This is a single article from *Research Matters: A Cambridge Assessment publication*. http://www.cambridgeassessment.org.uk/research-matters/
© UCLES 2018

Ofqual. (2016). *Summer exam entries: GCSEs, Level 1/2 Certificates, AS and A levels in England. Provisional figures April 2016*. Coventry: Office of Qualifications and Examinations Regulation.

Ofqual. (2017). *Provisional summer 2017 exam entries: GCSEs, AS and A levels*. Coventry: Office of Qualifications and Examinations Regulation.

Porter, N. (2015). *Crossing the line. Improving success rates among students retaking English and Mathematics GCSEs. A Policy Exchange Policy Bite*. London: Policy Exchange.

Smith, J.A. & Todd, P.E. (2005). Does matching overcome LaLonde's critique of non-experimental estimators? *Journal of Econometrics, 125*(1–2), 305–353.

Ward, H. (2017). DfE adviser calls for maths GCSE resits U-turn. *Times Education Supplement*. (2017, May 26). Retrieved from https://www.tes.com/news/school-news/breaking-news/exclusive-dfe-adviser-calls-Mathematics-gcse-resits-u-turn.

Wolf, A. (2011). *Review of Vocational Education – The Wolf Report* (DfE-00031-2011). London: Department for Education.

How many students will achieve straight grade 9s in reformed GCSEs?

Tom Benton Research Division

Introduction

"It's difficult to make predictions, especially about the future" (Danish proverb)

As General Certificate of Secondary Education (GCSE) qualifications are reformed in England, the grading scale is changing from students being awarded grades A*-G to being awarded grades 9–1, with grade 9 representing the highest grade and also relating to a level of achievement above that of the existing grade A*. This process began in practice in summer 2017 when Mathematics, English Language, and English Literature GCSEs were awarded on the new grading scale. The majority of subjects with large entries will be switching to the new grading scale as part of awarding in summer 2018 and the remainder will be switching in summer 2019[1].

This article attempts to predict the number of pupils who will achieve a perfect set of grade 9s in whichever reformed GCSEs they choose to take. This question sprang to prominence in the media in April 2017 when Tim Leunig, the then chief scientific advisor of the Department for Education (DfE), tweeted that he expected only two pupils to achieve grade 9 in all of their GCSEs. This led to contact between the TES and Cambridge Assessment and, subsequently, to the author giving his own alternative view that 'hundreds' of pupils will achieve grade 9 in every GCSE that they take[2]. For the remainder of this article we will refer to this accomplishment as achieving 'straight' grade 9s.

This article gives more details of how such a prediction might be made. As well as the evident interest in this question externally, it may be of substantive importance as it relates to the extent to which reformed GCSEs, and grade 9 in particular, will be able to discriminate between the very highest performing students.

Since making the original forecast of 'hundreds' of pupils to achieve straight grade 9s in April 2017, more information about attainment in reformed GCSEs has been published by both The Office of Qualifications and Examination Regulation (Ofqual)[3] and the DfE[4]. Naturally, this article makes use of this later information but the rationale is the same as for the earlier predictions. Note that, at the time of writing, the latest national pupil level data available to the author dates from summer 2016.

One method of making the prediction would be to retrospectively set the grade 9 boundary in all existing GCSEs using the formula used to define how many should achieve grade 9 in each subject (see Benton, 2016). It would then be a simple task to just count how many pupils actually attained notional grade 9s in all of the GCSEs they had entered. However, it was not possible to access the raw marks achieved by pupils on a national level, and the techniques employed in this article are entirely based upon data regarding the grades achieved by pupils.

Some simple methods of estimation

To begin with, we consider a very simple way to estimate the number of pupils who will achieve straight grade 9s to illustrate how it might be possible to reach a prediction of around two pupils. The first step is to consider the number of students who achieved straight grade A* in all of their GCSEs historically. Based on Gill (2017), who provides numbers based on students taking at least 5 GCSEs in June 2015, this value might be taken to be 3,300. Next, using an early proposal for the definition of grade 9 (Ofqual, 2014, p.20), we might assume that in every GCSE, around half of those awarded grade A* would be awarded grade 9. Thus, we might assume that, amongst those achieving straight grade A*s, half of these would fail to achieve grade 9 in the first GCSE we consider. This leaves just 1,650 candidates. Applying the same idea to the second GCSE again reduces the number by half to 825 pupils. If we continue with this process of halving the values until we reach 10 GCSEs, then we end up with a prediction of just 3 pupils to achieve straight grade 9s.

However, there are a number of flaws in the above calculation. Firstly, in each subject, the percentage of candidates who will be awarded grade 9 as a percentage of those who would have been awarded grade A* is a little higher than 50 per cent. It varies between subjects, as the percentage who will be awarded grade 9 is tied to the percentage historically awarded grade A or above rather than A* (see Benton, 2016).

1. https://www.gov.uk/government/publications/get-the-facts-gcse-and-a-level-reform/get-the-facts-as-and-a-level-reform
2. https://www.tes.com/news/school-news/breaking-news/exclusive-major-exam-board-predicts-hundreds-will-get-straight-grade
3. https://www.gov.uk/government/news/guide-to-gcse-results-for-england-2017
4. https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2016-to-2017

 © UCLES 2018

However, on average it is slightly over 60 per cent. Secondly, the logic ignores the possibility that the more grade 9s you have already achieved, the more likely you are to get the next one. For example, although we might only expect 60 per cent of those who have achieved the equivalent of grade A* or above in an individual subject to be awarded grade 9, the percentage of candidates who will be awarded grade 9 out of those who have achieved the equivalent of grade A* in this subject *and achieved grade 9 in all of their other subjects* should be somewhat higher. Taking account of this fact is crucial if we are to make an accurate prediction. Finally, the calculations in the preceding paragraph assume that all of the students we are interested in took 10 GCSEs. In reality, the number of GCSEs taken will vary between candidates.

