Research *Matters*

Issue 28 / *Autumn 2019*

Proud to be part of the University of Cambridge

Established over 150 years ago, Cambridge Assessment operates and manages the University's three exam boards and carries out leading-edge academic and operational research on assessment in education. We are a not-for-profit organisation.

Citation

Articles in this publication should be cited using the following example for article 1: Benton, T.(2019). Which is better: one experienced marker or many inexperienced markers? *Research Matters: A Cambridge Assessment publication*, 28, 2-10.

Credits

Reviewers: Lucy Chambers, Filio Constantinou, Sylwia Macinska, Tom Sutch and Joanna Williamson, Research Division, Cambridge Assessment
Editorial and production management: David Beauchamp and Anouk Peigne, Research Division, Cambridge Assessment
Additional proofreading: Sarah Taylor
Cover image: John Foxx Images
Design: George Hammond
Print management: Canon Business Services

All details are correct at the time of going to print in October 2019

If you would like to comment on any of the articles in this issue, please contact Tom Bramley, Director, Research Division. Email: researchprogrammes@cambridgeassessment.org.uk

This and all previous issues of *Research Matters* are available to download, in full and by individual article, from our website: www.cambridgeassessment.org.uk/research-matters

Research *Matters* / 28

A CAMBRIDGE ASSESSMENT PUBLICATION

Foreword

There is much talk about raising attainment in education systems, and accompanying discussion about equity. Each is important, and public policy debate becomes particularly interesting when approaches to improving both are the focus of attention. For sure, we have seen responses which improve one and not the other, or indeed improve one with deleterious impact on the other. Raising attainment and improving equity is not a Holy Grail – it can indeed be achieved, as Eric Hanushek's seminal work[1] on setting and streaming shows.

But when UK universities try to widen participation, they run straight into the issue of subject choice at A Level – the subject of Carmen Vidal Rodeiro's reprise analysis. Choose the wrong subjects, and some courses will be closed to you, for entirely understandable reasons. With schools judged largely by their grade outcomes, and immediate learner preferences often driving choices (Martin Bloomer's 'local rationality' argument[2]), premature closing of options is bad for individuals and bad for our society and economy. And despite gender balance and participation improving in A Level physics, it is still the case that in 2018 only 22% were female. Subject choice remains adversely affected by many factors – and feeds straight into inequality in options and progression. Aimed squarely at enhanced equity, Government policy on post-16 Maths and English has signalled clearly the link between good attainment in these subject and life chances (and thus social justice) but Jo Ireland's article shows clearly the critical practical issues which need to be worked through to enable the policy to achieve its laudable ambitions. Social justice questions also permeate the question of predicted grades – The Government's 2011 analysis and Tim Gill's previous analysis have indicated that differences in social background previously have affected predictions, while his latest analysis suggests a genuine effort by professionals to enhance their accuracy.

Attainment and equity; there are many things which affect each. Enhancing both without improving one at the expense of the other requires the kind of analyses we present here, and continued sophisticated public policy effort.

Tim Oates, CBE *Group Director, Assessment Research and Development*

Editorial

How should we define what is the 'correct' mark to give the response to an exam question (and the paper as a whole)? Should it be what the most experienced marker would give it, or the average 'wisdom' of a (small) crowd of less experienced markers? This fundamental question is addressed by Tom Benton in the first article in this issue of *Research Matters*. How we choose to define 'correct' has implications both for how we mark, and how we monitor marking.

The second article by Tom Gallacher and Martin Johnson takes a critical look at how some of the recent literature about 'learning progressions' fits into the larger picture of academic thinking about teaching, learning and curriculum design.

The third article by Carmen Vidal Rodeiro presents some key findings from a larger study exploring what HE courses are taken, and at what kinds of HE institution, by students with different subject choices at A Level.

There has been a lot of debate recently about the merits or otherwise of making students who do not achieve a grade C or 4 at GCSE in English or Maths continue to study these subjects as part of their post-16 curriculum. Jo Ireland draws out some themes from the different aspects of this debate in our fourth article.

In our final article, Tim Gill reports on a survey of a relatively small number of schools in three different subject areas aimed at finding out how they went about making their predictions of A Level results for individual students, how accurate those predictions were, and how they compared with the findings from a similar survey carried out before the reform of A Levels.

The five articles in this issue thus cover a variety of topical issues in education and assessment – I hope you enjoy reading them.

Tom Bramley *Director, Research Division*

1. Hanushek, E., & Wossman, L. (2006). Does educational tracking effect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal, 116*, C63–C76.
2. Bloomer, M., & Hodkinson, P. (2013). Learning careers: continuity and change in young people's dispositions to learning. *British Educational Research Journal, 26*(5), 583–597.

Which is better: one experienced marker or many inexperienced markers?

Tom Benton Research Division

Introduction

At present, the ultimate arbiter of the mark that should be awarded to any exam script is the principal examiner (PE). The reasons for identifying a single individual as the ultimate judge are mainly practical. For example, it means that when marking is reviewed, there is a single person with the authority to determine the correct mark in difficult cases where there is a disagreement over the marks that should be awarded. Furthermore, if the PE marks a number of scripts at the beginning of the marking process, then the extent to which other examiners are able to independently replicate these marks on the same scripts provides a good indicator of their suitability to begin marking. Marks from the PE are used in this way during both practice and standardisation of examiners at a time when no other marking has been completed. A similar process can also be used to monitor the quality of marking of different examiners right from the start of marking and throughout.

However, beyond these practical considerations, this approach also indicates a set hierarchy of markers. It implies a predetermined axiom that the best way to mark a given assessment is defined by the most senior marker (the PE). Therefore, the whole machinery of marking should be about communicating the PE's approach to marking to more junior markers, and them being evaluated with the regard to the extent to which their marks are in line with those that would be awarded by the PE.

At the top of the hierarchy is the PE. At the next layer of the hierarchy are team leaders – fairly experienced examiners who are responsible for supporting and monitoring teams of more junior markers. At the bottom of the hierarchy of markers are assistant examiners (AEs) who comprise the vast majority of the individuals involved in marking. Whilst junior, AEs are professionals who are appropriately trained and standardised before beginning marking. However, where marks awarded by AEs differ from those awarded by the PE, the default assumption is that it is the AEs that are wrong. Even where numerous AEs have a shared professional opinion of the mark that should be awarded a given script, the PE's mark is still assumed to be the correct one.

This article explores the evidence for the assumed supremacy of the PE over groups of more junior markers. Understanding the truth of this assumption has important implications both for the way in which markers are monitored, and for the way that quality of marking is reported at a national level.

Quality of marking, both for individual markers and at a national level, is usually evaluated via a marker monitoring process. This process typically collects data about each marker as follows. For each examination, before the main body of marking begins, a small number of scripts are pre-marked by senior examiners (usually including the PE), either in consultation with one another or acting alone. These scripts are known as "seed scripts" or "seeds", and the marks that the PE has agreed should be awarded to them are called the "definitive marks". During live marking, these seeds are placed at various random intervals in each marker's script allocation (Ofqual, 2016). Each marker marks these scripts "blind" (i.e., unaware that they have already been marked before and without knowledge of the pre-determined definitive mark). The marks awarded to seed scripts by individual markers are compared to the definitive marks and are continuously analysed for any signs that examiners' marking may be becoming inaccurate. After the end of an examination series, this same data is also used by The Office of Qualifications and Examinations Regulation (Ofqual) to produce reports on the quality of marking nationally (e.g., Ofqual, 2016). The implicit assumption both in the monitoring process, and in the post-series analysis, is that marks awarded by the PE (possibly in consultation with other senior examiners) can be treated as being absolutely correct (Suto, Nádas, & Bell, 2011) and that, as such, differences from these represent errors that must be eradicated.

With this backdrop in mind, this article aims to investigate the accuracy of definitive marks using empirical data. Specifically, it addresses the question of whether marks awarded by a PE are genuinely more useful than those derived by combining the marks of several, less experienced, AEs. If the PE's marks are truly "better" then they should be at least as predictive of wider achievement as the marks pooled from the AEs. This article will test whether this is in fact the case.

As noted above, definitive marks may, in some cases, be assigned by a team of senior examiners rather than by a PE working alone (Black, Suto, & Bramley, 2011). However, in order to simplify the language of the remainder of this article, we shall treat them as if they are generated by a single individual. For some definitive marks this is literally true, but no formal record is made of which senior examiners were involved in assigning particular definitive marks so the extent of this is not known. However, even if such marks involve some collaboration between senior examiners it is hard to believe that this would result in lower accuracy than would be achieved by a PE working alone. With that minor caveat in mind, we will proceed to the analysis.

