The Covid pandemic and the cancellation of state examinations caused unprecedented turmoil in the education systems on both sides of the Irish Sea. As the policy of calculating grades using purpose-built algorithms came undone in the face of a barrage of appeal, protest and legal action, the context in which the policies had been devised collapsed. The British and Irish governments had initially adopted similar approaches to issuing examination grades, but then diverged into different stratagems pre- and post-results, with significantly different outcomes. The Irish examination system emerged relatively unscathed, while the system in England suffered what was probably its greatest policy failure of modern times. This article examines and memorialises how and why this happened, and draws lessons for a future in which school closures and substitute examinations become the ‘new normal’.

Keywords: national outcomes; learning outcomes; international context; political economic and cultural contexts; policy; examinations

Introduction

It was the best of times, it was the worst of times... it was the spring of hope, it was the winter of despair; we had everything before us, we had nothing before us. Charles Dickens, opening lines of A Tale of Two Cities (1859)

The spring of 2020 saw the prolonged closure of schools and the cancellation of all state examinations in the UK and Ireland due to the Covid-19 pandemic. This caused unprecedented turmoil on both sides of the Irish Sea. The trigger for the disruption was sudden, but the longer-term consequences were severe, especially in relation to A-Level examinations and their Irish equivalent, the Leaving Certificate (LC), and progression to higher education. In England, A-Level examinations, upon which university entrance is predicated, were cancelled in March 2020 and the UK government announced that the Office of Qualifications and Examinations Regulation (Ofqual) would instead issue ‘calculated grades’ using a standardisation algorithm. Ireland followed suit on 8 May when it announced plans for a similar calculated grades...
replacement. In both jurisdictions, these grades were to be based on teacher predictions moderated using the algorithm to prevent inflation, a major plank of UK government policy since 2010, as it was for the Irish government. The UK Secretary of State for Education, Gavin Williamson, made it clear that his priority was to ‘ensure that no young person faced a barrier when it came to moving on to the next stage of their lives’ and he asked examination boards to ‘work closely with teachers, who know their pupils best, to ensure their hard work and dedication is rewarded and fairly recognised’ (Parker, 2020). A similar approach was adopted in Dublin by Norma Foley, Minister for Education and Skills, in terms of consultation and ‘fairness’ in recognising student achievement. Both jurisdictions shared a concern about the differential impact of the cancellation by social class.

As their respective approaches to calculating replacement grades using purpose-built algorithms came undone in the face of a barrage of appeal, protest and legal action after the grades were issued in autumn, the policy context collapsed. The two governments had initially adopted similar approaches, but had diverged into different stratagems pre- and post-results with significantly different outcomes. The Irish examination system emerged relatively unscathed, while the system in England suffered its greatest policy failure of modern times. This article looks at how and why this happened.

The grades standardisation algorithm for England

Ofqual considered 11 possible algorithms that could combine teacher-predicted grades—subsequently called ‘Centre Assessed Grades’ (CAGs)—with historical data on school performance and cohort-level data on the prior performance of the 2020 cohort at GCSE in 2018 (Ofqual, 2020a). The algorithm that was eventually chosen—the ‘Direct Centre Performance Model’—was relatively simple. The school or ‘centre’ provided a list of CAGs in each subject with the students ranked in order. System-level data on the link between GCSE grades and A-Level grades was then applied. Firstly, the historical distribution of grades for the years 2017, 2018 and 2019 was calculated for schools. These were compared to previous years’ distributions of predicted grades for students at the same schools, relative to their prior attainment at GCSE. This was intended to represent how accurate previous predictions were at each school; if a school had a track record of overly optimistic predictions, this would be expected again in 2020. Finally, the predicted distribution of grades at each school, based on school-level prior attainment at GCSE, was examined, so that a cohort with a high percentage of grades 9 and 8 (say) at GCSE could expect a high percentage of grades A* and A (say) in the 2020 A-Level exams (Ofqual, 2020b). The teacher-predicted grades and teacher rankings submitted by each school were raised or lowered to fit this distribution, and other adjustments were made for cases where prior attainment data was not available. The algorithm was only used for cohorts of 15 or more. For cohorts with fewer than 15 students, the raw CAGs were used.

It was a simple algorithm, but there were issues. For one thing, the modelling of prior attainment at national level caused schools that did better than the national average to lose out. Data published by the Fischer Family Trust showed that even
when no students had failed in the previous 3 years, it was still possible for the percentage ‘expected to fail’ to be non-zero (Taylor, 2020).

The Direct Centre Performance Model was primarily designed to combat A-Level grade inflation, although there has always been significant year-on-year variation in results at the school level in England where (unlike Ireland) considerable churn is created by students changing school after GCSEs. As early as March 2020, the minister had instructed Ofqual to ‘ensure, as far as is possible, that qualification standards are maintained and the distribution of grades follows a similar profile to that in previous years’ (Adams et al., 2020), and he issued a ministerial directive to that effect under the Children and Learning Act 2009. Ofqual took the view, as did the Irish Department for Education and Skills, that simply asking teachers to estimate student grades would create inflation because they would, as they always did (sic), consistently overestimate performance. There was also the question of timeliness: Ofqual needed to process the results quickly. This would also be a factor in the Irish calculated grades system, which ‘operated under unavoidable time pressures’ (Lawlor, 2020), although the Irish government had the advantage of later publication of results.

Details of the Ofqual algorithm were not released until after the results were published in August 2020, when some of the details, but not all, were made public (Harkness, 2020). This was perceived by disgruntled students and their parents—and there were many—to be a ‘black box’ of politically motivated manipulation. Ofqual, which is not an independent agency but a government department whose leadership is in the gift of the minister, seemed to be caught between the politicians and its public remit. Although it had given clear guidance on how schools were to use past performance in generating CAGs in a ‘fair and objective’ way, they were explicit in their expectation that different schools could and would use the guidance differently (Taylor, 2020), despite the fact that schools in England have been giving out predicted grades for decades as part of the annual university application system (UCAS). In fact, schools had submitted predicted A-Level grades a few months previously, on 15 January, before the pandemic took hold and months before schools were closed, and these had already been shared with students. It is strange then, given the considerable experience that schools in England had in predicting grades, but their lack of experience in ranking students, that the latter was taken into consideration, but not the former. Additionally, the process of ranking students, while designed to avoid clustering, was known to create a serious issue for small cohorts where teachers, inexperienced and conflicted, were forced to separate those students on the boundary by a full grade.