One way to account for the correlations between achievement in different subjects, and thus the fact that those getting grade 9 in some will be more likely to get grade 9 in others, is to assume that all candidates have an underlying level of general ability that influences their achievement in all of the GCSEs that they take. This idea has been prevalent in psychometrics for more than 100 years (see Spearman, 1904) and has previously been used to help analyse possible differences in difficulty between subjects (Coe, 2008). The idea is that each person has an, unmeasured, level of ability from somewhere on the normal distribution that influences their achievement in any assessment. In theory this ranges from minus infinity for people with zero chance of answering any question to plus infinity for anyone who is all-knowing. However, for practical purposes, nearly all people would be within the range -4 to +4 and we focus on working on the probability of people in this range getting straight grade 9s in all their GCSEs. Although there are some weaknesses in this approach that will be described later, the calculations are simple enough to be performed using no software more complicated than Microsoft® Excel, and will also serve to illustrate some of the difficulties involved with predicting how many pupils will get grade 9 in all their GCSEs. The details of the calculation steps are shown in Table 1. Where applicable, the Excel formulae that were used to complete calculations are shown.

To begin with, we specify the percentage of candidates we would expect to achieve grade 9 in any subject if the given subject was taken by every eligible student in the country. This percentage is set to be 3.1 to match the average percentage awarded grade 9 across the three reformed GCSEs awarded in summer 2017[5]. Next, we find the equivalent cut point in the normal distribution (1.87). This means that, if nationally available raw marks from each GCSE were transformed to a scale with a standard normal distribution with a mean of 0 and a standard deviation of 1, then the grade 9 boundary would be located at this point.

5. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639824/GCSE_results_2017_infographic.pdf.

Table 1: Calculation steps based on assuming a single underlying latent trait

				Percentage of candidates who will get grade 9 or above in each subject nationally *(pc9)*	3.10%
				Normalised cut-off for grade 9 *(cutoff)* (=NORMINV(*pc*9,0,1))	1.87
				Correlation between general and specific abilities *(correl)*	0.75
A. Normalised general ability	*B. Weight*	*C. Expected specific normalised score (=A x* correl*)*	*D. Standard deviation of score given general ability (=SQRT(1-*correl*^2))*	*E. Probability of getting grade 9 in any one subject (1-NORMDIST(*cutoff,C,D,TRUE*))*	*F. Probability of getting 10 x grade 9s (=E^10)*
4	0.01%	-3	0.66	0.00%	0.00%
-3.5	0.04%	-2.63	0.66	0.00%	0.00%
-3	0.22%	-2.25	0.66	0.00%	0.00%
-2.5	0.88%	-1.88	0.66	0.00%	0.00%
-2	2.70%	-1.5	0.66	0.00%	0.00%
-1.5	6.48%	-1.13	0.66	0.00%	0.00%
-1	12.10%	-0.75	0.66	0.00%	0.00%
-0.5	17.60%	-0.38	0.66	0.04%	0.00%
0	19.95%	0	0.66	0.24%	0.00%
0.5	17.60%	0.38	0.66	1.21%	0.00%
1	12.10%	0.75	0.66	4.57%	0.00%
1.5	6.48%	1.13	0.66	13.12%	0.00%
2	2.70%	1.5	0.66	28.99%	0.00%
2.5	0.88%	1.88	0.66	50.52%	0.11%
3	0.22%	2.25	0.66	71.91%	3.70%
3.5	0.04%	2.63	0.66	87.43%	26.10%
4	0.01%	3	0.66	95.67%	64.26%
Total expected to get straight grade 9s					0.0248%
Expected number to get straight grade 9s out of 500,000					124

The next part of the calculation requires us to specify the expected correlation between each student's underlying general ability and their normalised scores in each individual subject. This figure is chosen as 0.75 as previous research has shown that for large scale GCSEs a typical correlation between subject-grade and mean GCSE grade is 0.75 (Benton & Sutch, 2013, p.7, Table 2). As such, it provides a reasonable idea of the link between general ability and specific ability in particular subjects.

To complete calculations, we assume that underlying general ability follows a normal distribution nationally. Then, for each possible level of underlying ability (column A in Table 1) we can calculate:

- the proportion of candidates we expect to have this level of ability (column B);
- the expected normalised score of candidates with this ability in each individual subject (column C) and how much uncertainty there is in scores in specific subjects given general ability (column D);
- using C and D, the probability of the candidate getting grade 9 in an individual subject given their level of general ability (column E); and
- the probability that this achievement of grade 9 will be repeated across 10 different GCSEs (column F). This is calculated as just column E to the power of 10, the assumption being that *given candidates' underlying ability* achievement in separate GCSE subjects is independent.

Note that, although in reality we expect general ability to form a continuum, this is approximated by just 17 points on this scale between -4 and +4. This method of approximation, known as quadrature, is commonly used within psychometrics and the weights shown in column B are just set to be proportional to the density of the standard normal distribution, but also to sum to 100 per cent.

By taking a weighted average (weights in column B) of the values in column F we can estimate that less than 0.03 per cent of candidates (that is, less than 3 in 10,000) would be expected to achieve straight grade 9s across 10 GCSEs. If we imagine a GCSE cohort of 500,000 candidates this would mean that just over 100 of them would achieve straight grade 9s.

There are a number of flaws in these calculations, but before discussing these it is worth noticing the values in column F. For example, it is interesting to note in these calculations that even a candidate with an underlying general ability of 2.5, which would be enough to place them in the top 1 per cent of performers, still has a vanishingly small chance of achieving straight grade 9s across 10 subjects. In fact, nearly all of the candidates predicted to achieve straight grade 9s come from the final three rows of ability – that is candidates at or above the 99.9th percentile of general ability. It is very rare that calculations in education need to focus upon such extreme values. As such, the predicted numbers of candidates to achieve straight grade 9s are very sensitive to the assumptions underlying the model.