Data and Method

This analysis makes use of seed scripts from all of OCR's GCSE, AS and A Level papers from summer examination sessions between June 2015 and June 2018. For the purposes of this research, seed scripts are valuable as they provide a small number of scripts for every examination that have both been marked by the PE, and have been marked by several AEs.

Analysis is restricted to all seed scripts that had been marked by at least five AEs. It is also restricted to those examination components (papers) where at least ten such seed scripts were available. This left a total of 724 papers for analysis across the four examination sessions. Further details regarding the data set used in analysis are given in Table 1.

Table 1: Details of data used in analysis

Number of seed scripts	10,786
Number of examination components	724
Number of AEs	20,827
Median number of seed scripts per component	15
Median number of markers per component	21

To begin with, the mark that should be awarded to each seed script was derived in three different ways.

1. **The definitive mark.** As mentioned earlier, such marks are usually decided upon by the PE themselves, although other senior examiners may also be involved in the discussion.
2. **The mean mark awarded across all AEs that marked the script.** Note that if an AE does a large amount of marking then they may mark all of the available seed scripts once, and then (having completed this), any further seed scripts they are assigned will be ones they have marked before. It is also possible that during marking, they may be given feedback based on the marks they award seed scripts. For this reason, to keep this measure as free from the influence of senior markers as possible, only the first attempt each AE made at marking each seed script was included in this measure. Note that, for the purposes of this research study, markers were included in the generation of this mean mark even if they were later deemed to be inaccurate and stopped from live marking[1]. This decision was taken to ensure that this research gave a pure idea of the accuracy of multiple junior markers without the influence of more senior markers in choosing who should be included. The mean was used as it is the simplest possible method for combining the marks of various AEs into a single score. Also, given that for any individual whole script, the marks assigned by different examiners tend to follow a roughly normal distribution (see Ofqual, 2016, Figure 11), the mean might reasonably be assumed to be the most efficient estimate.
3. **The median mark awarded across all AEs that marked the script.** As above, but using the median mark across all markers rather than the mean. Possible advantages of the median include resistance to outliers and the fact that it is more likely to result in a mark that is a whole number.

For the purposes of brevity within this article, the second and third methods of assigning marks to each script will be referred to as *collectively assigned* marks.

In order to evaluate the predictive power of different ways of generating marks for these seed scripts, I used the *external* ISAWG (Benton, 2017; Benton, 2018) for the candidates providing the seed scripts. The ISAWG is a standardised measure of each candidate's achievement that summarises their performance across all of the assessments they have completed within a given examination session. For further details on how it is calculated, see Benton (2017). The external ISAWG is based on the same calculation but only using assessments other than the one being studied. It can be interpreted as a very general measure of ability across different subjects. It was used in this analysis, as it was easily available for nearly all the candidates from whom seed scripts were selected.

The predictive value (or concurrent validity) of the three different ways of generating marks for each seed script was evaluated by the Pearson correlation of the marks with the external ISAWG. That is, the analysis identified which of the three different ways of generating the marks for each script was most predictive of candidates' wider achievement.

Note that, for each individual examination being studied, only 10 to 20 seed scripts were typically available for analysis. As such, the individual correlations, based on such a small number of candidates, were almost meaningless when considered individually. However, by analysing them as a group across the hundreds of examination papers included in the analysis, we hope to identify some clear trends with regard to whether definitive marks from a PE or pooled marks across many AEs have more predictive power.

Results

Comparison of predictive power of definitive marks versus consensus marks

Table 2 provides a summary of the results across the 724 components included in the analysis. The central result from this table is that collectively assigned marks for seed scripts (whether calculated using the mean or the median) were slightly more predictive on average of external achievement than the definitive marks.

The fact that the difference in predictive power between collectively assigned and definitive marks is small (just below 0.02) is not unexpected. Definitive marks are from the most senior marker and should be very accurate. As such, there is likely to be relatively little room for improvement. Nonetheless, the results show that pooled marks from several AEs have more predictive power than the marks from PEs. Whilst we can invent fanciful stories about how these results might have occurred (e.g., "assistant examiners are more influenced by signs of general ability whereas PEs recognise abilities specific to the assessment being studied"), the most straightforward interpretation of these results is that collectively assigned marks from multiple examiners are simply a more accurate indicator of a candidate's performance. That is, the many junior examiners outperform the single PE.

Table 2 splits the summarised results into components within Science or Mathematics, those within Modern Foreign Languages (MFL) and all of the rest including Humanities, Expressive Subjects and Classics. For all subject groups, the collectively assigned marks had more predictive power than the definitive marks. However, the gain in predictive power for Science, Mathematics and MFL (less than 0.01) was less than the gain for other subjects (close to 0.03). This result mirrors the general finding from most previous research that marking is most reliable in "exams containing structured, analytically marked questions" (such as are likely to occur in Science, Mathematics and MFL) and generally less reliable in "exams containing essays" (such as Humanities); see Bramley (2008). Similarly, the subjects where collectively assigned marks have the greatest superiority in predictive

1. Where markers are stopped from marking, all of the live scripts they have already marked are then allocated to a new marker to be remarked. However, for the purposes of this research, the marks awarded by such markers are available within our systems.

Table 2: Summary of correlations of different ways of assigning marks to seed scripts with external achievement

Subject	*Number of components*	*Mean correlation with…*			*Median correlation with…*		
		Definitive mark	*Mean mark*	*Median mark*	*Definitive mark*	*Mean mark*	*Median mark*
Science and Mathematics	302	0.788	0.793	0.793	0.812	0.823	0.822
MFL	33	0.715	0.725	0.722	0.747	0.753	0.755
Other	389	0.744	0.771	0.767	0.790	0.815	0.808
Total	**724**	**0.761**	**0.778**	**0.776**	**0.796**	**0.814**	**0.811**

power are also those where differences between markers are less likely to be "procedural" or "attentional errors" and more likely to be due to "inferential" or "definitional uncertainty" (Ofqual, 2018b).

It is worth noting that, although collectively assigned marks were allowed to be non-integer numbers (i.e., decimals), rounding the marks before calculating correlations had virtually no impact on the results in Table 2. The largest change from rerunning the analysis with rounded marks occurred for the mean of the correlations of mean marks with external achievement in MFL and was still only a change of 0.004 (a drop from 0.725 to 0.721). The median change of those figures in Table 2 that changed at all was just 0.0005. In other words, it is the fact that marks are derived from a collection of individuals that drives the good performance of collectively assigned marks and not simply the fact that they can include non-integer numbers.

As can be seen from Table 2, there was very little difference in predictive power regardless of whether collectively assigned marks were calculated using the median or the mean. For this reason, the remainder of this article will focus upon the use of the mean to generate such marks.

A visual representation of the difference in the predictive power of definitive and collectively assigned marks is given in Figure 1. For each of the 724 examination papers in the analysis, the chart plots the correlation of external achievement with the definitive marks against the correlation with the mean mark. Although the individual points in this chart are based on very small samples, and thus, nearly meaningless, it is the overall pattern that is of interest. To help discern this overall pattern, the plot includes a dotted line representing a line of equality and a solid blue line which represents a regression

