Are exams getting easier?

And can Comparative Judgement help us answer this question?

DAISY CHRISTODOULOU

Are exams getting easier? Are grades being inflated? Are the kids of today doing better or worse than ever before?

Too often this debate turns into a slanging match between the generations.

Is there any data we can bring to bear on these questions?

Why monitoring standards over time is hard

Working out the relative difficulty of different exam papers from different eras is actually surprisingly hard to do, for a variety of reasons.

- It's hard to tell just by looking at a question or an exam paper exactly how hard it is.
- Even if you do look at two exams and decide that one is much harder than the other, even that gut instinct judgement doesn't tell you that much, because it may well be that students did very badly on the hard exam and much better on the easier exam.
- So it's not enough to just look at the exam paper in isolation: you need data on how students responded to the paper.
- And when you are comparing exams over time, the content and syllabus often changes - sometimes in ways that make it extremely hard to make a fair comparison.

How Comparative Judgement can help

Is there anyway around these problems? In 2016, my colleague Chris Wheadon and three co-authors published a paper that addressed exactly this issue. It uses some very clever techniques

to solve these problems, and won the 2017 British Educational Research Journal's Editor's Choice award.

They analysed an exam where the content had not changed too dramatically - pure mathematics A-level papers. They used 66 archive responses to exams from 1964, 1968, 1996 and 2012. These archive responses were ones that had been graded at A, B or E. They split each archive response into individual question responses, giving a total of 546 question responses.

They then used Comparative Judgement to assess the quality of each question response. Comparative Judgement is an innovative assessment technique which relies on the fact that humans are much better at making comparative judgements than absolute ones. So instead of looking at individual questions and deciding how hard they were, judges were given pairs of responses and had to say "which student you think is the better mathematician". Here's an example of a comparison.

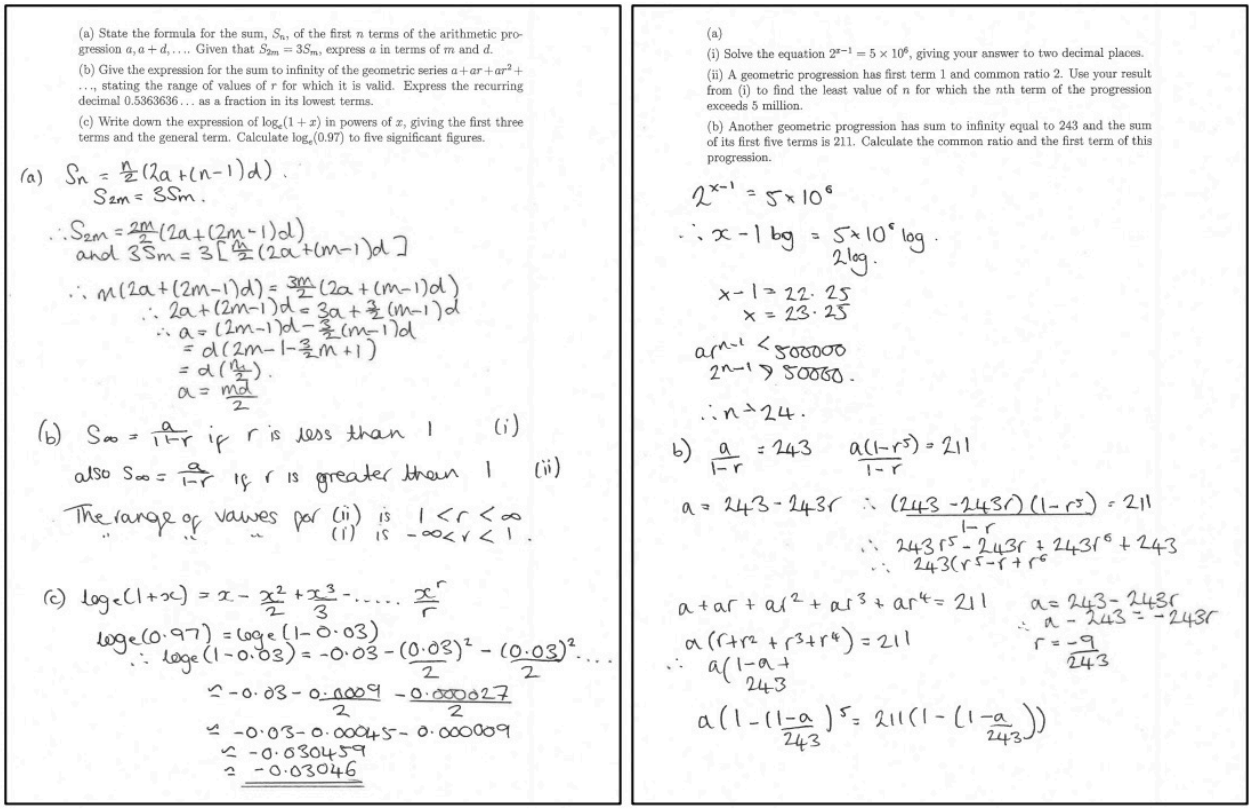


Figure 1: Example pairing of questions and responses. Experts were tasked to decide which candidate was "the better mathematician" of many such pairings.

The judges were all PhD maths students who had passed a maths test in order to be able to judge. The overall reliability of their judging was acceptably high.

At the end of this process, all 546 question responses were given a score, and then all of the original exam papers could be assigned an overall score by aggregating each individual response score.