Table 2 begins to show some of this sensitivity. Moving from top to bottom allows us to see the impact of the assumed number of GCSEs on calculations. As can be seen, the more GCSEs are taken by candidates, the fewer the number of candidates we expect to achieve straight grade 9s. After all, it is easier to get straight 9s in 8 subjects than in 10. In the first row, the number expected to get straight grade 9s across 3 subjects is included as this allows direct comparison with the 2,050 candidates known to have achieved straight grade 9s in all of Mathematics, English Language and English Literature GCSEs in 2017[6].

Moving from left to right shows the impact of the assumed correlation between individual GCSEs and general ability. If this level of correlation were to drop in reformed GCSEs from its historical level to a substantially lower value of 0.7, then our prediction of the number of straight grade 9 candidates would decrease by nearly two-thirds. In contrast, if it were to increase to 0.8, then the number of straight grade 9 candidates would nearly double. In fact, analysis of data collated from Ofqual's web analytics page[7] suggests that the correlation between English Language and Mathematics grades in reformed GCSEs is close to 0.7 which would imply a correlation of both with underlying general ability of around 0.8 (the square root of 0.7). With this in mind, it is of no surprise that it is this value that gives the closest match to the known actual number of straight grade 9s for the three reformed GCSEs in 2017 (i.e., the published number of 2,050). However, this can not necessarily be taken to imply that this value is the most appropriate one for predictions into the future.

Table 2: Predicted number of candidates who will achieve straight grade 9s from a cohort size of 500,000 depending upon correlations between general ability and individual subjects and the number of GCSEs taken

No. of GCSEs being taken by each candidate	*Expected correlation between general ability and scores in individual GCSEs*		
	0.7	*0.75*	*0.8*
3	1073	1560	2260
...	...	...	...
8	82	196	440
9	60	154	365
10	45	124	309

Although interesting, these calculations still have some weaknesses. For example, they assume that all candidates take the same set of subjects, thus ignoring the impact of subject choice on candidates' likely achievement (see Benton & Bramley, 2017). They also assume a common correlation between all subjects and general ability rather than noticing that some subjects (e.g., English Language and English Literature) are more strongly correlated to each other than to others. Finally, they again assume that all candidates take the same number of GCSEs. For these and other reasons, and in order to improve accuracy, we move from predictions based almost entirely on theory to predictions built directly from pupil level data.

Empirical estimates based on data from 2016

The data set

The data for analysis was taken from the National Pupil Database (NPD), which is held by the DfE and consists of results for all students in all subjects in schools and colleges in England. The analysis focussed upon the GCSE results of all candidates in Year 11 in the academic year 2015/16. This was the most recent set of national data available to the

6. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/639824/GCSE_results_2017_infographic.pdf.
7. https://analytics.ofqual.gov.uk/apps/2017/GCSE/9to1/

© UCLES 2018

author at the time of writing. International GCSEs were counted as GCSEs in the analysis as they were a common alternative qualification that is judged to be equivalent (Gill, 2017). However, the NPD only included the results of international GCSEs which had been accredited for use in state schools. There were also several non-accredited international GCSEs which were taken by some students attending independent schools and which, therefore, were not in the NPD. As such they could not be counted in the analysis in this article.

For this particular cohort of students, early entry was fairly common in some subjects such as English Language and Mathematics. To ensure that the cohort of students with results in Mathematics, English Language, and English Literature used in our analysis was of roughly the same size as the cohort taking the three reformed subjects in 2017, it was necessary to include early entries. However, if there were students with multiple entries, rather than simply take each student's highest grade in each subject, results within GCSEs were taken in preference to results in international GCSEs (as this article is really concerned with GCSEs), and results from June 2016 were taken in preference to any earlier results. Entries to the (now discontinued) combined GCSE English Language and Literature were counted as if they were English Language entries.

In order to restrict the analysis to a manageable number of GCSE subjects, our analysis first looked at the most common subject choices amongst students who achieved straight grade A* grade across at least five different GCSEs. Table 3 shows all GCSE subjects taken by at least 40 straight grade A* candidates. Subsequent columns show the percentage of straight grade A* candidates taking each subject compared to the percentage of all candidates taking each subject (of those who took at least five GCSEs). There are some very large differences in subject popularity between these two groups. For example, whilst more than half of all students chose to study Combined Science (Science [Core] and Additional Science), only just over one in twenty straight grade A* candidates chose these subjects, whilst the popularity of the separate sciences (Biology, Chemistry, Physics) was much higher amongst the straight grade A* candidates. Both Ancient and Modern Languages were far more popular amongst the group of candidates who achieved straight grade A*s than amongst GCSE candidates as a whole. Indeed, it is notable than GCSE Chinese, taken by less than 0.5% of candidates nationally, was taken by almost 7% of straight grade A* candidates. Similar comments, though to a slightly lesser extent, could be made about both GCSE Russian and GCSE Italian. The biggest differences of all can be seen for GCSE Latin and GCSE Classical Greek taken by around a third and a tenth of straight grade A* candidates respectively compared to around 1% and 0.2% of candidates nationally. To illustrate this further, the final column of Table 3 shows that almost 40% of *all candidates* who entered GCSE Classical Greek actually achieved straight grade A*s across at least 5 GCSE entries.

Analysis was restricted to candidates taking at least one of the subjects in Table 3 and at least three GCSEs in total. Having made these choices, some descriptive information about the data set used for analysis is shown in Table 4. In particular, for those candidates with results in all of Mathematics, English Language, and English Literature, some comparisons are made to data published by Ofqual after summer 2017. After making this comparison, there was some concern over the fact that the percentages of candidates achieving grade A (equivalent to grade 7 in 2017) and above in these subjects (and in Mathematics in particular) were larger in the analysis data set used in this article than in Ofqual's published data from summer 2017. The reasons for these differences are not known. Potentially they may relate to decisions within high-performing independent schools to continue to use unreformed international GCSEs rather than switch to using reformed GCSEs at this stage. However, regardless of the reasons for the differences, to improve the comparability of the two data sets, 8,000 candidates who had taken all three subjects of interest and had achieved grade A in Mathematics were randomly selected for removal from the data set. As shown by the final column of Table 4, although this step reduced the overall number of candidates used in analysis, it ensured that the number of high-performing candidates in each subject was closer to that within Ofqual's published data.