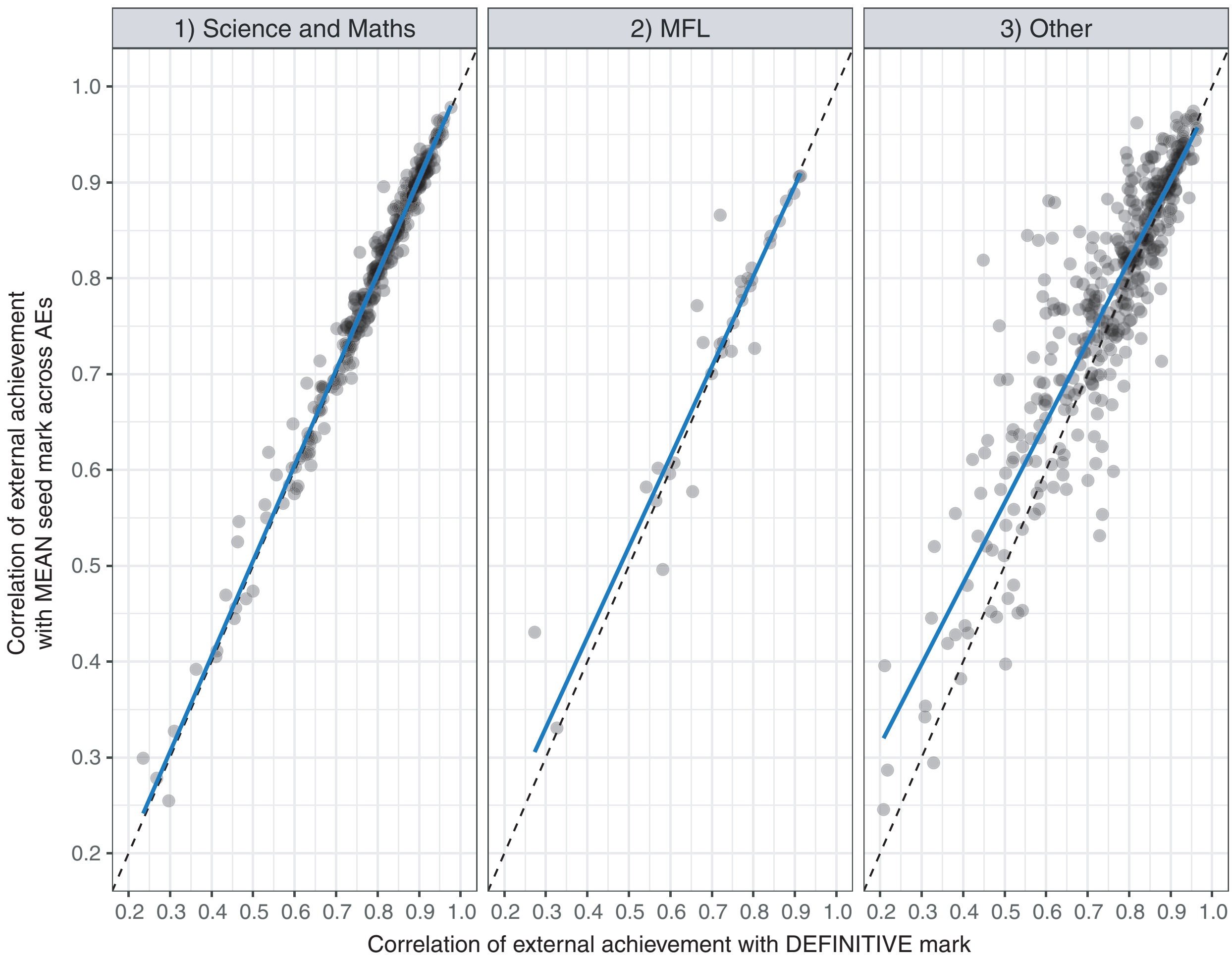


Figure 1: Comparing correlations of external achievement with mean mark awarded by AEs and with definitive marks awarded by PEs.

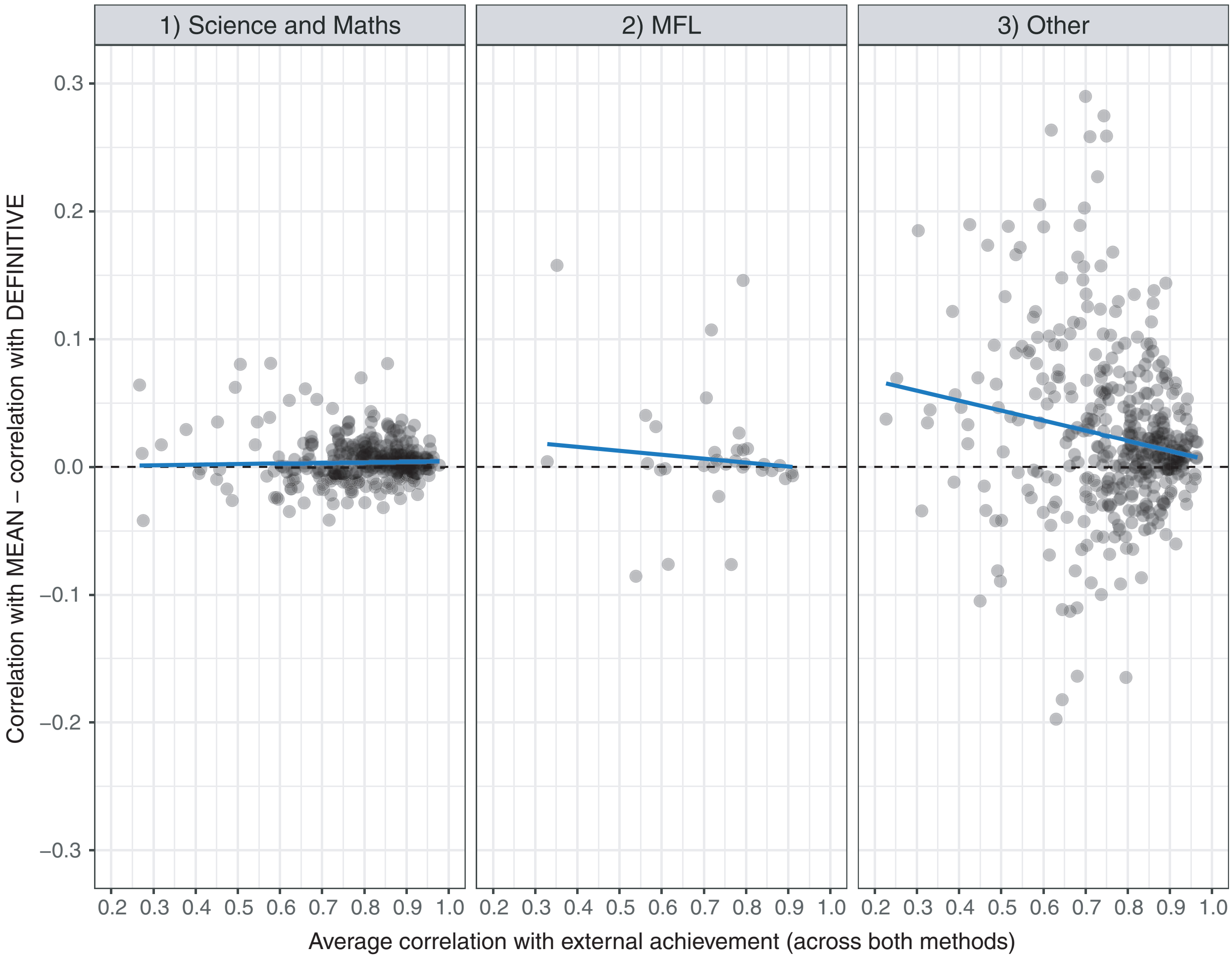


Figure 2: Mean-difference plot showing how the difference between the predictive power of collective and definitive marks relates to the average predictive power of both methods.

line[2]. As can be seen, for Science, Mathematics and MFL, there is barely any gain from using collective marks rather than definitive marks. However, for other subjects the blue regression line is clearly above the dotted line of equality. In other words, there are obvious gains from using collective marks rather than definitive marks.

Another way to view this same data is given in Figure 2. Using a Tukey mean-difference plot[3] allows a more detailed view of the differences between methods. Of particular interest is the right hand panel. This shows that the gain from using collective marks is largest when the predictive power of both methods is relatively low. Remember that each point in the chart is based on analysis of a very small number of scripts; as such, if the selected candidates happen to perform very differently in other examinations to the examination being studied, these correlations can be quite low. On the other hand, if the selected candidates happen to perform similarly well in other examinations then these correlations may be high and the potential for improvement is much lower. Nonetheless, regardless of the size of the gain, the central point from this analysis remains – the average mark from many junior examiners is superior to the mark from a PE.

2. Due to the very small sample sizes used to generate each correlation in the plots, and the associated increased risk of outliers, all regression lines in this article were created using robust regression. Regression lines were created using the function *rlm* from the R package *MASS* (Venables & Ripley, 2002).

3. Popularly known in biomedicine as a Bland-Altman plot.

The impact of the structure of exam papers

So far, we have seen that collectively assigned marks have more predictive power than definitive marks, and that this appears especially true in subjects that are likely to require more professional judgement in marking. However, this last point has only been crudely demonstrated by splitting assessments of Science, Mathematics and MFL from the remainder. This next section attempts to find a more universal measure of the likely level of subjectivity required to mark each exam paper.

The principle used to derive this measure is based on data on marking consistency across all awarding organisations combined reported in Ofqual (2016). Figures 4 and 5 of Ofqual's report (pages 10 and 11) show that (across many items) the scale of inconsistency between markers tends to increase in a way that is roughly proportional to the number of marks that are available on an item. That is, whilst it is accepted that there are many additional features of items and mark schemes that may affect marking consistency (Bramley, 2008; Black, Suto, & Bramley, 2011), to a large extent, the scale of marking consistency for markers can be explained by the number of available marks on the items they are marking.

