

Algorithms, Ofqual and Regulatory Independence

by ukcivilservant
27 August 2020

The A Level drama exposes an uncomfortable issue:- the balance between regulators' operational independence and decision making, and political accountability.

It is a tricky issue that has been around for decades. Most regulators would want decisions that impose significant distributional effects between classes or types of the public to be subject to greater political accountability. Politicians, of course, would prefer the opposite. This blog seeks to illuminate the debate.

Let's begin by summarising regulatory best practice.

Some regulators are more independent than others

It is in everyone's interest that regulators who make significant economic decisions (such as whether to permit a company merger or to allow gas prices to rise) do so without any political interference. But it is clearly not possible for politicians to wash their hands of responsibility for the quality of education and health services, so the regulators of those services tend to operate in a more overtly political environment.

All regulators need to communicate effectively with both the public and the main political parties.

Regulators take many decisions which deeply affect the lives and finances of millions of people. They should not be swayed by party political considerations, but they do need to anticipate criticism and explain their decisions in clear uncomplicated language. They should have clear and honest channels of communication with interested journalists. And they should brief the relevant government departments – and offer to brief Opposition spokespersons – whenever they announce particularly interesting regulatory decisions.

- Early/mid-morning press-conferences, meetings and conversations will ensure that the regulator's reasoning is widely understood whilst opinions are being formed, even if their judgment is not fully accepted.

Models, Spreadsheets, Algorithms

Regulators are often required to forecast the result of their decisions, and deploy various types of model to help them do so. But models are guides and simplifications of real life – they are not real life. Decision makers need to be careful not to over-rely on them and fall into the seductive trap of relying on model outputs as “truth” when judgment and pragmatism would frankly give better and more legitimate results.

Numerical outputs are best seen as approximations and can be rough and ready in their predictive power. Most analysts who develop models will tell you that 80 per cent accuracy is about as good as you will ever get for a complex social science question. It’s not physics and exact science we are talking about here.

Let’s also remember – to quote Timandra Harkness – that algorithms are ‘prejudice machines’ – which is another big subject in itself.

Consultation

High quality decision-making must be underpinned by effective, open and transparent consultation. It should seldom be necessary to depart from this process:-

- Ask questions, seek information and seek views from any interested party.
- Debate the issues with experts and representatives of key organisations as well as with others who offer interesting and/or challenging views.
- Announce a provisional (‘minded to’) decision and seek comments on it.
- Make a final decision in the light of stage 3 representations.

This process can take several months if the subject is important enough. Equally, it can be carried out in a matter of days if there is a need for urgency.

It is particularly important that the process should not be distorted by the volume of similar consultation responses, often encouraged by lobby and pressure groups. Indeed, it is perfectly possible that just one isolated submission, making a vital and otherwise overlooked point, can help steer the regulator away from a faulty conclusion.

It is also vital that regulators should not be upset by vitriolic criticism but should remain willing to listen to that criticism in case it contains an essential truth that had been overlooked. This is of course much easier to say than to do, especially in today’s world so dominated by social media. But

it helps greatly if there is a mature, mutually respectful relationship between the department and regulator.

Ofqual's Consultations – and the Algorithm

It is not yet possible for anyone outside the government and Ofqual to understand whether they followed the above best practice, but here are some provisional observations.

First, let's not forget that Ofqual were tasked with solving a fundamentally impossible problem in awarding grades to students who had completed neither course work nor formal exams. Its decisions would be very closely scrutinised, especially if they had distributional effects. There would need to be close collaboration with ministers and their officials.

It is often the case that regulators are faced with a choice between options which have different distributional effects: trade-offs where one group of the population benefits at the expense of another. Should gas/electricity standing charges be increased more than unit prices? Should regional pricing be allowed? Will increased competition benefit the affluent and IT-savvy, to the detriment of those who find it difficult to shop around?