Although we cannot assume this completely removed the differences in the characteristics of candidates in the two data sets, it should help to ensure we make valid comparisons. It should be noted that if these candidates were added back in to analysis, the later predictions of the number of candidates to achieve straight grade 9s would increase.

Table 3: The most popular GCSE subjects taken by candidates achieving straight grade A*s across at least five GCSEs in 2016

GCSE subject	*No. of straight grade A* candidates*	*% of straight grade A* candidates taking subject*	*% of all candidates taking subject*	*% of all subject entrants that get straight grade A*s*
English Literature	3,421	88.6%	93.9%	0.7%
Biology	2,386	61.8%	26.3%	1.7%
History	2,328	60.3%	45.6%	0.9%
Physics	2,225	57.6%	25.9%	1.6%
English Language	2,215	57.4%	92.6%	0.4%
Chemistry	2,164	56.1%	25.9%	1.5%
French	2,151	55.7%	26.0%	1.5%
Geography	2,029	52.6%	42.2%	0.9%
Mathematics	1,905	49.4%	95.7%	0.4%
Religious Studies	1,651	42.8%	48.0%	0.6%
Latin	1,325	34.3%	1.2%	20.3%
Spanish	1,161	30.1%	16.6%	1.3%
German	807	20.9%	9.3%	1.6%
Music	566	14.7%	7.5%	1.4%
Classical Greek	363	9.4%	0.2%	39.9%
Art & Design (Fine Art)	352	9.1%	8.8%	0.7%
Art & Design	325	8.4%	14.2%	0.4%
Computing	291	7.5%	11.3%	0.5%
Chinese	268	6.9%	0.6%	8.4%
Drama & Theatre Studies	251	6.5%	12.1%	0.4%
PE/Sports Studies	230	6.0%	20.5%	0.2%
Science (Core)	226	5.9%	69.6%	0.1%
Additional Science	221	5.7%	63.2%	0.1%
D&T: Resistant Materials	177	4.6%	8.4%	0.4%
Statistics	162	4.2%	8.8%	0.3%
Italian	162	4.2%	0.8%	3.9%
ICT	161	4.2%	13.6%	0.2%
Business Studies	134	3.5%	13.4%	0.2%
Russian	127	3.3%	0.3%	6.9%
D&T: Product Design	85	2.2%	6.6%	0.2%
Classical Civilisation	84	2.2%	0.6%	2.5%
Economics	81	2.1%	1.7%	0.9%
D&T: Textiles Technology	66	1.7%	3.9%	0.3%
Astronomy	60	1.6%	0.3%	3.2%
D&T: Food Technology	60	1.6%	6.0%	0.2%
D&T: Graphic Products	47	1.2%	4.7%	0.2%

Table 4: A comparison of the analysis data to figures published by Ofqual for results from summer 2017

Statistic	*Ofqual data (2017)*	*Initial analysis data set (2016)*	*Final analysis data set (2016)*
Total number of pupils	-	574,879	566,879
No. of pupils with grades in all of Maths, English Language, and English Literature (3-subject candidates)	508,950	506,226	498,226
No. of 3-subject candidates achieving grade A/7 or above in Maths	102,950	111,061	103,061
No. of 3-subject candidates achieving grade A/7 or above in English Language	84,750	89,311	85,292
No. of 3-subject candidates achieving grade A/7 or above in English Literature	96,050	98,121	93,883

Having collated the data for analysis, three different methods were applied to attempt to make predictions of how many candidates would achieve straight grade 9s. In each case, predictions of how many of the candidates who had taken all of Mathematics, English Language, and English Literature we would expect to achieve straight grade 9s *in these subjects only* were also produced so that these could be compared to the known results from summer 2017. Finally, as it may be relevant to performance tables, predictions were also made for how many candidates we would expect to achieve a perfect score in their *Attainment 8*[8] measure used for accountability.

Prediction using uni-dimensional item response theory method

The first set of predictions were made using a very common method with psychometrics – that of item response theory (IRT). As with the last set of theoretical calculations shown earlier, the method assumes that all of the relationships between achievements in different subjects can be explained by a single underlying latent trait (or general ability). However, the form of IRT model we used (the graded response model) allows for the fact that some subjects may be more strongly related to this underlying general ability than others (perhaps such as Art) which may consist of more specific skills. In addition, it is not assumed that, if all students took all subjects, then the grade distributions would be the same in every one of them.

Specifically, we define Y_{ij} as the grade achieved by the ith student in the jth GCSE subject being analysed. Then the probability that they achieve grade k or above is calculated as:

$$P(Y_{ij} \geq k) = \frac{\exp(\alpha_j\theta_i - d_{jk})}{1 + \exp(\alpha_j\theta_i - d_{jk})}$$

In this notation θ_i is the ability of student i. Abilities are defined to follow a normal distribution nationally with a mean of 0 and a standard deviation of 1. The α_j parameter defines the strength of the relationship between general ability and the grades achieved within subject j. In this notation, the d_{jk} (or difficulty) parameters specify the log-odds of an average ability student achieving below grade k. The values of d_{jk} increase for higher grades as, obviously, the chances of getting, say, grade A or above are lower than the chances of getting grade B or above.