Testing the English algorithm

A predictive algorithm would normally be tested prior to use by running it against the previous year’s data where the outcomes were known. The 11 algorithms that Ofqual considered were applied to the 2019 cohort, but the test was incomplete because there were no teacher-predicted rankings. Ofqual generated rankings retrospectively for the 2019 cohort test by placing students in order according to the marks they actually got in their exams (Harkness, 2020), but of course this weakened the test because the real grades and the estimated grades were based on the same set of marks, so that a high level of agreement was inevitable. And yet, even then, Ofqual (2020a: 7–8) had to
admit in its Interim Report that the accuracy of the chosen algorithm still varied hugely between subjects. Taylor (2020) claimed that at best, it was 68% accurate; at worst, 27%.

Effects of the algorithm

The 2020 A-Level grades were announced on 13 August. Overall, nearly 36% of grades emerged lower than the CAG and 3% were down two grades (among the three or four A-Level subjects usually taken). Students at small schools or taking minority subjects with fewer than 15 students, which is more common in small fee-paying schools, received higher calculated grades than their CAGs because these cohorts traditionally had a narrower range of marks and students were usually selected on ability. Conversely, students at large state schools with open-access policies saw their predicted grades plummet because that fitted their historic distribution curve. Students and parents perceived this to be unfair and there was huge upset and protest as a result of the perception that the system was disadvantaging pupils from lower socioeconomic backgrounds. Ofqual’s deputy chief regulator, Michelle Meadows, confirmed that pupils from lower socioeconomic backgrounds were ‘more likely to have seen a bigger downward adjustment’, but attributed this to the ‘tendency for more generosity in the predictions for students from lower socioeconomic status backgrounds’ and asserted that there was ‘no evidence of systematic bias’ in the algorithm (cited in Hazell, 2020). It was a disconcerting explanation—that poor students were more likely to be downgraded because teachers were more likely to have been generous—and one without any evidential basis, but there was a simpler technical explanation that pointed to a flaw in the very notion of a calculated grades algorithm; namely, that bias is inherent in the (legitimate) assumption that statistical adjustment is methodologically unsound for small cohorts, and since smaller classes are much more common at private schools than at state schools, results from private schools were less likely to be downgraded. Ofqual was willing to lay the blame on overly generous teachers, but not on its own algorithm.

Following the publication of results, serious political pressure was applied to the minister, from within his own party and from opposition politicians, to reverse his decision to use the algorithm. On 12 August, he announced a ‘triple-lock’ appeal system that would allow school ‘mock’ exams to be taken into account. Three days later, when the appeals procedure was published, it varied considerably from what he had earlier promised, but the minister stated that there would be ‘no U-turn and no change’ (Walawalkar, 2020). He criticised the Scottish government for its U-turn the previous week10 and restated his belief that awarding unmoderated grades would be ‘unwise’ and would ‘cause rampant grade inflation’. The Prime Minister then weighed in, stating that the results were ‘robust and dependable’, but later the same day Ofqual suspended the system (Guardian, 2020a). Two days later, on Monday 17 August, Ofqual announced that students would be awarded their CAG grades instead of those moderated by the algorithm, and a week later, on 25 August 2020, the Head of Ofqual, who oversaw the development of the algorithm, resigned (Richardson, 2020). Three days later, the Permanent Secretary at the Department for Education (DfE), its most senior civil servant, followed suit.
Initially, the algorithmic results for A-Levels—that is to say, the moderated or adjusted CAGs—showed a slight increase in grades compared to 2019. A total of 39% of grades (approx. 280,000) had been lowered from their CAGs and 2.6% had been raised. In 58% of cases there was no change to the grades predicted by teachers. The top grades (A* and A together) increased by 2.4% to their highest level in recent years, but Ofqual proudly asserted that if it had not moderated the CAGs, that percentage would have surged by a whopping 12.5% (Clark & Gilbert, 2020; Hazell, 2020). Given that the moderated results were abandoned within a week (on 17 August) and the unmoderated CAGs awarded instead, we therefore know that 2020 has produced the most incredible grade inflation ever seen in the system, despite the fact that the explicit policy aim of the algorithm was to prevent it.

It later emerged, after the chair of Ofqual had unwisely accused the Royal Statistical Society (RSS) of spreading ‘misunderstanding and suspicion’ about Ofqual’s work, that the RSS had offered to help with the construction of the algorithm, but withdrew their offer when they saw the nature of the non-disclosure agreement they would have been required to sign. Ofqual was not prepared to enter into discussions with the RSS and delayed replying for nearly 2 months (Murray, 2020). It also emerged that OCR, one of England’s main examination boards, told the minister that the algorithm was producing some rogue results, but the minister and his department were told by Ofqual that these would be corrected by the appeals procedure. OCR replied that it was ‘more than a few results’ and that ‘patterns could be observed’, but this was also ignored (Adams & Stewart, 2020). On 2 September, Ofqual’s chairperson, Roger Taylor, appeared before the Education Select Committee and apologised to students, parents and teachers.

University entrance

In the midst of all this turmoil, UCAS announced on 19 August that some 15,000 students who had been rejected by their first-choice university because of their lower algorithm-generated grades, could now reapply using their CAGs instead. Universities were left to deal with the mess, but (like over-booking on airlines) they no longer had room for everyone now qualified to get in. 90% of these students were holding offers for the most selective courses in selective universities. The situation with medicine was especially problematic, with an estimated 2,000 extra students entitled to places, but with capacity for only 600 (Adams et al., 2020; Guardian, 2020b). The junior minister with responsibility for universities wrote to vice-chancellors on 20 August promising funding for more medical school places, but only if the Department of Health could guarantee clinical placements during their courses and training places afterwards (Taylor, 2020). At the time of writing, this has not yet been resolved.

Meanwhile, in Ireland . . .

