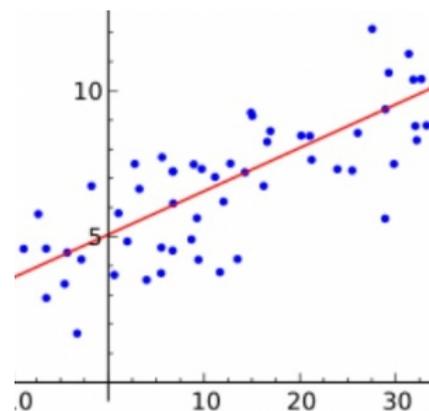


Replication of government research uncovers shaky evidence on relationship between school and degree performance.

LSE blogs.lse.ac.uk/impactofsocialsciences/2014/07/30/replicating-government-commissioned-research/

7/30/2014

Interested in the statistical analysis used to justify the Department of Education's reforms, [Ron Johnston](#), [Kelvyn Jones](#), [David Manley](#), [Tony Hoare](#) and [Richard Harris](#) requested the data related to school performance and degree results via a Freedom of Information request. One year later the dataset was finally made available and they were able to identify some substantial flaws in the government research including sample bias, missing data and poor research design. Rather than confirming that AS level results 'add a negligible degree of accuracy to university performance prediction', the team found that by scrapping AS-levels for examination criteria, 1-in-5 students capable of getting a good degree at the institution of their choice might not be given that chance.



Alongside the Department of Education's push to reform AS-level exams as the first stage of the A-level exam, Minister of State for Schools David Laws commissioned research to form 'part of the evidence base as to whether AS level results are necessary to predict University results, or whether GCSE results alone would suffice'. This involved an in-house statistical analysis of some 88,022 students who graduated in 2011 having finished their A-levels in 2008 and GCSEs in 2006. It sought to predict degree grade – a binary (whether or not students got a 2:1 or 1st-class degree – often termed a 'good degree' and required for entry to many post-graduate degree courses) from the GCSE and AS-level points scores both separately and together (using the standard procedures for converting the grades at each examination into an interval number). The results were published in May 2013 as [A comparison of GCSE and AS level results as a predictor of getting a 2:1 or above at university](#).

The research came to four conclusions, reproduced here in full (for more on the background of the report and the policy decision-making implications, a [summary piece can be found here](#)):

- Neither GCSE or [sic] AS results predict whether a student will get a 2:1 or better with great accuracy (approximately 70% accurate)
- GCSE results are marginally better at predicting whether a student will go on to get a 2:1 or above than AS level results (69.5% accuracy compared to 68.4%)
- The effect of combining GCSE and AS level results adds a negligible degree of accuracy to the predictions
- Without AS level results, we can still predict degree performance to a similar level of accuracy based on GCSE grades alone.

The final conclusion is presumably the evidence base on which the ministers decided to act.

But – seeking replication: obtaining the data

The evidence base appears unproblematic and strong, justifying the proposed change. Or is it? To explore that, we set out to replicate the study. Recognising the government's commitments to both evidence-based policy and open data, in June 2013 we submitted a Freedom of Information request for the data, on the assumption that it could readily be redacted so that confidentiality would not be threatened: all we needed for each of the 88,022 students was their GCSE score, their AS score, the university they attended, and their degree classification. We were told that

it was not possible to release the data then, because it was derived from two separate sources (the department itself, presumably from the National Pupil Data census, and from the Higher Education Statistical Authority). It was hard to see why this would cause delay, since presumably the DfE's statisticians had already combined the two data sets, otherwise how could they have conducted their analysis? However, we were assured that action was being taken to facilitate such requests and it was hoped our application could be acceded to 'soon'.



Image credit: [Testing Times Ahead](#) by [Pete](#) (Flickr, [CC BY 2.0](#))

Following that original response we sent reminders at about monthly intervals. The response was always the same, along the lines of: thank you for your patience, we are still working on the issue and hope to be able to respond in a few weeks. All that was being done was that a form was being produced to allow a single application to be made for data derived from two separate sources – which we were told is legally complex! There was no guarantee that we would gain access to the data – only that we could apply for access! And so at the end of November 2013 we wrote to David Laws asking him to expedite release of the data: weeks later, his reply was a virtual facsimile of what his FoI office was saying to us. (We copied the letter to Kevin Brennan MP, Shadow Minister for Schools, to whom David Laws wrote when the report was published: he never replied.)

Eventually in early 2014 the – lengthy – form appeared and we filled it in. In April we were sent five data sets, but several were undecipherable; there was no data catalogue and we were unable to identify many of the variables (one of the data sets contained 515 variables for 978,344 observations – presumably individual students; another had only 29 variables – but 7,259,538 observations!). We asked for further information; what we got was of little help so we asked if our queries could be directed to the individual(s) who undertook the analysis. In May 2014 they agreed to send us the dataset they prepared and analysed and it arrived in June – one year after our original request. We were able to replicate their analyses exactly, and started to undertake our own, exploring the data further and applying better modelling strategies.