The α_j and d_{jk} parameters for each subject were estimated using the national data set of GCSE results from 2016 described earlier and the R software package mirt (Chalmers, 2012). Although it is possible to also produce direct estimates of each individual candidate's ability (θ_i), such estimates are unlikely to properly reflect the full national distribution of candidate ability. For this reason, the method of plausible values is used instead (Wu, 2005). This method is commonly used in the analysis of large-scale international surveys such as the Programme for International Student Assessment (PISA). Rather than trying to get the most accurate estimate of ability for each individual student, these values are chosen for each student to be within the likely range of their true values given their GCSE grades in each subject, but also so that they are likely to have the correct distribution across the population as a whole (Marsman, Maris, Bechger, & Glas, 2016).

The combination of item parameters and (plausible) ability estimates allow us to simulate the likely grades of any students in any combination of GCSE subjects. However, so far, because grade 9 had not been defined in 2016, such simulations only go up as far as grade A*. In order to go beyond this, one more step is required – the imputation of *plausible marks*.

One way to view the graded response IRT model, as previously defined, is that each student's achievement in any particular GCSE subject follows a logistic distribution centred around their scaled ability $\alpha_j\theta_i$. Specifically, we might define M_{ij} as some monotonic transformation of the marks that candidate i achieves in subject j. Then, we can say that:

$$M_{ij} = \alpha_j\theta_i + \varepsilon_{ij}$$

where ε_{ij} is the candidate's overachievement in subject j relative to their general ability and the values of ε_{ij} in any two subjects are independent. We also define the ε_{ij} to follow a logistic distribution. Finally, we use the values of d_{jk} to be the grade boundaries on this mark scale so that a candidate's grade is greater than or equal to k if $M_{ij} \geq d_{jk}$.

Using this formulation, it is possible to simulate plausible marks for all candidates in all of the subjects that they have taken. Simulated plausible grades can be created by using the d_{jk} parameters as grade boundaries. Note that, since we are working from simulated marks, each candidate's simulated grade may not match their actual grade in each subject. Next, we identify a grade 9 boundary on this newly defined mark scale. In order to do this, we first calculate the percentage of candidates that achieved grade A or above in the subject. Now, according to the definition of grade 9, the percentage of candidates who achieve grade 9 is derived from the percentage achieving grade A or above using the tailored formula specified in Benton (2016). Once this percentage is calculated, we simply identify the location of a grade 9 boundary on the simulated mark scale so that the percentage of candidates achieving grade 9 matches the intended number.

With simulated marks and grade boundaries in place it is then possible to calculate exactly how many students achieved straight grade 9s given their simulated marks and this provides the basis of our main prediction.

However, some initial checks on these predictions led to some

8. Attainment 8 is calculated by looking at each candidate's grade in Mathematics, their best grade from English Language and English Literature, the best three grades they achieve in any EBacc subjects (Science subjects, Computer Science, History, Geography, and Languages), and their best three grades from GCSEs not already used within previous categories. For further details see https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/676184/Secondary_accountability_measures_January_2018.pdf.

© UCLES 2018

concerns over their accuracy. For example, when used to simulate grade A*s (i.e., before the grade 9 boundary had been defined) the method predicted that more than 6,300 students would achieve grade A* in all of Mathematics, English Language, and English Literature, when in reality fewer than 5,900 did so. Of slightly greater concern was the fact that the model predicted that more than 2,800 candidates would achieve grade 9 in these same 3 subjects when statistics published by Ofqual have revealed that only 2,050 candidates achieved this in summer 2017.

Attempts to correct these overestimates by fitting more complex, multi-dimensional IRT models were unsuccessful. For this reason, an alternative method that was not reliant on IRT was attempted.

Prediction using plausible marks from logistic regression and actual grades

Rather than relying on a particular IRT model, an alternative was to attempt to derive plausible marks using logistic regression. The first step was to model each candidate's probability of achieving a grade A* in each GCSE subject as a function of their mean GCSE grade in all of their other GCSEs. The formula for logistic regression defines the probability of achieving a grade A* as:

$$P(Y_{ij} \geq A^*) = \frac{\exp(\beta\, MeanGCSE_i\ \alpha_j)}{1 + \exp(\beta_j MeanGCSE_i - \alpha_j)}$$

As can be seen, this formula is very similar to the one for the graded response model outlined earlier. However, rather than relying on a latent variable (θ) that is assumed to follow a normal distribution, probabilities are modelled based on an observed variable (mean GCSE grade). The β_j parameters defined the strength of the relationship between grades in a particular GCSE and the mean grade in other GCSEs. Only a single intercept parameter (α_j) is defined as this model just focusses upon grade A* as this is the most informative existing grade for the research question being studied.

The fitted logistic regression model can be used to produce plausible marks in each GCSE as before using the equation:

$$M_{ij} = \beta_j MeanGCSE_i + \varepsilon_{ij}$$

where ε_{ij} is the candidate's overachievement in subject j relative to mean GCSE grade in other subjects and is simulated from a logistic distribution. However, to improve upon this method, we make a further amendment so that for each candidate their simulated plausible marks will be consistent with the actual grade they achieved. This method is illustrated in Figure 1. The top two panels show the possible score distributions from standard simulation for two candidates who, in reality, both achieved grade A* in Mathematics. The candidate on the left had an average grade of A in their other GCSEs, whereas the candidate on the right achieved straight grade A*s. The red dotted line shows the grade A* boundary on the scale of plausible marks. Notice that both candidates have a high chance of being simulated a plausible mark below this boundary even though in reality both achieved grade A*. To address this we can instead use conditional simulation as illustrated in the bottom two panels of Figure 1. In this method, each candidate's plausible mark is selected from the truncated part of the distribution above the grade A* boundary. Note that the mean of this truncated distribution for the candidate who had achieved straight grade A*s in all of their other GCSEs remains (slightly) higher than the mean for the candidate who had only averaged at grade A. In this way, the simulation ensures that even within candidates who have achieved a grade A* in a given subject, we expect the highest marks to occur amongst students with high attainment elsewhere.