More specifically, Ofqual's research shows that the standard deviation of marking differences (between individual markers and definitive marks) is roughly proportional to the number of marks available of the item. That is, *standard deviation of marking differences on item i* $\approx k * Max_i$

where k is a fixed constant[4] and Max_i is the maximum number of marks available on item i. This formula gives us a straightforward way of estimating the likely scale of marker inconsistency at item level. To estimate the scale of marker inconsistency at the level of whole scripts we use standard statistical rules. These tell us that, if marker differences on different items are independent (and we would hope that they are), then the standard deviation of marker differences at whole script level should be the square root of the sum of the squared standard deviations of differences across each item. That is, using equations similar to some given on page 15 of Ofqual's report,

Standard deviation of whole script marking differences

$$= \sqrt{(SD\ of\ diffs\ on\ Q1)^2 + (SD\ of\ diffs\ on\ Q2)^2 + (SD\ of\ diffs\ on\ Q3)^2 + \cdots}$$

$$\approx \sqrt{(k * Max_1)^2 + (k * Max_2)^2 + (k * Max_3)^2 + \cdots}$$

$$= \boldsymbol{k} * \sqrt{\textbf{\textit{Sum of squared item maxima}}}$$

The final term is the square root of the sum of the squared item maxima. From now on, we will refer to this as the RSSIM. Although, to some readers, the equations above may appear complex, the calculation of the RSSIM is relatively straightforward and provides a single metric which is proportional to the likely extent of disagreement between markers. For example, suppose an exam requires candidates to answer three questions in an exam with 10, 20 and 30 marks available respectively. In this case the RSSIM=$\sqrt{100+400+900} = \sqrt{1400} = 37.4$. On the other hand, if another 60 mark assessment consists of 60 items each with one mark available then RSSIM= $\sqrt{1+1+1+\cdots} = \sqrt{60} = 7.7$. Thus, we would expect the amount of variation between markers to be roughly five times higher in the former 60 mark test than in the latter one.

4. Using Figures 4 and 5 of Ofqual (2016) we can see that k is approximately equal to 0.12.

Returning to our main research question, we can investigate how the gain in predictive power from using collectively assigned rather than definitive marks relates to the RSSIM. This analysis is provided in Figure 3. As expected, the greatest gain from using collectively assigned marks is found for those exams where marking consistency is hardest to achieve – that is, those with the largest RSSIM values. It can also be seen that the RSSIM also serves to explain the differences between the subject groupings used earlier. This is important as it means that the RSSIM can be used as a shortcut to classify exam papers in terms of the likely difficulty of marking without the need to manually decide upon how to group different subjects.

How many ordinary markers is the one good marker worth?

Having seen that collectively assigned marks have greater predictive power than definitive marks, the aim of this section is to determine how many AEs are needed in order for this to be the case. In order to answer this question the following procedure was used for each assessment within the study:

1. For each seed script in turn, randomly select n AEs to contribute to the collectively assigned mark.
2. Calculate the collectively assigned (mean) mark of each seed script based on the selected markers only.
3. Calculate the correlation of the (new) collectively assigned marks to all of the seed scripts with external achievement.
4. Repeat steps 1 to 3 500 times and record the average correlation across all of the replications.

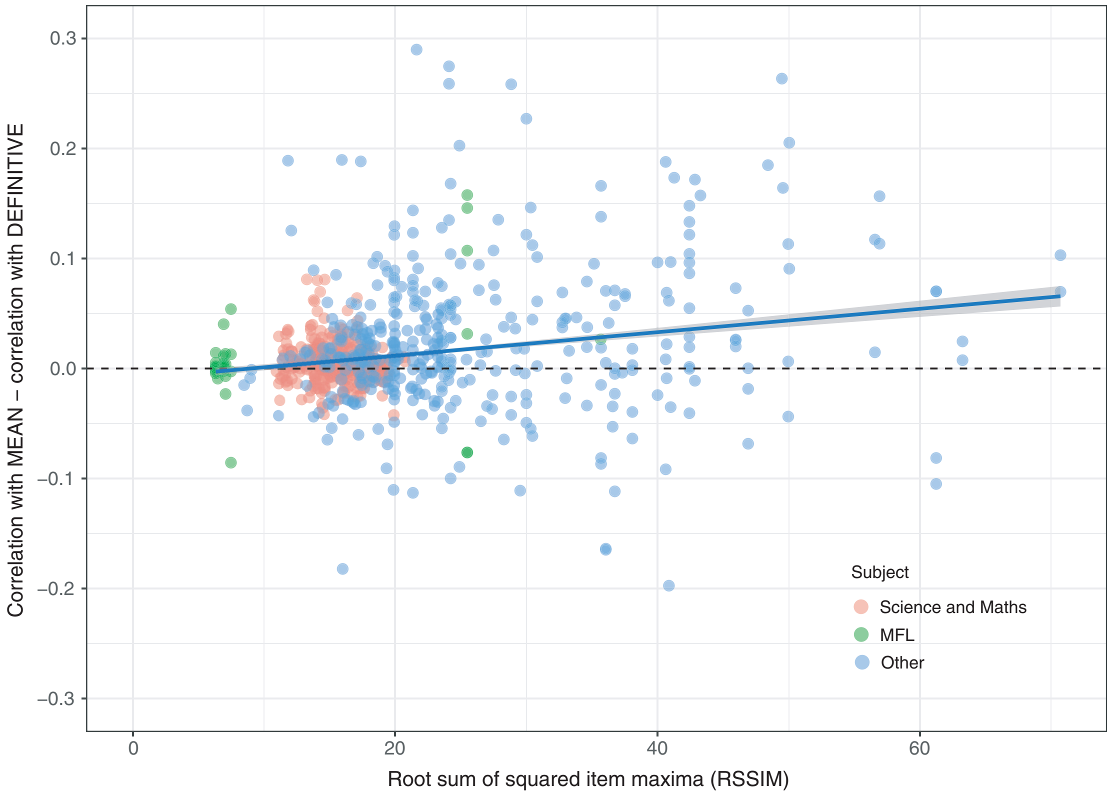


Figure 3: Relationship between the root sum of squared item maxima (RSSIM) and the gain in predictive power from using collectively assigned rather than definitive marks.