Problems – data and modelling

We encountered a number of issues that led us to query the nature of the analyses undertaken, and so wonder about the veracity of the reported findings.

First, the explanatory commentary accompanying the statistical results is so devoid of any surplus flesh as to be

positively emaciated. The researchers report few caveats regarding their data – too few, in our view. For example, we have to assume that the 88,022 graduates included in the study – covering the entire UK – excluded those doing unclassified degrees (notably medicine). All we are told is that the study excluded those taking four-year degrees, which are concentrated in the sciences and engineering – a potential source of ‘sample bias’ that the researchers do not recognise – and are also the most common form in Scotland across all disciplines. Similarly, they presumably excluded all students who did not take GCSE and AS examinations, which would involve virtually all students at schools in Scotland plus those there and elsewhere who took another route to university entry, such as the IBac. (For example, of the 88,022 students whose results were analysed only 68 were at the University of Edinburgh, 8 at the University of Stirling and 5 at the University of Strathclyde; the data are clearly unrepresentative of Scottish universities which graduate very many more each year! Similarly, Imperial College London, a predominantly science and engineering institution, has only 236 graduates in this analysis: the majority of its graduates in 2011 were presumably either medical students or on four-year degree programmes.)

HESA data indicate that there were some 350,000 first degree graduates in the UK in 2011, so the issue of the representativeness of the data deployed in the research is important. No data set is perfect, but analysts should recognise the full limitations of those they deploy, since any selectivity in the population analysed can potentially impact on the validity of the results.

A further problem is that of the 88,022 students in the data base, 1,506 have a zero score for their GCSE performance – suggesting either that they did not take the exams in 2007 or that they never did – and 1,442 have a zero score for their AS-level performance in 2008; 25 had a zero score on both. As far as we can tell, these 2,923 were almost certainly included in the DfE statistical analyses. Further, as well as a variable indicating whether the students got a ‘good degree’ or not there is another giving further details on their degree classification – and for 5,013 there is none. Again, as far as we can tell, these 5,013 – like the 2,923 – were included in the DfE analyses despite the problem of missing data: as around 300 of those for which there was no information on degree classification also had no data on GCSE and/or AS-level scores, it seems that those analyses included a little under 8,000 observations for which at least one of the key items of information was missing (i.e. some 9 per cent of the total)!

The nature of the statistical analyses is also problematic. It involved binary logistic regressions; the dependent variable was the degree classification (scored as 1 if the student obtained a 2:1 or better degree and 0 otherwise). The independent (predictor) variables were GCSE and/or AS score plus a series of dummy variables, with each of 151 separate institutions attended by students within the sample coded either 0 or 1 depending on whether the student graduated there. Manchester – with the largest number of graduates in the dataset – was used as the comparator in these ‘fixed-effect’ computations; the regression evaluated the probability of all of the other students getting a 2:1 or better in each separate institution compared with a Manchester student with the same GCSE and/or AS score.

This simple research design is flawed, not least because – as already noted – the 151 institutions varied substantially in their number of graduates in 2011. The DfE researchers note that because ‘degrees are not standardized, and universities have different standards of intake, it would not be fair to compare two universities based on their coefficients in this study’ [so much for the vaunted external examiner system which is supposed to ensure consistent standards!]. Looking at those reported coefficients raises substantial concern, however. Is it really likely to be the case, for example, that a student at Imperial College with a particular GCSE score is 40 per cent less likely to get a 2:1 or better than a similar student at Manchester, or that of two students with the same GCSE score one attending either Leeds College of Music or the University of Exeter is 2.3 times more likely to get a ‘good degree’ than a comparable Manchester student? Indeed, the regression coefficients for some of the individual institutions can only be described as nonsensical (even if statistically insignificant) – such as -21.347, -21.666, -21.192 and -21.408 for four small Scottish institutions by number of graduates in the survey (plus -21.593 for Birkbeck College and -22.563 for the Institute of Education); all of those coefficients have standard errors over 10,000(!) in the GCSE analysis.

Many, but not all, of the apparent anomalies apply to small institutions – and the research design gives them equal weight to the larger places, providing a misleading portrayal of the university sector as a whole. So has the DfE’s overall finding – that they can predict degree performance with about 70 per cent accuracy – much value at all, let alone as evidence on which to base policy?

Starting a re-analysis

The correct research design for such a data set would reflect its hierarchical structure – of students nested within institutions: the most commonly-used relevant technique (indeed, it was developed for just such applications in educational research) is multi-level modelling (MLM).

Initial replications using MLM – omitting the students for whom there is either no GCSE or no AS-level score, or no degree classification, and thus reducing the total to 80,420 – largely confirm the DfE’s findings with regard to predictability: using GCSE scores plus institution we correctly predicted the degree classification for 69.7 per cent of the students, and using institution plus AS-level score the percentage correctly predicted was 68.7. This apparently bears out the Department’s conclusion that degree result can be as accurately predicted by the former as by the latter. At first sight, therefore, abolition of AS-levels should not impede universities in making their offers of places on the basis of expected degree result.

