Fiasco," *British Broadcasting Corporation*. https://www.bbc.co.uk/news/education-53923279
18. Del Bono, Emilia, Fumagalli, Laura, Holford, Angus, and Rabe, Birgitta (2021). "Coping with School Closures: Changes in Home-Schooling

during Covid-19," *Institute for Social and Economic Research*, University of Essex. https://www.iser.essex.ac.uk/files/news/2021/little-inequality-homeschool/coping-with-school-closures.pdf

19. Children's Commissioner (2020). "School Attendance since September," London: Children's Commissioner for England. https://cco-web.azureedge.net/wp-content/uploads/2020/12/cco-briefing-on-school-attendance-since-september.pdf
20. Oxford Cambridge and RSA (2021). "Consultation on How GCSE, AS and A Level Grades Should Be Awarded in Summer 2021: The OCR Response," *OCR*. https://www.ocr.org.uk/Images/609621-summary-of-the-ocr-response-to-dfe-ofqual-consultation-how-gcse-as-and-a-level-grades-should-be-awarded-in-summer-2021.pdf
21. McCall, Alastair, Griffiths, Sian, and Rodriguez, Nick (2022). "Private Schools 'Gamed' Covid Rules to Give Their Pupils More Top A-Levels," *The Times*. https://www.thetimes.co.uk/article/6bf66e7c-85fd-11ec-9c9e-c5f8451b2970?shareToken=7a9ad0e8241d545e98106d6ee8317344
22. Lumby, Tommy (2022). "GCSE and A-Level Grade Appeals Soared in 2021, Ofqual Data Shows," *Wales Online*. https://www.walesonline.co.uk/news/uk-news/gcse-level-grade-appeals-soared-23128647
23. In the UK, "middle-class" can be broadly understood to refer to those who work in managerial or professional occupations and "working-class" to those who work in routine and manual occupations. The statistic is taken from Cullinane, Carl and Rebecca Montacute (2020). "Covid-19 and Social Mobility Impact Brief: School Closures," *The Sutton Trust*. https://www.suttontrust.com/wp-content/uploads/2021/01/School-Shutdown-Covid-19.pdf
24. Hutchinson, Jo, Mary Reader, and Avinash Akhal (2020). *Education in England: Annual Report 2020*. London: Education Policy Institute. https://epi.org.uk/publications-and-research/education-in-england-annual-report-2020/
25. The attainment gap between more and less privileged students grew further during the first year of the pandemic such that it reversed almost ten years' progress. See Renaissance Learning and Education Policy Institute (2021). "Understanding Progress in the 2020/21 Academic Year: Complete findings from the Autumn term." London: Department of Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1062287/Understanding_progress_in_the_2020-21_academic_year_Complete_findings_from_spring_term_Oct2021.pdf

26. One example of this is that Anders, Jake, Macmillan, Lindsey, Sturgis, Patrick, and Wyness, Gill (2021). "The 'Graduate Parent' Advantage in Teacher Assessed Grades," *UCL*. https://blogs.ucl.ac.uk/cepeo/2021/06/08/thegraduate-parentadvantageinteacherassessedgrades/
27. Freedman, Sam and Hares, Susannah (2021) [Twitter] 31 May. Available at: https://twitter.com/samfr/status/1399485039440760841?lang=ar (Accessed June 1, 2021).