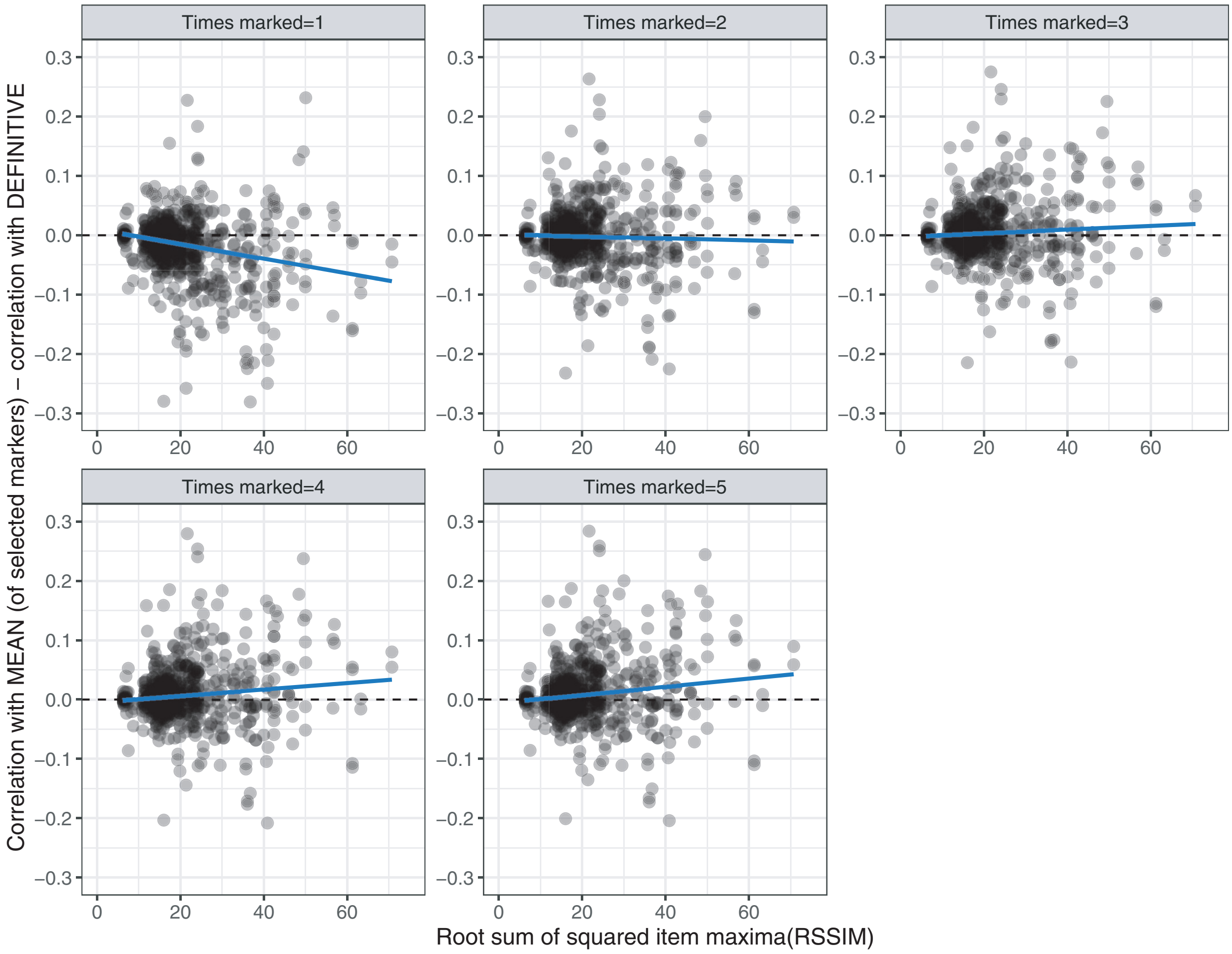


Figure 4: Relationship between the root sum of squared item maxima (RSSIM) and the gain in predictive power from using collectively assigned marks from between 1 and 5 AEs rather than definitive marks.

The above procedure was used to investigate the predictive power of collectively assigned marks from 1, 2, 3, 4 and 5 AEs (i.e., $n = 1, 2, 3, 4$ and 5).

The results of this analysis are shown in Figure 4. This chart shows that if only one AE marks each script (not really a collectively assigned mark at all) then the definitive marks tends to have noticeably more predictive power. If the mark awarded to each seed script is derived from two AEs (double marking) then definitive marks were still superior (but only just). However, marks collectively assigned by three or more AEs tended to have more predictive power than the definitive marks. Thus, it appears that a PE is worth between two and three AEs. This result is broadly consistent with results reported for marking of an English Language essay in Benton and Gallacher (2018).

How accurate are definitive marks?

For the final section of this article, we change direction slightly. We have seen in the previous section that definitive marks (awarded with the involvement of the PE) have greater predictive power than those awarded by individual AEs. In other words, PEs are very good markers. However, the preceding sections showed that the collectively assigned marks from multiple AEs have even greater predictive power – that is, they are better still. Taking these two facts together means that the data analysed in this article provides an opportunity to explore the accuracy of definitive marks themselves.

To gain some idea of the accuracy of definitive marks we look at the scale of differences between definitive marks and collectively assigned marks. Before doing this, it is important to put the two sets of marks on a common scale. Collectively assigned marks will tend to have a lower standard deviation than definitive marks. Broadly speaking, this is because it is harder for a candidate to convince two examiners that their essay is a work of genius than it is to convince one alone. Similarly, and on a more positive note, you would be unlucky to find two examiners in succession who both think your essay is entirely without merit. To account for this fact, before doing this analysis, the collectively assigned marks in each assessment were rescaled to have the same mean and standard deviation as the definitive marks.

This change in scale did not matter for the earlier analysis of correlations because correlations are scale free measures of association. However, in this analysis we are investigating the actual differences between methods in terms of marks. As such, accounting for changes in scale is important.

The results of this analysis are shown in Figure 5. Each point in this chart represents an assessment. The RSSIM of the assessment is plotted against the average across all seeds of the absolute difference between (rescaled[5]) collectively assigned marks and definitive marks. Once again, it is worth reiterating that each individual point is based on a very small number of scripts and, as such is not too meaningful in itself. Furthermore, it is worth remembering that all available AEs have been retained in this analysis including those that were ultimately stopped

5. Skipping the step of rescaling does not make an enormous difference to the results. The scale of differences between definitive and collectively assigned marks just becomes slightly larger.

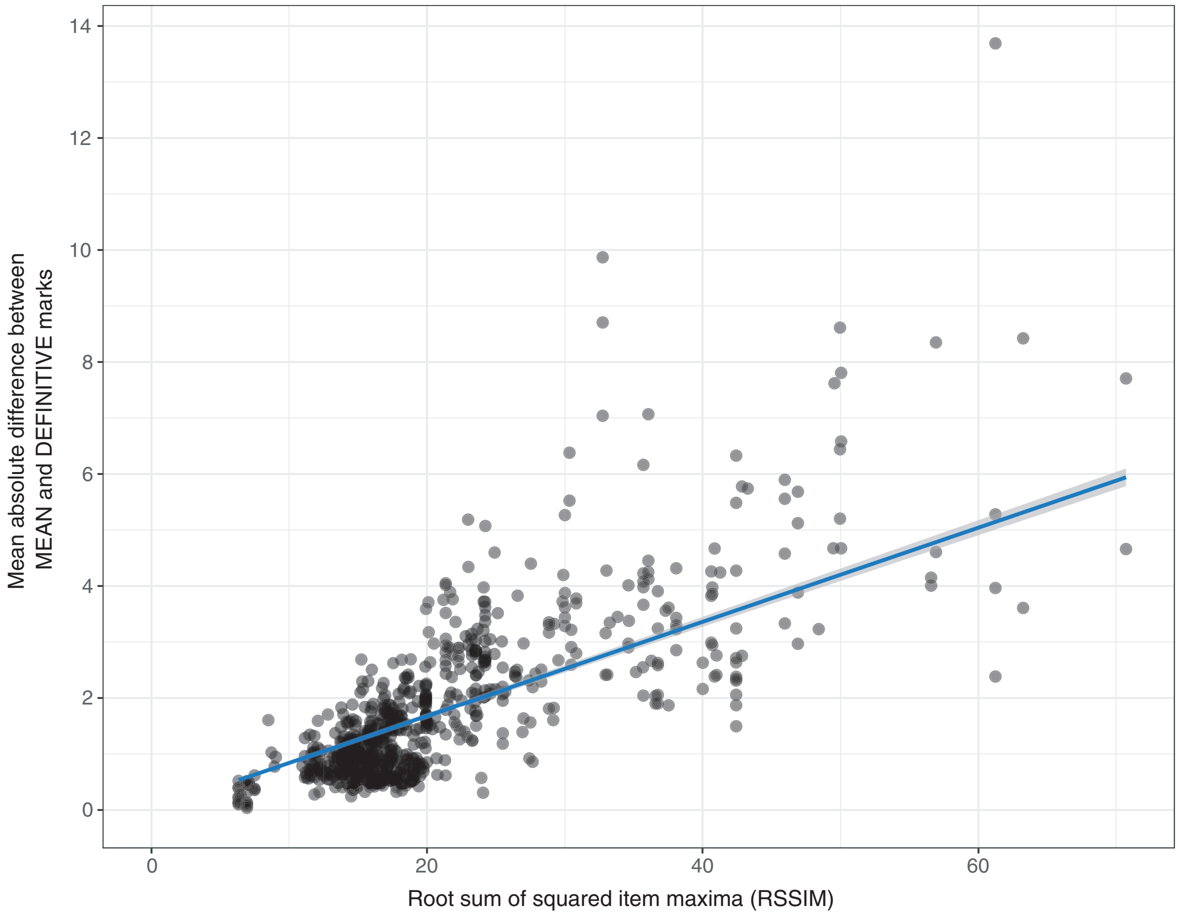


Figure 5: Relationship between the root sum of squared item maxima (RSSIM) and the mean absolute difference between collectively assigned and definitive marks.

from marking. For this reason, occasional markers who displayed very large discrepancies from definitive marks are included within the analysis leading to the possibility of outliers.

However, the main purpose of Figure 5 is to look at the overall trend. This is captured by the solid (robust) regression line. For the purposes of this chart only, this regression line was defined so that if extended it would pass through the origin. This decision was partly driven by common sense (if there are zero marks available on an assessment then it is impossible for any disagreement between markers to emerge), but was also made to allow us to derive a very simple rule for the likely accuracy of definitive marks. Specifically, the regression line shows that we expect the mean absolute difference between definitive and collective marks to be roughly one-twelfth of the RSSIM.