It is an article of faith amongst regulators that politicians and not regulators must own such distributional effects. Regulators can advise but such effects are by their nature political and require an element of democratic legitimacy. Increasingly, though, politicians have started asking regulators to take essentially political decisions, most obviously in energy regulation where [Ofgem has been forced to accept a substantial proliferation in the number of general duties](#). Since the 1986 Gas Act, the number of duties has risen from eight to twenty-one. And the position is broadly similar for electricity. The prioritisation of these often conflicting duties should not in principle be left to the unelected regulator.

Part of the impossibility of the task facing Ofqual was that, as noted above, models are fine when all the decision maker needs is an 80% approximation. But, when you are talking about exam grades, every individual matters – you can't simply say 80% is good enough. It leads to too much rough justice and unfairness, exacerbated in this case by teachers being forced to rank students who were in reality indistinguishable. Ofqual's board and DfE might therefore have been expected to ask (and might well have asked) whether alternatives to the model would have been acceptable and less prone to rough justice. Could teacher assessed grades be used subject to Ofqual checks – or even peer review from teachers in other schools using whatever evidence was available?

It is hard to understand why the model and its assumptions were not shared transparently before the results. Ofqual published a 319-page document explaining its methodology only after the A-level results had been published. It is not clear why this could not have been published much

earlier. This would always be good regulatory practice, and was even more vital given the novel and political contentious nature of the task. It would have helped draw out the sharp edges earlier and hopefully driven different decision making.

The fact that Ofqual could not find a way to take up the offer by the Royal Statistical Society to review the model was particularly odd. I have never heard of a regulator requiring its advisers to sign Non-Disclosure Agreements.

And yet ... there was clearly enough information in the public domain to cause concern amongst those that tried to understand it. So did Ofqual find it difficult to accept that non-expert criticism might be well-founded? This is often the case with experts, of course, who feel that their professional honour is impugned whenever an amateur or outsider seeks to contribute to a debate.^{[1] [2]} But it does appear that input from former DfE official Jon Coles, IT consultant Huy Duong, and his statistician sister, as well as the Education Select Committee amongst others, should have given the regulator pause for thought.

On the other hand, there was an 11-person external advisory group (including Tim Leunig, a reportedly somewhat maverick Treasury Economic Adviser) whose discussions have been reported as 'robust' but leading to consensus.

If there were no apparently acceptable alternative, it follows that the key question is whether Ofqual and the DoE really understood *and* then fully discussed ahead of time *and* co-owned:

- the way the algorithm locked in the school's previous history,
- the implicit skew in the model that favoured small tuition groups (most often found in the private sector),and
- the forcing of lower grades (including 'U's) on those at the bottom of teachers' ranking.

And, then, most crucially of all, had the Secretary of State been made fully aware of this model output well before results day? It was pretty clear that this was a political nightmare waiting to unfold.

Department-Regulator communications certainly seem to have broken down once the Secretary of State decided that he would no longer rely on the algorithm. I can only imagine the near-panic in both organisations at that time, possibly worsened by interference from No.10. But it didn't look

good, and the supposed decision to use mock A-level results was clearly rushed and ill thought through

Two Lessons?

The FDA's Dave Penman was surely correct when he said that 'I don't think it's fair that civil servants are attacked ... but I don't think ministers should be either. The government needs to find out what went wrong [and learn from it] not make a knee-jerk decision to abandon officials'. But I have two broader suggestions.

First, I suggest that it would be sensible for all departments and their regulators to sit down together to ensure that they have a shared view of which sort of decisions should be taken by the regulator, and which by Ministers. They should also agree on how they would jointly handle criticism of contentious decisions, bearing in mind that agreed division of responsibility.

Second, this episode contains clear lessons for the whole of the public sector. As Stephen Bush has commented: "As politics becomes increasingly dominated by algorithms, what will matter ever more is transparency – about who is writing them, what goes into them, and what they mean for us all."

Martin Stanley

Editor – [Understanding Government](#)

[1] A classic example was the Navy's wartime resistance to the suggesting that convoys might reduce the losses being experienced by merchant ships crossing the Atlantic. (War memoirs of David Lloyd George p1149)

[2] One wonders whether something similar happened in the UK in the early stages of the current COVID-19 pandemic.