Note that, as shown in Figure 2, the method of simulation we have described makes almost no difference to the overall distribution of

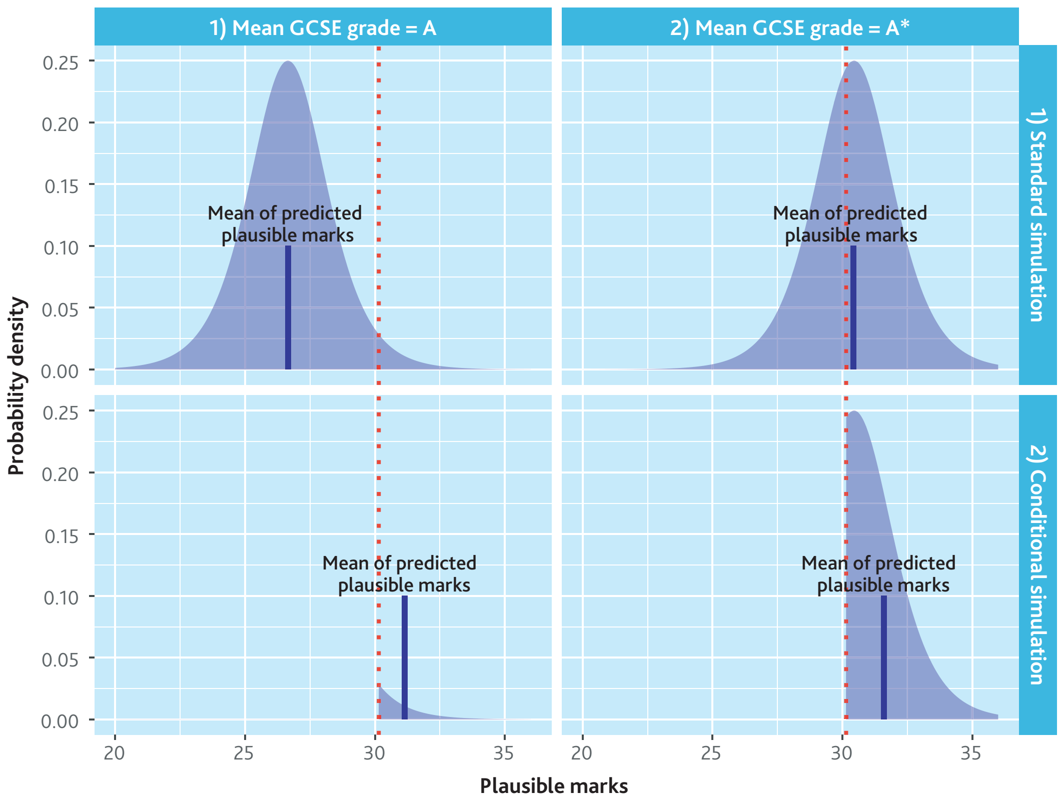

Figure 1: Illustration of different methods of simulating plausible marks for two candidates who had achieved grade A* in Mathematics

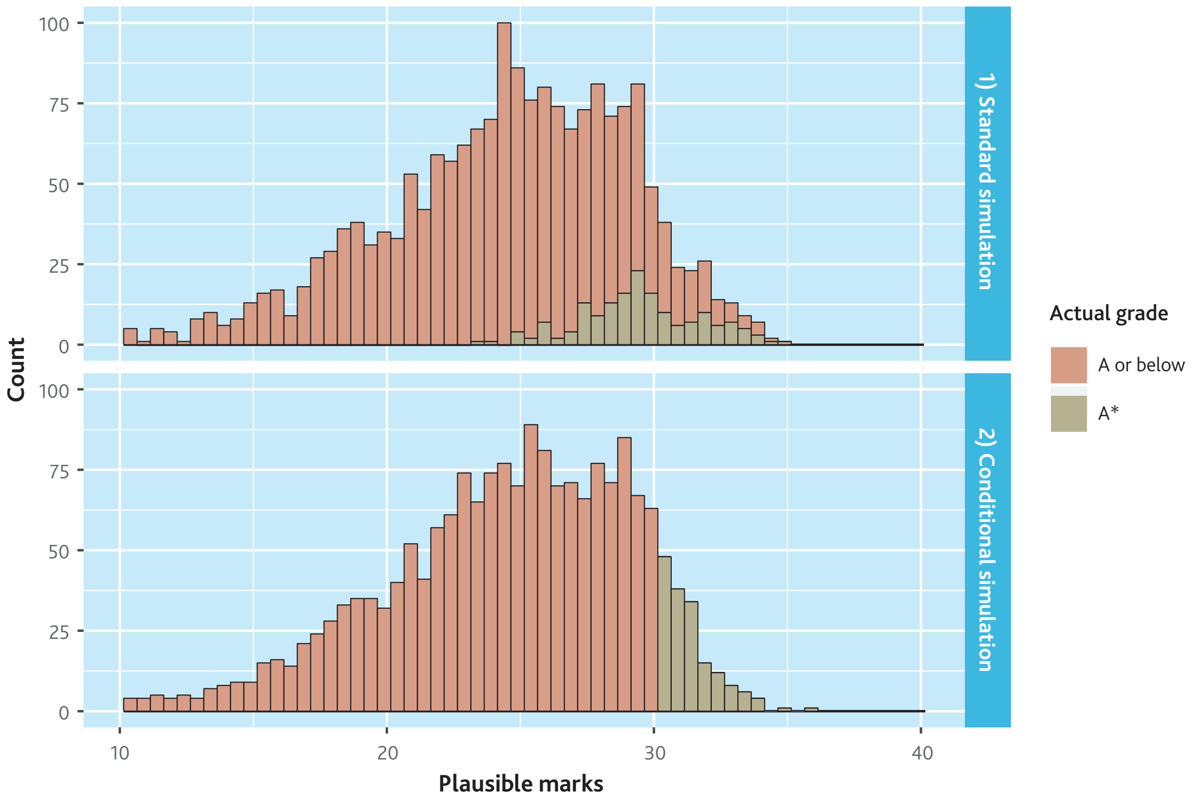

Figure 2: Overall distributions of plausible marks in Mathematics based on standard simulation and simulation conditional upon the actual grade candidates achieved