For historical reasons, the education system in Ireland is similar to that in England and the two share an imperative, as do all developed economies, to do well in international comparison tests like the OECD’s Programme for International Student Assessment (PISA). Nevertheless, although high pupil attainment is considered both
an input and an outcome of economic success, both countries had in recent years shared a concern for grade inflation and maintaining the integrity of their respective public examination systems. During the period of the Covid pandemic, amidst the preparations for calculating proxy grades for the cancelled examinations, these concerns were made very explicit by ministers in their respective jurisdictions. As things turned out, they both failed, but to varying degrees. The number of top grades in the LC\textsuperscript{13} increased. They usually follow (approximately) the bell curve to ensure that there is no significant deviation from year to year, but in 2020 the official figure for inflation in the final awarded grades across all subjects and across all levels was 4.4\% (O’Brien, 2020a). In the circumstances, this was modest and approximately one-third of that in England. Additionally, for the vast majority of subjects, the modal grade stayed the same, at H4.\textsuperscript{14} The extent of the inflation in top grades varied by subject of course, but again, although the ‘hawks’ were not happy that there was any inflation, it was relatively modest in the circumstances: year-on-year, H1 in Mathematics rose by 2\%, H1 in Irish rose by 3\%, H1 in English rose by 1.3\% and H1 in Geography rose by 2.1\%.

It became clear during Ireland’s moderation/calculation process that teachers had overestimated their students’ grades at all levels, especially at the top end where the initial teacher-predicted grades were sometimes two, three or four times higher than normal, but these predicted marks were adjusted downwards by the algorithm. In total, 17\% of teacher-predicted grades were pulled downwards to increase parity with previous years, but most teacher predictions remained intact (O’Brien, 2020a) and this has been a source of encouragement for the profession and the public in Ireland. Overall, 34\% of the 2020 cohort of (60,000) students did not have any of their teachers’ grades lowered by the algorithm; 32\% (18,584) had one of their predicted grades lowered; 20\% (11,663) had two grades lowered; 9\% (5,288) had three grades lowered; and a further 3\% (1,885) had four grades lowered. As one member of the Irish parliament, Jennifer Carroll MacNeill, noted, it was ‘painful for these students and their families, but the numbers are small in the overall context of 60,000 students’ (O’Brien, 2020a).

Initially, the Irish algorithm proposed using four data sources:

- Category 1: Teacher-calculated grades and teacher rankings of students, as supplied by schools.
- Category 2: Junior Certificate (JC) prior attainment at the school level.
- Category 3: Prior attainment at the school level in the previous 3 years.
- Category 4: Historical system-level profile data on a subject-by-subject basis.

On 24 August, the Assistant Secretary General of the Irish Department of Education and Skills (DES), Dalton Tattan, issued a departmental Memorandum to the Independent Steering Committee for Calculated Grades (ISCCG) ‘setting out changes to its terms of reference’. It made explicit reference to the UK (sic) debacle and was

\ldots conscious that the use of school-by-school historic data has been criticised in the public discourse about calculated grades and has led in the UK context to accusations that students attending disadvantaged schools were at risk of being unfairly treated\ldots [and] aware of the loss of public confidence in the moderation process and in the results issued by the authorities that occurred in the UK as a result. (Tattan, 2020)
A less direct reference to the experience in England was also contained in the *1st Letter of opinion regarding the calculated grades system* written to minister Norma Foley by the chair of the ISCCG, Dr Aine Lawlor, on 2 September 2020, which stated that consideration was given by the minister and the working groups to the public disquiet ‘in other jurisdictions where versions of calculated grades have been transacting’ (Lawlor, 2020: 8).

The Tattan Memorandum stated that while the minister acknowledged that the original purpose of running the algorithm was to arrive at ‘the most likely outcome of results if the students had undertaken the traditional examinations’, she had decided, in light of the imperative to ‘treat outlier candidates and small classes fairly’, to run the standardisation model without Category 3 data; that is to say, without prior attainment at the school level. The minister further acknowledged that the public was unsettled at the prospect of adjusting grades to ‘rigidly maintain year-to-year comparability in national standards’ (i.e. to prevent grade inflation) and consequently that Category 4 data (historical subject-level data at the system level) should also have ‘greatly diminished importance’. The Memorandum therefore reduced the algorithm to Category 1 data (teacher-predicted grades and within-class teacher rankings) and Category 2 data (JC prior attainment at the school level).

In relation to grade inflation, the *1st Letter of opinion* (Lawlor, 2020) made clear that the overall settled aim was to ‘maintain a sense of fairness’—an interesting turn of phrase as it relates not just to fairness but to the *perception of fairness*—and acknowledged that the ministerial decision to amend the algorithm to ‘a less strict application of the original model’ meant that grade inflation was inevitable; in other words, that 2020 would have ‘considerably stronger results’ compared to previous years and that this was acceptable. The original (pre-Tattan) imperative was simply to ‘maintain standards’ and prevent grade inflation, but that was subsequently adjusted to include the proviso that it had to be ‘proportionate to the sense of fairness for all’ and in particular that ‘stronger performing students in traditionally lower-performing schools’ could ‘attain beyond the normal range of grades’ (Lawlor, 2020: 8). It made a virtue of political necessity, predicted from ‘the UK experience’, but the concern for students in ‘traditionally lower-performing schools’ was well founded and shared across the two jurisdictions. It is known that socioeconomic status is an important factor in attainment and there had been fears in Ireland that students attending schools in disadvantaged areas would be unfairly penalised by the algorithm. Largely as a result of the Tattan adjustments—specifically ditching Category 3 and Category 4 data—these fears were fortunately not realised. Data subsequently released by the DES shows that while 16.8% of grades were downgraded by the algorithm, the figure for disadvantaged schools—schools participating in the Delivering Equality of Opportunity in Schools (DEIS) programme—was significantly lower at 13.6%. And while 3.9% of grades awarded by schools were raised by the algorithm overall, that figure was higher for DEIS schools, at 5% (BBC, 2020). The only (as yet) unaddressed concern relates to gender. The ISCCG found that girls on average outperformed boys *to a greater extent* in 2020 than in previous years, which, as the algorithm did not control for gender, was ‘attributable to the school estimates conferred on students’ (Lawlor, 2020: 11). This needs to be followed up, although it is worryingly similar to Ofqual’s earlier assertion that poor students in England were more likely to be downgraded because teachers were more likely to have been generous in their grading.