This result provides a lower bound for the level of consistency we can expect between markers. On the one hand, for each examination, we have taken our most experienced marker: the best of the best PE. On the other hand, we have a set of marks that the evidence suggests is even better than that – the collectively assigned marks. The analysis shows that even for these two most accurate sources of marks the difference between them tends to be about one-twelfth of the RSSIM. Thus, we cannot reasonably expect that the average difference between any two independent examiners, whether the PE or anyone else, would be less than this. For example, this implies that for an assessment with an RSSIM of 40 (as, for example, would occur for an exam out of 80 consisting of four 20 mark items), we cannot reasonably expect the average difference between the marks awarded to whole scripts by independent examiners to be less than 3.3 marks. It is also worth noting that this is an average difference. Broadly speaking, half of the differences between markers would be greater than this and half would be less.

Although the RSSIM is relatively easy to calculate, it is not particularly intuitive. To simplify matters Table 3 shows how, according to this analysis, the expected best possible absolute marking difference between examiners would vary, as a percentage of the paper maximum, according to the number of items in the paper if all items were worth the same number of marks[6]. This shows that for examination papers where candidates are asked to answer small numbers of items, the lower bound for the average difference between examiners is likely to exceed 5 per cent of the paper total. On the other hand, for papers where candidates are asked to respond to large numbers of items, the average difference may be less than 2 per cent of the paper total.

Also shown is an estimate of a lower bound for the standard deviation of marking differences. This metric is included to allow comparisons with results reported in Ofqual (2018a). By assuming that marking differences follow a roughly normal distribution, this is calculated by multiplying the mean absolute marking differences in the previous column by 1.25[7]. The values in this column are relatively consistent with findings recently reported across all awarding organisations in Ofqual (2018a). For example, Figure 4 on page 14 of Ofqual (2018a) shows that the standard

6. This can be calculated as $100/(12 * \sqrt{Number\ of\ items})$.

7. According to the properties of the half-normal distribution the expected absolute value of a normally distributed variable is the standard deviation of that normal distribution divided by $\sqrt{\pi/2}$ – that is, roughly 1.25. See https://en.wikipedia.org/wiki/Half-normal_distribution for more details.

deviation (SD) of marking differences for Physics components tends to be at around 2 per cent of the paper total. Given that Physics papers typically include roughly 30 items[8], this value is very close to the relevant value reported in the final column of Table 3. Thus, for Physics at least, marking accuracy is already very close to the best that can reasonably be expected of markers.

Table 3: Expected differences between collectively assigned and definitive marks as a percentage of the total mark available for an exam assuming that all items are worth the same number of marks.

Number of items	*Mean absolute marking difference as percentage of paper total*	*SD of marking differences as percentage of paper total*
1	8.3%	10.4%
2	5.9%	7.4%
3	4.8%	6.0%
4	4.2%	5.2%
5	3.7%	4.7%
10	2.6%	3.3%
20	1.9%	2.3%
30	1.5%	1.9%
40	1.3%	1.7%
50	1.2%	1.5%
100	0.8%	1.0%

Discussion

The analysis in this article has shown that marks averaged across multiple junior examiners have greater predictive power than the definitive marks agreed by senior examiners. This is particularly evident in subjects such as Humanities where differences in professional opinion are more likely to arise. By extension, this implies that many junior examiners are better than a PE alone. Building on this evidence, the analysis has suggested a lower bound for the reliability of definitive marks themselves. These results have some practical implications.

Marker monitoring

Given the evidence that, particularly in subjects where professional judgement is an important element of marking, averaged marks are more accurate than definitive marks, it seems reasonable that, where possible, the former should form the basis for evaluating the performance of markers rather than the latter.

It is worth noting that the process of setting definitive marks remains useful as it may help to cement an idea of the marking standard that can be communicated to all markers. Furthermore, in practice, because exam boards require marker monitoring to be operational as soon as marking begins, and cannot wait until sufficient junior markers have marked seed scripts, marker monitoring would still need to work from a set of definitive marks to begin with. However, in rare cases where groups of junior examiners consistently disagree with the definitive mark, the evidence here suggests that the original definitive mark could reasonably be overwritten with the mean of the marks from AEs. At present where such situations occur, PEs are asked to review the original definitive marks to verify that they are the marks they intended. In practice, they can prove unwilling to shift their opinion of the correct mark for a script. The evidence in this article suggests that, since collectively assigned marks from AEs are generally better than those from a single PE, unless there was other evidence of a problem with the panel of AEs, the original definitive mark could be legitimately overwritten without the agreement of the PE. That is, if AEs as a group generally agree with the PE, then, in the rare cases where they disagree, the opinion of many AEs should take precedence.

Of course, if definitive marks were overwritten, it would be beneficial to ensure that any truly aberrant markers were excluded from this calculation. Alternatively, a robust mean, estimated using statistical methods that are resistant to outliers, could be employed.

Post-series evaluation

Recognising that marks from PEs are unlikely to provide an absolute truth with regard to the correct mark that each script should be awarded, and also noting the relatively small number of seed scripts used in analysis, has important implications for the use of such data for reviewing marking quality at a national level. In particular, it indicates that findings with regard to the overall accuracy of marking for individual exam components based on this data should be treated with some caution. Thus, although it may be desirable that analysis of such metrics will allow exam boards to "channel additional resource and support to those components or qualifications which most need improving" (Ofqual, 2018a, p.35) it is also possible that such metrics could be subject to random variation from year to year depending upon the exact selection of seed scripts.

Furthermore, since the evidence suggests that collectively assigned marks are superior to definitive marks, it would make sense for the post-award marking metrics generated by Ofqual to be based around the former rather than the latter. That is, rather than focusing on differences with definitive marks, these metrics should evaluate differences from the (more accurate) collectively assigned marks.

Marker accuracy

Finally, by looking at the difference between (very accurate) marks from PEs and (even more accurate) marks defined collectively by groups of AEs, we can derive a lower bound for the extent of difference we should expect between examiners. Furthermore, this article has shown how this lower bound for marking accuracy relates to the structure of the exam paper. The relevant formulae in this paper may help to set the tolerances for marker monitoring.

These results are important as they may help to manage expectations regarding the level of consistency between marks that it is possible to achieve through training and increased experience. After all, as we have seen, it is likely that even the best and most experienced markers, the PEs themselves, display some level of inconsistency for examination questions requiring professional judgement. If we are dissatisfied with the estimated levels of agreement between (principal) examiners suggested in this paper then the route to improve reliability is unlikely to lie in greater marker training or making slight tweaks to mark schemes. To dramatically improve reliability, far-reaching changes, such as altering the structure of assessments, or the number of markers who mark each exam, may be required.

Kathleen Tattersall, the very first chief regulator of Ofqual, warned against simplistic expectations that the marking system should be perfect (Clark, 2008). The research in this article provides rather more

8. Based on a quick review of a number of GCSE Physics papers available online.

detail about the likely scale of imperfection. Having an honest understanding of what can be achieved is important if we are to ensure that the demands placed upon an assessment system are realistic.

References

Benton, T. (2017, November). *Pooling the totality of our data resources to maintain standards in the face of changing cohorts.* Paper presented at the 18th annual AEA-Europe conference, Prague, Czech Republic.

Benton, T. (2018, November). *Exploring the relationship between optimal methods of item scoring and selection and predictive validity*. Paper presented at the 19th annual conference of the Association for Educational Assessment in Europe, Arnhem-Nijmegen, The Netherlands.

Benton, T., & Gallagher, T. (2018). Is comparative judgement just a quick form of multiple marking? *Research Matters: A Cambridge Assessment publication, 26*, 22–28.

Black, B., Suto, I., & Bramley, T. (2011). The interrelations of features of questions, mark schemes and examinee responses and their impact upon marker agreement, *Assessment in Education: Principles, Policy & Practice, 18*(3), 295–318.

Bramley, T. (2008, September). *Mark scheme features associated with different levels of marker agreement.* Paper presented at the annual conference of the British Educational Research Association, September 3–6, in Edinburgh, UK.

Clark, L. (2008, May 15). Pupils should not expect to receive 'accurate' results from exams, warns new watchdog. *Mail Online*. Retrieved 29th May, 2019, from http://www.dailymail.co.uk/news/article-1019938/Pupils-expect-receive-accurate-results exams-warms-watchdog.

Ofqual. (2016). *Marking consistency metrics*. Ofqual report, Ofqual/16/6121. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_da a/file/681625/Marking_consistency_metrics_-_November_2016.pdf.

Ofqual. (2018a). *Marking consistency metrics: An update*. Ofqual report, Ofqual/18/6449/2. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_da a/file/759207/Marking_consistency_metrics_-_an_update_-_FINAL64492.pdf.