plausible marks. The overall distribution of plausible marks remains the same but we have ensured that, at least up to grade A*, the combinations of grades achieved by candidates will be consistent with reality. This means that, for example, candidates simulated to achieve straight grade 9s will always come from amongst those who actually achieved straight grade A*s. In this way we ensure that, at least up to grade A*, relationships between subjects are preserved. As such, the simulation is purely being used to ensure that, amongst candidates with a grade A* in any subject, it is recognised that the highest marks are likely to be achieved by candidates who have achieved good grades elsewhere. However, it should be remembered that measured achievement in other GCSEs is still capped at grade A*, and so the method is still likely to underestimate the true level of correlation between the marks of candidates within the highest grades in different subjects. This means that the predictions from this method are likely to be a lower bound to what should be expected in future rather than an exact prediction.

By design, the number of candidates predicted to get straight grade A*s via this method matched the actual number achieving this within the data set. More importantly, the predicted number to achieve straight grade 9s in Mathematics, English Language, and English Literature was very close at 1,829 to the actual number (2,050) reported by Ofqual for summer 2017. Furthermore, the fact that the prediction was marginally lower than the true number fits with the expectation that this method should provide a lower bound.

Prediction using a hybrid method

As we have mentioned, the weakness of the method based on logistic regression was that it only accounted known abilities up to grade A*. As such, it could not simulate the likely effect of very high ability students being likely to achieve well above the grade A* boundary across many subjects. In an attempt to address this, a second version of simulation based on logistic regression was run. However, rather than using the mean GCSE grade as the predictor of plausible marks in each subject, the simulated plausible values of general ability derived from the original IRT analysis were used. Unlike the mean GCSE, these simulated ability values were not capped and so, when used to further simulate plausible marks, could allow for higher correlations between different subjects at the top end of achievement. As we described for the previous method, plausible marks were simulated conditionally upon the actual grade that students achieved in each subject.

Again, by design, with the hybrid method the number of candidates predicted to get straight grade A*s matched the actual number achieving this within the data set. Also, as expected given that the cap on general ability had been lifted, the predicted number of students to achieve straight grade 9s in Mathematics, English Language, and English Literature rose considerably to 2,607 which is higher than the officially published number of 2,050. This may potentially indicate continuing weaknesses in the method. However, as shown earlier in Table 4, we already know that the characteristics of the 2017 cohort analysed by Ofqual differ a little from the 2016 data used in this analysis. Although some allowance has been made for this by removing a number of grade A and above Mathematics candidates, it is possible that this has not fully accounted for the differences between the data sets. For example, at the time of writing, the full extent to which high-performing independent schools have or have not transferred their entries from existing international GCSEs to reformed GCSEs is not known. If many of these centres had not switched over, it may explain the fact that some of our predictions are higher than currently published results. In theory, if more of such centres were to move to taking reformed GCSEs in future, it could substantially increase the number of candidates achieving straight grade 9s.

© UCLES 2018

Summary of results

Predictions from each of the three methods are summarised in Table 5. This table includes comparisons to actual results in terms of straight grade A*s in 2016, published results from Ofqual, and also to statistics published by the DfE regarding the number of candidates achieving grade 9 in Mathematics and then grade 9 in either English Language or English Literature in summer 2017. However, it should be noted that this last published statistic is limited to pupils in state schools and so will be lower than the full number. With this in mind, the results confirm the suggestion that the predictions based on logistic regression using the mean GCSE will be too low and that the predictions using either IRT or the hybrid method will be too high. With these caveats in mind, the results suggest that:

- between 1,000 and 2,000 candidates will achieve straight grade 9s in at least 3 GCSEs;
- if we restrict ourselves to those taking at least 5 GCSEs, between 600 and 1,500 candidates will achieve straight grade 9s;
- of those taking at least 8 GCSEs, between 200 and 900 candidates will achieve straight 9s;
- of those taking at least 10 GCSEs, between 100 and 600 candidates will achieve straight 9s; and
- thinking about the Attainment 8 accountability measure, we should expect more than 2,000 candidates to achieve a perfect score and that the number may be as high as 4,000.

Given the difficulty of attaining grade 9, the size of this last prediction is a particular surprise. The cause for the increase is that, although achieving a perfect score in Attainment 8 requires students to achieve at least eight grade 9s, it does allow for them to achieve below grade 9 in at least some GCSEs. A similar increase can be seen historically in statistics published in Gill (2017) which show that in June 2015, although only 3,300 achieved straight grade A*s, more than 8,500 achieved grade A*s in 8 or more subjects. On a similar theme, although Ofqual statistics show that only 2,050 pupils achieved straight grade 9s in Mathematics, English Language, and English Literature, the DfE's statistics show that, in state schools alone, more than 6,000 pupils achieved perfect scores across both the Mathematics and English 'pillars' of Attainment 8.

Final predictions

Overall, the analysis in this article confirms the initial prediction made in April 2017 that 'hundreds' of pupils will achieve straight grade 9s. If we restrict to candidates taking at least 8 GCSEs then the prediction is that between 200 and 900 of them will achieve straight grade 9s. If we take a purely literal definition of straight grade 9s, and include all candidates regardless of how few GCSEs they have taken, then the number is likely to be greater than 1,000. This article also provides a new prediction that at least 2,000 pupils will achieve perfect Attainment 8 scores in their GCSEs, and that this number may be as high as 4,000.