© 2021 The Author. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.
In retrospect, having taken the reasonable decision on public health grounds to cancel the final public examinations, it was unreasonable also to exclude other forms of assessment. Actually, in England, Ofqual had advised against cancelling the examinations, suggesting that holding them in a socially distanced manner was the better option, but the UK government cancelled anyway, as did the three devolved administrations. Ireland followed suit shortly afterwards, despite the fact that the epidemic was less keenly felt there and the Irish government was more effective in introducing containment measures at an earlier stage. In other European countries, school-leaver examinations were held as normal, or delayed slightly. Germany’s 16 states, which decide their own education policies within a federal structure, were initially divided over whether or not to cancel the Abitur examinations, but agreed in late March that they should go ahead as planned, despite nationwide school closures. They were held with smaller numbers of students sitting socially distanced in well-aired classrooms, rather than in large assembly halls, although of course the pandemic was less keenly experienced in Germany than in the UK. And in Italy, although the written papers of the high-school Esame di Maturità examinations were cancelled, half-a-million students took the oral part in mid-June (Henley et al., 2020).

In both England and Ireland, the cancellation of examinations and the grade inflation that followed caused massive turbulence, undermining public confidence in the assessment system and introducing an element of randomness (or at least the perception of it) into educational outcomes. However, the educational establishment in Ireland came out of the crisis better than its UK counterpart. The mess that was the calculated grades process in England proved useful for Ireland’s DES, which had a few extra weeks to learn from British mistakes and adjust their strategy accordingly (O’Brien, 2020b). It was a simple choice, as the Irish policy-makers viewed it: either accept relatively generous grades for the 2020 cohort, or hold a hawkish line on grade inflation to maintain comparability with previous years. They opted for the former, having initially planned for the latter, with officials saying that they were now ‘prioritising fairness for the class of 2020 over eliminating grade inflation’ (O’Brien, 2020b). The government simply changed policy, even though the algorithm, designed with the latter policy in mind, was up and running.

Similarities and differences

The most important similarity between the English and Irish secondary school systems is their dependence on a final high-stakes examination. Ireland has been like this since 1925, so the hands of policy-makers, universities and teachers were tied in what they could tweak in response to the pandemic. England was similarly tied, but need not have been! Its A-Level system had been reformed in 2002 by the first New Labour (Blair) government to reduce dependence on a set of final examinations—it broke A-Levels down into ‘AS’ and ‘A2’ components, examined respectively in the first and second years of senior cycle—but this modular approach was abolished in 2016 by the Conservative education minister, Michael Gove. So if the Covid pandemic had struck before 2016, students would already have banked half their marks, and their final grades could then have been extrapolated easily and robustly, and the current Conservative government would not have been caught out so badly. Defeat was snatched from the jaws of victory!
Of course, there are also major differences between the Irish and English systems, particularly in the way they use their respective examinations for university matriculation. The Irish exam system is more fine-grained than A-Levels, in that the best six subjects from seven, rather than three subjects from three or four, are finely tuned using eight grades, rather than six (see note 5). This fine grain allowed teachers in Ireland more latitude in predicting grades and ranking students, so the Irish system could tolerate some inflation between sub-grades without massive overheating of the system when it came to university entrance, which is completely demand-managed. The admission thresholds for third-level courses in Ireland rise and fall annually with demand, with grades converted to tariff points and thresholds calculated after the full results from the entire cohort nationally are known. In contrast, the English system operates a conditional admission system on the basis of teachers’ predicted grades, even though pupils from poorer backgrounds tend to have lower predicted grades18 (Murphy & Wyness, 2020) and usually lack the confidence to apply to elite universities (Dillon & Smith, 2017; Campbell et al., 2019).19 Applicants receive an ‘offer’ of a place contingent on obtaining certain grades, which then becomes legally binding if the student achieves those grades, but in many cases, even when the predicted grades fail to materialise, the student is admitted anyway because to do otherwise would create huge churn in the clearing system and within universities. There is a small measure of demand management in the English system, but nothing like the Irish system. In England, offer grades are pitched so that, on the basis of results from previous years, courses will not be too over-populated with successful applicants, in the manner of airlines over-booking their seats, but in most years, courses are overfilled because of grade inflation. In Ireland in 2020, as a result of the 4.4% grade inflation, university admission thresholds rose, but fortune favoured policy-makers, because those higher tariffs were in part mitigated by a lower number of international students applying to study in Ireland, and by the astute provision of an additional 1,250 third-level places (McGuire, 2020). Similarly, in France, the 7% increase in the pass rate for the 2020 Baccalauréat was in part mitigated by the French government creating 10,000 extra university places (Henley et al., 2020).

As discussed already, teacher rankings in both the English and Irish systems were intended to avoid clustering, but the process created a serious issue for small cohorts where teachers were forced to separate students on the boundary by a full grade. This is always troublesome, as the lowest ranked student in the ‘B’ or ‘H3’ category is the one most likely to end up in the ‘C’ or ‘H4’ category. This is critical for university admissions in England, but not so much in the more fine-grained points-based system in Ireland. In the English system, if an A were required in a certain subject for admission to a course—for example, an A in chemistry to study medicine—then a B in the subject does not suffice.20 In the Irish system, a reduction from H2 to H3 is a loss of 11 points, which may or may not make any difference to admission because the threshold might fall, but additionally there is always the opportunity to compensate for that loss by getting higher points in another subject. Sometimes in England, A-Level grades are converted into UCAS tariffs (where an A* grade is worth 56 points, an A is worth 48, a B 40, a C 32, and so on), so a university can demand either specific grades like ‘ABB’ or specific tariff scores like ‘128 points’, but only one-third of universities use this system (UCAS, 2020). This inconsistency and lack of demand-responsiveness in the English
system made it harder to be flexible during the Covid crisis. There was little choice but to stick to the fight against inflation and when that dam was breached, the system lost credibility. Secretary of State Gavin Williamson had only been in office for 8 months, since the end of July 2019, so in fairness, it was a fault of long-standing in the system he inherited.