Ofqual. (2018b). *A study of hard-to-mark responses: Why is there low mark agreement on some responses?* Ofqual report, Ofqual/18/6449/5. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_da a/file/759216/HardtoMark_-_FINAL64495.pdf.

Suto, I., Nádas, R., & Bell, J. (2011). Who should mark what? A study of factors affecting marking accuracy in a biology examination. *Research Papers in Education, 26*(1), 21–51.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S. Fourth Edition*. New York: Springer.

"Learning Progressions": A historical and theoretical discussion

Tom Gallacher GL Assessment[1] and **Martin Johnson** Research Division

Introduction

Learning Progressions (LPs) are a relatively recent approach that aim to support three aspects of education: teaching and learning, assessment, and curriculum design. According to Schmidt, Wang, and McKnight (2005) the effectiveness of these three aspects of education may be increased by better coherence, and the LP approach claims to improve coherence by providing frameworks of knowledge and skills called "LP models". These frameworks describe the progression that can be expected of learners through their education (Gotwals & Songer, 2013). LP approaches are popular and influential across the fields of education and curriculum development, with discussion being carried out across a number of international contexts (Australian Council for Educational Research, 2018; E. M. Kim, Haberstroh, Peters, Howell, & Nabors Oláh, 2017; H. Kim & Care, 2018). This suggests that the consideration of the approach is topical.

This article outlines the specific objectives of the LP approach, the mechanism by which LP models may attain these objectives, and finally, the likelihood of this attainment (based on previous evidence). LPs should only be expected to achieve their aims if the assumptions of the LP approach are correct; however, our view is that the evidence suggests that the assumptions embedded within the frameworks are overly simplistic. Education is complex and the implementation of the LP approach to teaching and learning, assessment, or curriculum design may have unintended consequences when implemented without consideration of other possible approaches.

Proponents of the LP approach display a minimal engagement with previous theories of learning, and their ideas have been criticised as being "the latest manifestation of a much older idea, that of regularity in the development of students as they learn a certain body of knowledge or professional practice" (Wilson, 2009, p.716). This suggests that LP proponents should also consider the similarities of their theory with previous work to derive an approach that is most likely to attain its desired objectives.

Objectives of LPs

In order for LPs to benefit teaching and learning, assessment, and curriculum design, the approach needs to have a theory of learning that satisfies the practical and theoretical demands of the professionals involved in all three areas (Black, Wilson, & Yao, 2011).

1. The work was carried out when the first author was a member of the Research Division.

Moreover, this central theory needs to be robust to the criticism of evidence so that it can satisfy its objectives of explaining important phenomena. As stated, the LP approach aims to improve teaching and learning, assessment, and curriculum design, by providing frameworks that model the process of change that learners go through when engaging in education. These frameworks cut across the three aspects, and are claimed to have benefits to each independently:

Firstly, in order to benefit teaching and learning, the LP framework aims to provide detailed instruction on the optimal order for presenting material within a subject. This structure can then support lesson planning, helping teachers to track student progress and identify actions that support the learners' learning (Alonzo & Gearhart, 2006, p.100).

Secondly, to support assessment, the LP approach aims to provide a framework for comparing different learners in order for the results of such comparisons to be useful for learners (Catley, Lehrer, & Reiser, 2005). This framework would also provide a validity argument for assessments (Gotwals & Songer, 2013). In addition, changing the emphasis of assessments so that they are demonstrations of problem solving that correspond to the way that an expert behaves (called "learning performances"), would provide rich and useful information on the abilities of learners (Coppola, 2006).

Thirdly, to support curriculum design, the LP approach aims to provide a method of refining the material presented to learners (Corcoran, Mosher, & Rogat, 2009). By empirical observation and research, a curriculum may be optimised to enable learners to derive the best possible education (Smith, Wiser, Anderson, & Krajcik, 2006).

The commonalities between these three areas are collectively referred to as the theory of learning that is prescribed by the LP approach (Black et al., 2011). It is this set of assumptions that can be tested against the evidence already accumulated within fields of teaching and learning, assessment, and curriculum. If the theory of learning is not contradicted by the previous findings, then we can presume that the theory works well enough to suggest that the implementation of an LP approach would be useful for achieving the stated aims. If the theory cannot account for previous findings, or worse, predicts the opposite, then we can conclude that the theory developers would benefit from more engagement with prior literature.

In our critique of the core issues around the LP approach, our intention is to contribute to the debate around conceptualisations of learning progressions and to suggest that it is also important to look at other areas of curriculum theory for insight.

An outline of the LP approach's theory of learning

As mentioned, the theory of learning that is prescribed by the LP approach underpins the three aims of the approach since it is this theory that allows the coherence between the three areas. The LP approach can broadly be summarised into four points:

- LP models are domain-content specific. Subjects like Science or Mathematics have distinct ways of thinking and distinct bodies of material that need to be taught to be understood (e.g., Smith et al., 2006). This means that there are central concepts and principles of a discipline, which ties the area to the notion of coherent "big ideas" (e.g., see Harlen, 2010).
- LP models incorporate knowledge and practice. Learners learn (and demonstrate through assessment) the "what" and the "how" of a subject domain-content (e.g., Lehrer & Schauble, 2015; Wilson, 2009).
- LP models are successive and progressive. If learners fail to master a particular first thing, they are not able to do a specific second or third thing (e.g., National Research Council, 2007).
- LP models are based on research about what learners can do at different ages and stages of progression (e.g., Duncan & Hmelo-Silver, 2009).

At the simplest level, therefore, LP theory analogises learning a subject to be like climbing a ladder. Climbing each rung is dependent on climbing the previous rung, and it allows the climber (learner) a better view of the subject. Therefore, the main priority for LP developers is to design the ladder appropriately.

For illustrative purposes, a simple LP model of two ideas ("matter" and "colour") is conceptualised (by the authors) and presented in Figure 1. "Naïve understanding" is taken as the starting point, so these learning progressions start with learners who have no prior empirical or "scientific" understanding (in a conventional sense) of matter or colour, but might have a variety of views about what matter and colour are.

In this article we argue that the LP theory of learning is made up of four simplifications, which we can unpick to begin evaluating whether the adoption of the LP approach at the expense of other approaches will meet the aims above. A lot of the arguments presented have not been addressed since the advent of previous theories of hierarchy development (Phillips & Kelly, 1975).

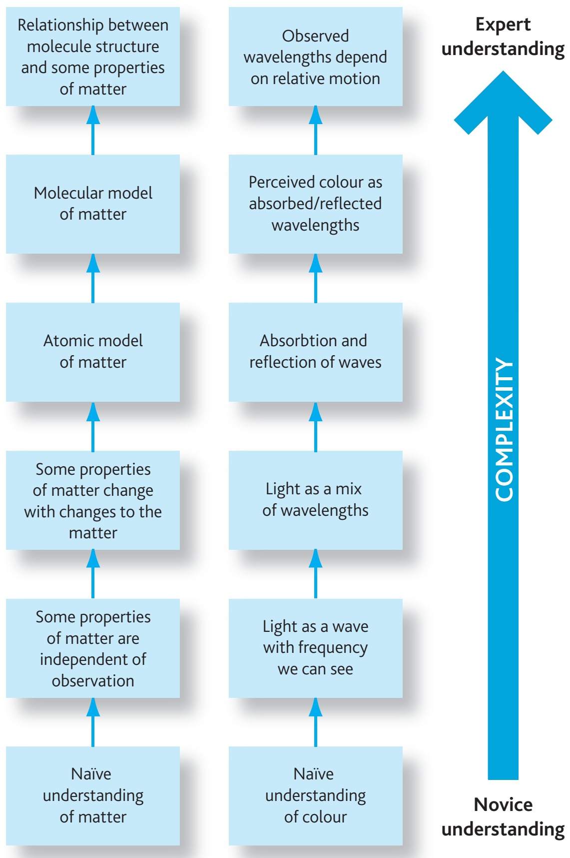


Figure 1: Illustrative Learning Progressions model of matter and colour (authors' own example)[2]

2. Please note that that the progressions presented here are open to discussion, for example, see Taber (2000) and his commentary on the "pedagogic pitfalls of the atomic ontology".

Simplification 1: LPs include a mix of cognitive processes

The first simplification inherent in the theory of learning proposed by the LP approach stems from the second point above: that knowledge and practice should be combined or bundled into a unit of "concept". Moreover, a learner's grasp of this concept is a part of "the developmental pathway in which students' understanding…become[s] more sophisticated over a long period of time" (Paik, Song, Kim, & Ha, 2017, p.4965). In the analogy of a ladder, these bundled concepts comprise the rungs of the ladder, since they are on the same level. Despite the legitimate concern that models can lack the sophistication to describe complex realities (e.g., Goldstein, 1998), some theorists have employed the metaphor of a ladder to exemplify the learning process (e.g., Hess, 2008; Masters & Forster, 1996; Vorst, 2018). Our concern is that the conflation of knowledge and practice into a concept may lead to an insecure inference about what a learner "knows" based on their performance.