As might be expected, there are a number of caveats to these predictions. Firstly, it should be noted that these predictions are based upon GCSE and international GCSE entry patterns from June 2016. If GCSE reform leads to major changes in the popularity of different subjects and, in particular, to the numbers of GCSEs taken by different students, then this may have a noticeable impact upon the actual results. In addition, the extent to which high-attaining independent schools, which have historically entered their students for international

Table 5: Predictions and comparisons to (some) known results from 2017

Statistic (No. to achieve …)	*Predictions*				*Actual results*		
	No. of candidates relevant to prediction (base N)	*Method*			*No.*	*… out of (No. of relevant candidates)*	*Source*
		Pure IRT	*Logistic regression-based plausible marks*	*Hybrid*			
… straight grade A*s in Maths, English Language, and English Literature	498,226	6,382	5,891	5,891	5,891	498,226	2016 analysis data set
… straight grade 9s in at least 3 GCSEs	565,431	2,045	1,077	2,054	-	-	-
… straight grade 9s in at least 5 GCSEs	535,216	1,563	620	1,516	-	-	-
… straight grade 9s in at least 8 GCSEs	382,278	894	216	817	-	-	-
… straight grade 9s in at least 10 GCSEs	131,876	619	110	508	-	-	-
… straight grade 9s in Maths, English Language, and English Literature	498,226	2,816	1,829	2,607	2,050	508,950	Ofqual analytics 2017[a]
… grade 9 in both the Maths and English pillars of Attainment 8	537,207	8,388	6,396	7,250	6,129[b]	527,859	DfE statistics 2017[c]
… a perfect Attainment 8 score	566,879	4,247	2,598	3,797	-	-	-

a. https://analytics.ofqual.gov.uk/apps/2017/GCSE/9to1/
b. The published figure from the DfE is restricted to state funded schools only.
c. See Characteristics national tables at https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2016-to-2017

© UCLES 2018

GCSEs, either have (or will) switch their entries to reformed GCSEs is not known. Whether such schools contribute to the national GCSE results will make a noticeable difference.

Although many GCSE subjects will have been reformed by summer 2018, the final test of these predictions will not be until after summer 2019. In particular, our analysis has shown that some minor Modern Languages (Chinese, Russian, and Italian) are very popular amongst candidates who have achieved straight grade A*s historically and so, only when results for the reformed versions of these subjects are available (summer 2019), will we know the accuracy of our predictions.

Regardless of whether the predictions are right or wrong, one thing is clear: Achieving grade 9 in any GCSE subject is hard. Congratulations to all those students who achieve it in any subject at all.

References

Benton, T. (2016). *A possible formula to determine the percentage of candidates who should receive the new GCSE grade 9 in each subject*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment. Retrieved from http://www.cambridgeassessment.org.uk/Images/298710-a-possible-formula-to-determine-the-percentage-of-candidates-who-should-receive-the-new-gcse-grade-9-in-each-subject.pdf.

Benton, T. & Bramley, T. (2017). *Some thoughts on the 'Comparative Progression Analysis' method for investigating inter-subject comparability*. Cambridge Assessment Research Report. Cambridge UK: Cambridge Assessment. Retrieved from http://www.cambridgeassessment.org.uk/Images/416591-some-thoughts-on-the-comparative-progression-analysis-method-for-investigating-inter-subject-comparability.pdf/

Benton, T. & Sutch, T. (2013). *Exploring the value of GCSE prediction matrices based upon attainment at Key Stage 2*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment. Retrieved from http://www.cambridgeassessment.org.uk/Images/181034-exploring-the-value-of-gcse-prediction-matrices-based-upon-attainment-at-key-stage-2.pdf.

Chalmers, R.P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48*(6), 1–29. Retrieved from http://www.jstatsoft.org/v48/i06/.

Coe, R. (2008). Comparability of GCSE examinations in different subjects: an application of the Rasch model. *Oxford Review of Education, 34*(5), 609–636.

Gill, T. (2017). *Candidates awarded A* and A grades at GCSE in 2015*. Statistics Report Series No. 111. Cambridge, UK: Cambridge Assessment. Retrieved from http://www.cambridgeassessment.org.uk/Images/352252-candidates-awarded-a-and-a-grades-at-gcse-in-2015.pdf.

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2016). What can we learn from Plausible Values? *Psychometrika, 81*(2), 274–289.

Ofqual. (2014). *Consultation on Setting the Grade Standards of new GCSEs in England*. Ofqual, Ofqual/14/5401, Coventry. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/630155/2014-04-03-consultation-on-setting-the-grade-standards-of-new-gcses-in-england.pdf.

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201–292. Retrieved from http://www.jstor.org/stable/i261624

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*(2–3), 114–128.

How do you solve a problem like transition? A qualitative evaluation of additional support classes at three university Biology departments

Simon Child Research Division, **Sanjana Mehta** Cambridge Assessment English, **Frances Wilson** OCR, **Irenka Suto** Research Division, and **Sally Brown** Cambridge Assessment Network

(The study was completed when the second and third authors were based in the Research Division, and the fifth author was based at OCR)

Introduction

Some university Biology departments have introduced additional support classes for students who struggle with the transition from school or college to higher education (HE). In this study, classes at three contrasting British universities were investigated. The structure and content of the classes were compared, and reasons for introducing the classes were explored. Data collection comprised linked observation and interview methods from three stakeholder perspectives: lecturer, undergraduate, and teacher. This article discusses the transitional challenges identified by the different stakeholders in relation to the recently completed reforms to General Certificate of Education (GCE) Advanced level (A level) qualifications in the Sciences.

Background

Many students experience difficulties in making the transition from school or college to university (Lowe & Cook, 2003; Pampaka, Williams, & Hutcheson, 2012) and lecturers have frequently expressed dissatisfaction with the skills and knowledge that new undergraduates possess when they first enter university following their A levels (Mehta, Suto, & Brown, 2012). In the Biosciences, report-writing and mathematical abilities have been identified as weak (H. Jones, 2011; Suto, 2012). Skills in practical Science, including the use of equipment and data analysis, have also raised concern (J. Wilson, 2008). Poor retention of basic biological concepts has been attributed to a reliance on surface learning during A level (pre-reform) and other pre-university

© UCLES 2018