Overall, in Ireland in 2020, some 54,000 students made university applications in the usual way and some 24,500 places were first-preference allocations. More than 80% of students were awarded one of their top three preferences (Lehane et al., 2020), but inevitably, there were individual losers in the system, just as there were in England. Aggregated trends and averages mask the personal despair and individual tragedy of futures lost or rendered uncertain. In Ireland, this applied to four groups.

- Those who sat the LC in 2019 but deferred for a (‘gap’) year. Up to 20,000 applicants applied for courses starting in 2020 on the basis of results achieved in 2019 or earlier. These applicants were disadvantaged because of grade inflation. The attainment of the ‘class of 2020’ devalued the currency that these older students had earned in previous years (O’Brien, 2020b).
- Students from non-Irish backgrounds. Some 2,000 students from non-Irish backgrounds sat exams in minority languages such as Croatian and Polish. About half of these (935 students) were unable to get a calculated grade because they were studying the subject outside school and did not have prior attainment in the JC (see note 3). The system responded in part by ensuring that any student who was relying on these subjects to satisfy university entry language requirements got an exemption (O’Brien, 2020b), but there was no concession regarding grades.
- Ireland offered 2020 LC students the choice of accepting their calculated grades or sitting the examinations in person in mid-November, but fewer than 3,000 students took up the offer (McGuire, 2020). Anyone sitting the November examinations received whichever was the higher grade from the two processes, and those who went on to receive a better (i.e. higher preference) university offer on the footing of their November results received deferred college places to start in 2021 (O’Brien, 2020a). The problem is that students disappointed with their calculated grades were ill-prepared for the November super-high-stakes examination ‘re-runs’, especially having taken the summer off from study, a necessity for poorer children. It was a generous ‘second bite at the cherry’, but not unproblematic.
- Finally, those students who had their grades moderated downwards to a significant extent, given the grade inflation, not only lost ground against their expectations but lost ground against the rising tide of demand-driven tariffs. Students who were expected to get 500 points (say) and thus admission to a prestigious course in a popular university, but were downgraded by the algorithm to 450 points (say), lost not only 10% against expectation, but an additional percentage against the rising tariff for that course. In effect, they lost an additional percentage of their attainment because the algorithm used aggregated data from their peers and allowed grade inflation.

Some of these groups won a Judicial Review in the High Court in Dublin based on the claim that the minister had acted unfairly by opting not to include historical data about schools’ past performance as part of the calculating process. The legal
cases raise some technical and theoretical issues with both English and Irish algorithms, which we will now consider.

**Technical and theoretical concerns**

Most of the problems with the algorithms arose from *policy* decisions that were made at the outset, rather than *technical* issues that emerged during the process. The Irish algorithm, which was developed by the Canadian company Polymetrika, was more robust and technically advanced than its English counterpart, but two coding errors were discovered in the algorithm after the final calculated results were published. The minister, Norma Foley, estimated that around 6,500 students (10%) had received a grade lower as a result. The first error related to how JC results were used in the calculation. Instead of choosing a student’s two *strongest* subjects (after English, Irish and Mathematics), the algorithm chose their two *weakest*. The second error was that the algorithm included JC marks from the subject ‘Civic, Social and Political Education’ (CSPE), which it should not have done. All erroneous grades were subsequently corrected and no student was disadvantaged by the mistake. The algorithm was later re-checked for similar coding errors, but none were found (Lehane *et al.*, 2020). No technical glitches were reported from the English algorithm.

There are some theoretical issues with algorithms that should be aired even though they were subsequently avoided when the English one was abandoned and the Irish one was radically altered. These issues are not the focus of this article—in extraordinary times, it is reasonable to tolerate some shortcomings—but it is prudent to think about them in the event that further exams are cancelled. One theoretical issue is around the multilevel nature of the different datasets used in the algorithms. Student attainment data is hierarchical: pupil-level data is nested within teacher-level data, which is nested within school-level data, which is nested within system-level data. It is important to preserve the level of the variance in any calculations and to have an appropriate amount of variance at each level. The algorithms used historical data at the school level. Individual student-level prior attainment was omitted, as was teacher-level data. We know from educational effectiveness research that the most important determinant of student attainment is prior attainment—that is to say, how well a student does at A-Level and in the LC depends significantly on how well the student did at GCSE and in the JC—but we also know that there is greater variance between students *within schools* than between schools, so there will always be a sense of personal injustice when an algorithm tries to reverse-engineer expected school-level aggregated performance back to the individual student level. As things turned out in 2020, this approach was subsequently abandoned by the Irish algorithm, although it was part of the initial design, and it didn’t happen with the English algorithm because that was abandoned in its entirety, but it was never defensible that one individual student’s grades would be ‘extracted’ at random from the cluster of others around him and adjusted to fit a higher-level aggregated profile (i.e. to choose one particular student for downgrading and not another student from within the same cluster). Using school-level characteristics for individuals is an arbitrary decision, no matter how robustly the different aggregated datasets are merged and irrespective of how they were weighted to produce national profiles. It is an ecological fallacy.\(^{24}\)
Having said that, it is surprising that universities in England and Ireland did not themselves propose working backwards to applicants’ GCSE and JC results, ‘skipping’ the A-Level and LC calculated grades, which (it was known at the time) were to be adjusted in line with them anyway. The two university admissions systems (UCAS in the UK and CAO in Ireland) are centralised and have all the necessary prior attainment data at the student level. In the terrible societal crisis faced by the two systems, where every way forward was problematic and imperfect, why not use student-level prior attainment for each individual applicant? It is not a perfect solution, but it is at least the student’s own actual attainment.

The second issue relates to validity checks. It is not known what checks were carried out on the English algorithm—the DfE was not as transparent as its Irish counterpart, the DES—but in Ireland it was admitted that checks were not widely carried out ‘due to time constraints’ (Lawlor, 2020: 11). This was not in keeping with the aims or spirit of the standardisation exercise, even if it is reassuring to see it acknowledged, and this will need to be addressed in the event of a recurrence.