But further exploration of the data threw up a major problem. We anticipated a close relationship between GCSE and AS-level performance – that students who performed well at the first exams would do well again at the second a year later – but that was not the case. An r-squared value of only 0.38 indicated that over three-fifths of the variation in AS-level performance could not be successfully predicted by performance at GCSEs only a year previously. This suggested that there may be a similar poor relationship between predicted degree performance based on GCSE scores and that based on AS-level scores.

To explore whether that was so, we re-ran the DfE’s models. We derived the probability that each student would get a 2:1 or 1st according, separately, to their GCSE and AS-level performance (with the institutional dummy variables). We then regressed those two sets of probabilities. The r-squared value was 0.57 – only just over half of the variation in one could be accounted for by variation in the other. This meant that, if universities were to make their offers on the basis of predicting which students would get a ‘good degree’ from their GCSE results their lists would include many who would not be included in lists based on their AS-level scores – and vice versa. How many? The following table compares those two sets of probabilities – for those students for whom we have data on all of the variables.

Table. The probability of getting a ‘good degree’ according to students’ GCSE and AS-level performance

GCSE	AS					
	0.00-0.24	0.25-0.39	0.40-0.49	0.50-0.59	0.60-0.74	0.75-1.00
0.00-0.24	95	91	44	24	10	0
0.25-0.39	41	2,036	1,727	1,306	671	95
0.40-0.49	4	1,666	3,025	2,950	2,306	352
0.50-0.59	1	819	2,693	3,986	4,635	1,099
0.60-0.74	1	353	1,965	4,369	9,953	5,756
0.75-1.00	0	36	325	1,312	6,312	20,245

The statistical procedure used by the DfE analysts classifies a student as likely to get a 2:1 or 1st if the predicted probability exceeds 0.5. The first row of the table refers to those for whom that predicted value, based on GCSE performance and institution attended, is less than 0.25. The columns refer to the probabilities based on AS-level performance plus institution attended. Of the 264 in that row – those least likely to get a ‘good degree’ according to their GCSE results – 34 would have performed that well if the prediction was based on their AS-level results instead

(i.e. those in the three right-hand columns, shown in bold); some 12.9 per cent of the degree outcomes are predicted differently by the two sets of school exam results. In the next row, the difference is 35.3 per cent – of the 5,876 predicted to get a 2:2 or lower according to their GCSE grades, 2,072 (those shown in bold) would have achieved a 2:1 or better according to their AS-level results. The differential is even greater in the next row; of those who would have missed a 2:1 or first according to their GCSE results (i.e. with a probability of 0.40-0.49), fully 54.4 per cent would have got one of those degree grades according to their AS-level performance.

The lower three rows of the table show those predicted to get a 2:1 or 1st according to their GCSE results; the figures in bold identify those who would not have done so if the prediction was based on AS-level performance alone. Of those who would have just made it according to their GCSE scores (probabilities of 0.50-0.59), 26.8 per cent would not if their AS-level performance had been the basis of the prediction: for the next two rows – those very likely to get a ‘good degree’ according to their GCSE results – the relevant percentages are 10.4 and 1.3. Having performed well enough at GCSE that they would probably get a ‘good degree’, many fewer would be pulled down by their AS-level results compared to the many not predicted to do well by their GCSE performance but who would have done based on their AS-levels; more students improve their performance over time than fall back.

Overall, for 18.5% of the students whether or not they would probably get a 2:1 or 1st depends on which set of prior exam grades is used. If – as Michael Gove and David Laws have ensured – AS-levels had been abolished by 2011 as exams that university admissions officers could take into account then, according to these figures, 7,714 of the 80,321 students for whom we made predictions might not get an offer on the basis of their GCSE performance because they were not expected to get a ‘good degree’; yet their improvement at AS-level indicates that they probably would. Is a policy that makes so many ‘mistakes’ a desirable one? If GCSE results are the sole guide then 1-in-5 students capable of getting a good degree at the institution of their choice according to their AS-level results might not be given that chance to fulfil their potential?

A complementary article on the policy implications of this research can be found at our sister blog [British Politics and Policy](#).

Featured and top left image credit: Linear regression by [Sewaqu](#) (Wikimedia, Public Domain)

Note: This article gives the views of the authors, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Authors

Ron Johnston is a Professor in the School of Geographical Sciences at the University of Bristol.

Kelvyn Jones is Professor of Quantitative Human Geography at the University of Bristol (since 2001), being Head of the School of Geographical Sciences from 2005-2009.

Dr David Manley is a Lecturer at the University of Bristol.

Dr Tony Hoare has been on the academic staff at the University of Bristol since 1976.

Dr Richard Harris is a Reader in Quantitative Geography at the University of Bristol.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.