The results

- We are now able to compare the standards from different years. Here are the key findings.
- Standards did not change much between 1964 and 1968.
- Standards declined between 1968 and 1996. The quality of work that would have produced an E in 1968 would have got a B in 1996.
- Standards did not change much between 1996 and 2012.

So what is happening?

Chris and his co-authors put forward some suggestions as to why this might have happened. I'd like to suggest a reason they don't consider: the expansion in numbers of students taking maths A-level. A-levels in general were taken by about 5-10% of students in the 1960s. Today, about half the cohort take them, and a lot of that increase happened between 1968 and 1996. Keeping the gradeset at the same standard after such an expansion in numbers would mean a large chunk of students failing the exam.

So does this prove kids of today just aren't doing as well as kids in the past?

At the start of this article, I posed three questions.

Are exams getting easier? Are grades being inflated? Are the kids of today doing better or worse than ever before?

These questions are all related, but they are not the same. I think this paper shows that grade inflation, for maths A-level at least, is real. The standard represented by a B grade maths in 1996 was not as difficult as the standard represented by a B-grade 30 years earlier. The currency represented by each grade has been devalued.

In the long run, are we are all doing better?

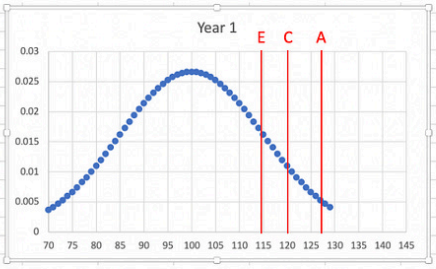
However, this does not necessarily mean that the kids of today are doing better or worse than in the past. To see why, think of an analogy with the economy. The pound is now not worth nearly

as much as it was 100 years ago. But the UK economy has grown enormously in that time and we are much wealthier on average.

Here's a worked example of how this might work for exams.

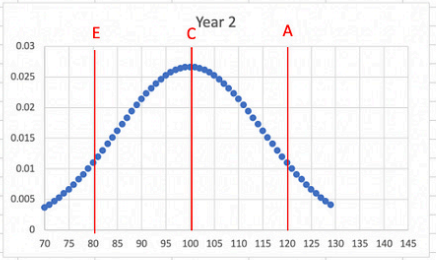
- Imagine an exam that in year 1 has a very tough gradeset - only 15% pass.
- Imagine that in year 2 there is no change to underlying attainment, but the markers decide to change the gradeset and make it less tough. Now all the pupils pass and the standard that would only have got you a C in Year 1 will now get you an A! Textbook grade inflation.
- But now look at Year 3. In Year 3, underlying improvement has improved significantly - by an entire standard deviation. We retain the inflated gradeset of Year 2.

Thus, when we compare Year 3 with Year 1, the following two things can be true: the cohort are much better at maths in Year 3 compared to Year 1, and the grades represent a lower standard in Year 3 compared to Year 1.



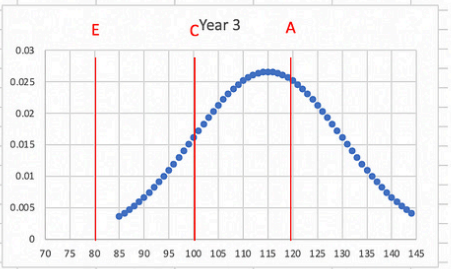
- A-E grades are set as follows

A = 127+
B = 125
C = 120
D = 117
E = 115+



- Underlying attainment is the same as it was in Year 1
- A-E grades are made more generous and are set as follows

A = 120+
B = 110
C = 100
D = 90
E = 80



- Underlying attainment is one standard deviation better than it was in Year 1 & Year 2
- A-E grades are set as they were in year 2

A = 120+
B = 110
C = 100
D = 90
E = 80

The above is a hypothetical worked example. I don't think it is what actually happened. I think it's quite hard to establish what has happened to underlying attainment in the past half century or so, but my best guess is it has probably stayed about the same.

There are other analogies here with economics. Keynesians and Hayekians love to argue about whether inflation boosts or hinders the economy, and you could imagine a similar argument about whether grade inflation motivates or demotivates students.

But I don't want to weigh in on that here. What I want to show with the above is that grades can be very misleading. They are essentially layered on top of underlying attainment scales, and yet very often we take them as direct measures of underlying attainment. We've written in more depth about the distortions this can cause here.

I would like to see some of the bigger national exams reported on a consistent underlying scale, which would make it easier to see what was happening to standards year on year. Comparative Judgement could play a part in making that happen.

Discussion about this post

Write a comment...

...

well it has got easier in all subjects. I looked at what my father did in his O levels took my GCSEs in 1996. I got for example A in math and science. but I struggled with a lot of what my father had to do. plus the grade boundaries where a lot higher. so not only was it harder work/questions but also harder to pass/get a decent grade. those that truly did well stuck out and did well in life.

Now doing my A levels a got B,C, D (and an E in General studies :s) I studied biology, math and chemistry. now I found those exams hard. then a few years latter they brought in modular testing, there where students getting 7, 8, 9 A-levels at A* At this point i had finished my first degree. I got a first (its worth knowing that you only need 40% to pass a degree granted its a 3rd but in all fairness employers don't actually care!)

Fast forward to now, I have helped my son with his GCSEs and A levels the questions where sooo much easier and they are handed far more in help than we where. this is not just isolated to GCSEs, A-levels etc. I did a HNC before doing my second degree as an older student I found it very easy (I was older than most that where teaching the modules) They confirmed many times that they have year on year dumbed down the syllabus, even making comments like you used to have to know how to work this out now you just need to know it exists, so no understanding regarding it.. also the number of