The third issue relates to sampling and quality assurance. It is not known what exactly was done in England, but in Ireland schools returned their data online to the Calculated Grades Executive Office (CGEO), who ‘carried out checks and balances to ensure that the data were entered correctly and accurately’. A quality assurance process was then conducted by the CGEO in its subsequent processing and this ‘included a sampling of completed paperwork from 30 randomly selected schools’ to ensure that ‘stability and integrity had been maintained, including the throughput of the estimated marks and the rank order’. This sample was too small for the task in question—there were 786 schools, 61,000 LC students and 440,000 grades to be predicted, calculated and adjusted—and in any case it is not clear that the sample population should have been ‘schools’ rather than ‘teachers’, whose predicted grades provided the major input for the model post-Tattan; or ‘subjects’, which themselves have bias regarding gender and socioeconomic status. The sampling should have been on a larger scale and iterative.

Unintended consequences

An unintended consequence of the grade inflation in the 2020 results—accepted in Ireland by Norma Foley as a quid pro quo for perceived fairness and accepted in England by Gavin Williamson as the result of system failure—will create a problem in 2021 if the system reverts to normal sitting. It has not been decided how 2020 will be treated compared to previous years in terms of the national, subject-level and school-level profiles when it comes to standardising the 2021 results so as ‘to maintain year-to-year comparability in the national standard of the examinations’. Nor is it clear whether the 2021 examinations will themselves need calculated grades and if so what will be used instead of the failed 2020 algorithm; and looking further ahead, pessimistically, if the 2022 examinations need calculated grades, what will be used for student-level prior attainment given that the (A-Level/LC) cohort will not have sat any previous public (GCSE/JC) examinations. The use of algorithms is, as Ireland’s ISCCG rightly stated, ‘not a perfect solution’ and only a realistic option ‘in exceptional circumstances’. This is a fair summation, but it begs the question as to how ‘exceptional’ the 2020 ‘circumstances’ turn out to be, if they are from necessity repeated in subsequent years.

© 2021 The Author. *British Educational Research Journal* published by John Wiley & Sons Ltd on behalf of British Educational Research Association.
Conclusions

The English and Irish algorithms probably tried to do too much in very difficult circumstances. In a severe crisis, some things simply cannot be done to the fullest extent. It would have been optimal, public healthcare permitting, not to cancel the public examinations but to hold them in a socially distanced way in multiple centres, with extra support for students by blending online teaching and home support with some socially distanced in-school teaching. However, having taken the decision to cancel them, it would have been safer to rely on teacher-predicted grades if the effort and resources that went into the algorithm had been put instead into detailed guidance and training about how to take ‘mock’ and house examinations into account and how to adjust predicted grades based on student-level prior attainment at JC and GCSE and subsequent student trajectories.

Attempts to ‘fire-break’ the spread of Covid in both jurisdictions have met with mixed results. Mutations have emerged that appear to be more infectious, though not necessarily more lethal, and the reproduction rate for the disease is fluctuating, so policy-makers should be prudent in preparing for some disruption to public examinations in 2021. There are alternatives to a grade-calculating algorithm, which should be considered:

• Although the current criterion-referenced assessment approach played no part in the crisis, policy-makers might consider using simple norm-referenced grades, awarding (say) a H1 or an A* to the top 4%, and so on. Between 1963 and 1986, A-Level grades in England were norm-referenced and if ‘normalising adjustments’ are going to be made anyway during these exceptional circumstances, then the arguments against it are weakened.

• If public examinations are held, university admissions in England could be changed to a system based on actual achieved grades, rather than predicted grades. Once they know their grades, students can then choose the best university available. The switch to ‘actual-grade admission’ would mean either: (i) applying before the examinations but receiving offers after results are known, if necessary by starting the examination season earlier in the summer term; or (ii) applying and receiving offers after results are known, if necessary by starting the opening university term for ‘freshers’ later in autumn. Both Ireland and England have inspectorates that could be involved, and have considerable experience from earlier moderations.

• Simply accept teacher grades as submitted—this is essentially what happened in England in 2020 anyway—but put resources in place to train teachers in predicting grades and establish teacher networks to moderate them. There are well-established systems in the university sector already for ‘blind second marking’, and these could easily be adapted at no expense. This would need extensive trialling and several simulation exercises during the current academic year to establish parity across exam boards, subjects and regions.

• Use a mixture of student-level prior attainment, continuously assessed coursework, formal house/mock exams and contemporaneous teacher rankings, suitably trialled and tested in the usual manner with data from previous years, to build an individual performance trajectory for every student. France used something along these lines. It was one of the EU countries to cancel all its (Baccalauréat) examinations. The
French education minister, Jean-Michel Blanquer, announced in April 2020 that the country’s 740,000 final-year students would be awarded an average grade for each subject based on coursework and earlier school tests. ‘Local juries’ assessed and adjusted student grades, albeit using national averages and school-level prior examination performance (Henley et al., 2020).

Whatever approach is adopted, it really all comes down to confidence. For an assessment system to work, it must have credibility among stakeholders: universities must have confidence that grades reflect academic attainment and the ability of applicants to pursue courses at third level; employers must have confidence that the award of a school certificate guarantees certain skills and competencies; parents need to believe that their children have achieved something important and have acquired some ‘capital’ from their years of formal schooling; and society needs to rely on the system to ensure the credibility of qualifications over time and compared to competitor nations. It is a pyramid of trust needed in all developed schooling systems: pupils trust their parents and guardians; these in turn trust teachers and school principals; these professionals trust the system and the relevant government department; and department officials trust policy-makers. Extensive consultation and buy-in is absolutely necessary for something like a calculated grades algorithm, but consumer confidence is as complex an issue in education as it is in economics. Confidence from one group flows through the system to other stakeholders. If school principals and teachers have confidence in the examination system, that transfers to students and to external stakeholders. However, it is not just public trust in the system that is at stake, but the perceived mistrust by policy-makers of teachers’ capacity for professional judgement. Indeed, algorithms are by their nature a sign of mistrust; the supposedly objective replacement of supposedly subjective professional judgement. The Covid examination crisis caused confidence to collapse in England and showed the system to be less trustful of teachers and less capable of adjustment even when the political will was there and necessity demanded it. The Irish government was lucky in that it could make decisions based on avoiding what had happened earlier in London, but in operating a single centralised state examination board, it was also better equipped to respond to system-wide challenges and its policymakers were more flexible in being able and willing to change direction. Arguably, the Irish minister’s decision to reduce the algorithm from four datasets to two, changing the basis of the grade calculation from that originally advertised, should have been preceded by a wide consultation with principals and teachers, but generally speaking, the public in Ireland retained its confidence in the integrity of the examination system, the professionals within it felt that they had been heeded, and the minister escaped censure. The same cannot be said of the English experience. Although there is no evidence that the UK minister was any less committed or any less competent, the system was exposed as flawed in several important respects. Going forward, both jurisdictions must work to ensure, in addition to some ‘hardware’ restructuring, that all stakeholders are extensively consulted and have confidence in whatever examination proxy is adopted.

Conflict of interest
The author has no conflict of interest of any sort, political or otherwise.
Ethical statement

This article follows BERA ethical guidelines (4th edition). The political actors have been treated with respect throughout the commentary.

Data availability statement

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

NOTES

1 The response to the pandemic varied across the four home nations of the UK because health and education are devolved matters. There are challenges for authors using adjectives like ‘British’ and ‘English’ in an education context because although many features are shared across England, Wales and Northern Ireland, there are important differences, and Scotland has its own system. This article focuses on a comparison between Ireland and England, so ‘England’ and ‘English’ are the preferred terms except when referring to something common to all four home nations.

2 Norma Foley TD has been ‘Minister of Education and Skills’ (now ‘Minister for Education’) since June 2020. The Minister in post during the early stages of the pandemic, who announced the cancellation of public examinations on 8 May 2020, was Joe McHugh TD (2018–June 2020).

3 The General Certificate of Secondary Education (GCSE) is the academic qualification taken by all 16-year-old state school students in England, Wales and Northern Ireland. State schools in Scotland use the Scottish Qualifications Certificate instead. The examinations are taken after 5 years in secondary school and 2 years before A-Levels. The ‘Junior Certificate’ (JC) is Ireland’s equivalent examination, taken 2 years before the Leaving Certificate (LC).

4 For context, the current numerical grading scheme at GCSE is part of a curriculum introduced in England in 2014 by Michael Gove when he was Secretary of State for Education. The highest grade is 9 and the lowest is 1. The old grades A* and A are covered by the numeric grades 9, 8 and 7. The old grades B and C are covered by the numeric grades 6, 5 (‘strong pass’) and 4 (‘pass’). The old grades D, E, F and G are covered by the numeric grades 3, 2 and 1. U (‘ungraded’) is effectively 0.

5 For context, there are six ‘pass’ grades at A-Level: 40% corresponds to an E grade; 50% to a D; 60% a C; 70% a B; 80% an A; and 90% an A*. There is one ‘fail’ grade, U. It varies from year to year, but approximately 8% of grades are A*, 18% are A, 27% B, 24% C, 15% D, 6% E and 2% U.

6 There is very little churn in the state system in Ireland (which includes fee-paying schools) as so few students change schools after Junior Certificate at age 16, and there is no ‘sixth form college’ (equivalent) sector.

7 Leaving Certificate results 2020 were published on Monday 7 September. A-Level results were published approximately 1 month earlier, on Thursday 13 August. In previous years, the two sets of results were issued 1 or 2 days apart.

8 The Universities and Colleges Admissions Service (UCAS) is the organisation that operates the centralised application process for UK universities. It is an independent charity, funded by fees charged to applicants and universities, plus advertising income.

9 UCAS predicted grades were submitted on 15 October 2019 for Oxford and Cambridge, and all medicine applicants.

10 Scotland downgraded approximately 25% of grades in order to ‘maintain credibility’, but following widespread criticism and a public apology (on 10 August) from the First Minister, the Scottish government agreed (on 11 August) to upgrade thousands of examination results and accept teacher-predicted grades. A week later (on 18 August) the Scottish Qualifications Authority announced that 75,000 new exam certificates would be issued. 2.2% were increased by one grade, 0.05% by two grades and 0.01% by three grades.

11 Of the 160,000 students whose results were re-upgraded, 100,000 had already been accepted by their first-choice university.

12 There are actually three Leaving Certificate programmes. (i) The traditional established Leaving Certificate, introduced in 1924, is by far the most popular. A minimum of six subjects, including Irish, are examined; most students take seven, from which their best six grades are taken. The LC is mainly used for university entrance. (ii) The Leaving Certificate Vocational Programme (LCVP), which is more a practical hands-on complement to the traditional LC, but otherwise the same. The student takes at least five traditional subjects, one of which must be Irish. Two subjects must be practical-based and another must be a modern European language. (iii) The Leaving Certificate Applied (LCA), which is designed for students who do not wish to proceed to further or higher education. It is taken to prepare the student for work. Calculated grades were used in all three programmes, but the data cited in this article refers mainly to the traditional academic LC programme.
Subjects in the LC are taken at one of three levels: Higher (H), Ordinary (O) or Foundation (F). Each is graded from 1 (highest) to 8 (lowest), with each carrying a points tariff for university entrance: H1 (90–100%) grade is worth 100 points; H2 (80–90%) is worth 88 points; and so on down to H7 (30–40%) worth 37 points; O1 is worth 56 points; O2 is worth 46 points; and so on down to O6 worth 12 points; F1 is worth 20 points and F2 is worth 12 points. 25 bonus points are awarded for Higher Level Mathematics for H6 grades and above, so, for example, a H2 in Higher Level Mathematics is worth 113 points.

A programme comprising 230 disadvantaged schools, of which 134 are in Dublin.

Grade inflation does seem to follow exam cancellation. Results were only marginally higher in the Abitur in Germany where examinations were held as normal, but in France, where the Baccalauréat exams were cancelled, the pass rate was 7% higher than in 2019 (Henley et al., 2020).

In England, AS is still offered, but as a separate qualification and AS grades no longer count towards a subsequent A-Level.

In the UK, only 16% of applicants typically achieve the A-Level grade points that are predicted by their teachers. 75% are over-predicted (i.e. higher than they actually achieve). The average student is over-predicted by 1.7 grades, but low socioeconomic status students fare worse—their predicted grades are not as over-predicted—which adversely affects their university entrance choices. Type of school attended is also a factor: high-achieving students from state schools receive lower predictions than those from private schools (Murphy & Wyness, 2020).

High-achieving but under-predicted candidates are 10% less likely to apply to the most selective universities, and approximately 7% more likely to enrol in a university where they have grades higher than their peers (Murphy & Wyness, 2020).

Typically, admission to Oxford and Cambridge universities, irrespective of course, requires ‘AAA’ in specific subjects. ‘AAB’ will mean the offer (and the place) is forfeited.

Minority languages are almost always sat by students who are bilingually raised in a family where at least one parent speaks the language. These subjects have the highest grades in the LC.

Held on weekday evenings, and on Saturday and Sunday mornings from 16 November to 11 December. All practical coursework was carried over from the 2019–2020 school year to the November sitting of the examinations (O’Brien, 2020b).

For example, Freddy Sherry (applicant) v. the Minister for Education and Skills et al. and the Attorney General (respondents) (O’Loughlin, 2020). Sherry’s was the lead test case for approximately 50 challenges by other LC students against the calculated grades process.

The ecological fallacy consists of thinking that relationships observed for groups necessarily hold for individuals. For example, just because a student attends a high-performing school with the highest average attainment in an area does not mean that the student is automatically a high attainer.

Despite good news on the development and roll-out of vaccines, all public examinations in the UK have been cancelled for 2021.

References

Adams, R. & Stewart, H. (2020, September 1) Williamson told about flaws in A-level model two weeks before results, The Guardian.

Adams, R., Elgot, J., Stewart, H. & Proctor, K. (2020, August 19) Ofqual ignored exams warning a month ago amid ministers’ pressure, The Guardian.

BBC (2020) Coronavirus: 17% of Leaving Cert grades predicted by school to be reduced. Available online at: www.bbc.co.uk/news/world-europe-53986003 (accessed 10 October 2020).

Campbell, S., MacMillan, L., Murphy, R. & Wyness, G. (2019) Matching in the dark? Inequalities in student to degree match. CEPEO Working Paper No. 20–01 (London, Centre for Education Policy and Equalising Opportunities, UCL).

Clark, A. & Gilbert, D. (2020, August 18) In charts: How the A-level algorithm U-turn fails to fix grade inequality, The Telegraph.

Dillon, E. & Smith, J. (2017) Determinants of the match between student ability and college quality, Journal of Labor Economics, 35(1), 45–66.

Guardian (2020a, August 15) Ofqual suspends A-level appeal criteria hours after announcing them, The Guardian.

Guardian (2020b, August 21) ‘We’re picking up the pieces’: A university vice-chancellor’s diary of A-level chaos, The Guardian.

Harkness, T. (2020) How Ofqual failed the algorithm test. Available online at: https://unherd.com/2020/08/how-ofqual-failed-the-algorithm-test/ (accessed 14 October 2020).

Hazell, W. (2020) A-level results 2020: 39% of teacher predicted grades downgraded by algorithm amid calls for U-turn. Available online at: https://inews.co.uk/news/education/a-level-results-
2020–grades-downgraded-algorithm-triple-lock-u-turn-result-day-578194 (accessed 20 September 2020).
Henley, J., Giuffrida, A. & Burgen, S. (2020, August 12) School exams and Covid: What could the UK have learned from EU?, The Guardian.
Lawlor, A. (2020) 1st Letter of opinion regarding the calculated grades system (Dublin, DES).
Lehane, M., Cunningham, P. & Doyle, M. (2020) Around 6,500 Leaving Cert students affected by calculated grades errors. Available online at: www.rte.ie/news/2020/0930/1168450-leaving-cert/ (accessed 15 October 2020).
McGuire, P. (2020, September 7) Leaving Cert results: Grade inflation evident across most subjects: Cancelled exams and calculated results lead to widely anticipated grade inflation, Irish Times.
Murphy, R. & Wyness, G. (2020) Minority report: The impact of predicted grades on university admissions of disadvantaged groups, Education Economics, 28(4), 333–350.
Murray, J. (2020, August 24) Royal Statistical Society hits back at Ofqual in exams algorithm row, The Guardian.
O’Brien, C. (2020a, September 28) Almost 100 Leaving Cert students had six or more subject grades lowered: Applications to sit the Leaving Certificate 2020 in November open this morning, Irish Times.
O’Brien, C. (2020b, September 7) Leaving Cert 2020: Winners and losers from calculated grades, Irish Times.
O’Loughlin, A. (2020, September 30) Belvedere College student’s case selected as ‘lead’ action over Leaving Cert standardisation, The Irish Examiner.
Ofqual (2020a) Awarding GCSE, AS, A Level, Advanced Extension Awards and Extended Project Qualifications in Summer 2020: Interim Report (London, Ofqual).
Ofqual (2020b) Requirements for the calculation of results in summer 2020 (London, Ofqual).
Parker, K. (2020) GCSE resits: Students to receive ‘calculated’ grades. Available online at: www.tes.com/news/gcse-resits-students-receive-calculated-grades (accessed 19 October 2020).
Richardson, H. (2020) Ofqual chief resigns after exams chaos. Available online at: www.bbc.co.uk/news/education-53909487 (accessed 14 October 2020).
Tattan, D. (2020) Memorandum of 24 August setting out changes to the terms of reference of oversight groups for calculated grades (Dublin, DES).
Taylor, P. (2020) Ofqual and the algorithm, London Review of Books, 42(17). www.lrb.co.uk/the-paper/v42/n17/paul-taylor/short-cuts
UCAS (2020) UCAS tariff points (Cheltenham, Universities and Colleges Admissions Service).
Walawalkar, A. (2020, August 15) Students in England must meet eight criteria to appeal exam results, The Guardian.