Catley et al. (2005) are very explicit about how concepts implicate the bundling of different activities together, stating that "we represent this blend of knowledge, skills and forms of activity that support the development of knowledge and skill as learning performances" (2005, p.5). Other authors are less explicit, but make some reference to differentiating cognitive processes, since all parts of understanding are "enacted" by a "learning performance" with the material (Smith et al., 2006).

Downplaying the differences in cognitive processes into one single unit allows a potentially problematic assumption about what learners can and cannot do. Any successful performance with learning materials can be taken to indicate, according to the LP ladder analogy, that the learners can demonstrate successful performances when the material is presented in different ways. This is because the learner is assumed to have gained understanding (have climbed that rung of the ladder). This assumption, as will be seen, is not always true or useful for educators in practice.

Some LP authors seem to have a preference for some cognitive processes over others, such that declarative memory recall is negatively contrasted by Smith et al. (2006, p.93) with "important aspects…of understanding and reasoning." Other authors are less explicit, except that by emphasising the development of problem-solving skills, there is little mention of developing the knowledge required by the beginner levels (Messick, 1984, p.216) where "in beginning or low-level achievement a major issue is the acquisition of a critical mass of information on the subject," with more advanced levels reflecting more complex cognitive processes.

Simplification 1 Evidence

Firstly, the claim that different processes can be effectively bundled together is considered a simplification given evidence from how the development of different processes happen at different points, and are likely to happen in cycles (Fischer, 2008). The specific cognitive development of learners might enable them to perform some tasks with the material, but not all, while failure at a task might be due to the failure of several different cognitive processes. This makes it difficult for teachers to identify how to help different learners who have a range of successes at different tasks. If learners are not consistently able to demonstrate a "concept" then the ladder may be a dangerous analogy on which to base decisions.

Within alternative theories of learning, memory is typically specified as crucial to "higher" application of knowledge, and so in Bloom's Taxonomy of Educational Objectives (Bloom, 1956) the cognitive processes are shown as a hierarchy or network with memory at the bottom. More recently revised versions of the taxonomy, such as by Anderson and Krathwohl (2001) and Webb (1997, 1999), retain a distinction between memory and problem solving. While recognising the foundational nature of memory for "higher" processes, years of teacher observation data suggests that memory, although insufficient of itself, is a necessary requirement for higher cognitive functions.

Studies of experts and non-experts show crucial differences in how memory – rather than understanding – is changed by learning (e.g., National Research Council, 2000, 2001), and a theory of learning that downplays these changes will not be able to account for such evidence. For these reasons, if LPs were to be implemented as a system of learning, we would expect insufficient consideration of the different cognitive processes that support learning, and therefore that the implementation might provide ineffective education. When "Assessment without levels" was introduced in England, the system "encouraged undue pace and progression onto more difficult work while pupils still had gaps in their knowledge or understanding" (Department for Education, 2015, p.17).

This simplification has focused on the rungs of the ladder, and is essential to understanding the next two simplifications, as will be seen.

Simplification 2: Hierarchies of concepts

The second simplification inherent in the theory stems from the third "successive and progressive" point, such that within a subject domain there is a hierarchy of "understandings" that proceed over the course of learning a subject. In the analogy of the ladder, some rungs are higher than others, with each successive rung being higher than the same set of previous rungs, plus one more. Within each LP, some concepts are more advanced, and are therefore closer to the concepts of experts within that field.

The purpose of education within the LP approach is to bring the understandings of non-experts closer to that of experts (Duschl, 2006; Lehrer & Schauble, 2015). In addition, the role of teachers is to mediate the material and to scaffold the learning so that learners are brought closer to the end goal of the LP (Duncan & Hmelo-Silver, 2009).

This simplification is an attractive one, since it implies the simple progression derived from learning (Fensham, 1994). Once a leaner has progressed beyond a stage of learning, their understanding of a subject is closer to that of an expert, and therefore they are able to solve more problems than before and are ready for the next stage. It has been noted that high performing educational jurisdictions incorporate such a process (Valverde & Schmidt, 1998), which might support a conclusion that such a process is useful for learning. However, this model assumes that subject experts have a monolithic set of concepts to be worked towards, that might not be applicable to all subject areas (Bernstein, 1999), while teaching such a view might damage the process of later learning (Efland, 1995).

Simplification 2 Evidence

The evidence for hierarchies of knowledge and skills is mixed. Gagné (1968) reviews the evidence to support the idea of "Learning Sets", which supports a theory of learning that claims that the optimal ordering of material can be found empirically. Like LPs, the theory is implicitly Vygotskian, in that the main determinant for whether a material can be learnt is the prior learning and knowledge, rather than any formal stage of cognitive development. For illustrative purposes, a Learning Sets curriculum of matter and colour that builds toward one idea is presented in Figure 2. Notice that Learning Sets allows connections between parallel curricula, allowing the possibility of a network analogy, rather than the ladder prescribed by Learning Progressions.

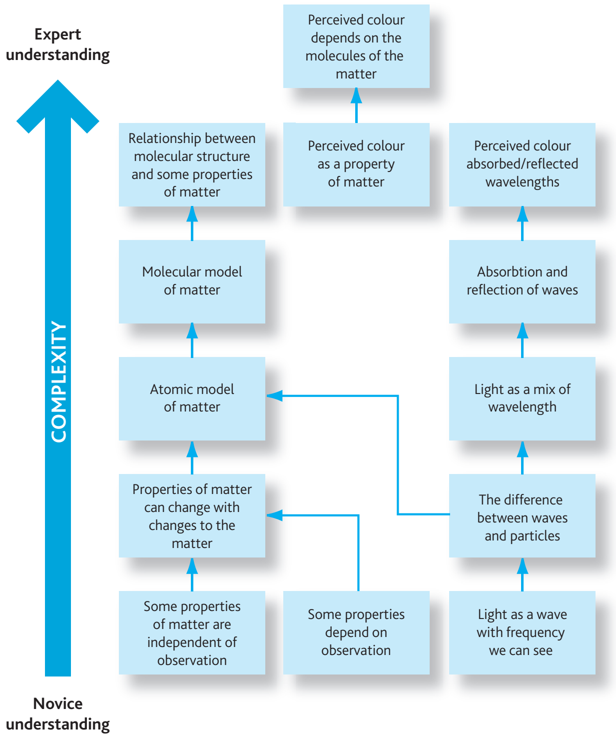


Figure 2: Illustrative Learning Sets model of matter and colour (authors' own example)

Gagné & Bassler (1963) found that the forgetting of subordinate learning sets may occur independently of, and without effect upon, retention of the total task which has been achieved through learning. This forgetting has implications for assessment if discrete task performance at a particular time is taken as a signifier of learning or ability. This undermining of the theory was also confirmed by Kolb (1967), particularly for declarative knowledge. This suggests that although an optimal method of presenting material could theoretically exist, the order will not determine the retention of material by the learner, undermining the analogy of a ladder since rungs are not retained in the order they were climbed. More recent evidence goes one step further to show that intermediate learners are dependent on the context and presentation of problems to guide how they apply the skills that they have learnt (Bao, Hogg, & Zollman, 2002). This inconsistency of applying a skill or knowledge generalises across different cognitive activities from declarative memory recall to problem solving beyond secondary education. Future skill development has been argued to *require* inconsistency as part of the process of consolidating learning (Fischer, 2008). This suggests that the trajectory of learning is less like a ladder, but more "three steps forward, one step backwards", and suggests that despite its use a s a model for learning (e.g., Hess, 2008; Masters & Forster, 1996; Vorst, 2018), the ladder analogy is inadequate to describe the complexity of the learning process.

Applications of the LP approach that adhere to the ladder-progression analogy are likely to be problematic if they do not consider the contexts in which different learners can or cannot demonstrate a technique appropriately. Such a problem has been observed in England where "teachers planned lessons which would allow pupils to learn or demonstrate the requirements for specific levels. This encouraged teachers to design and use only classroom assessments that would report a level outcome. As a result, formative classroom assessment was not always being used as an integral part of effective teaching" (Department for Education, 2015, p.13). It may be that the conclusion from Valverde and Schmidt (1998) failed to identify other differences between jurisdictions that contributed to the observed high performances in those jurisdictions.

Another problematic issue that may pertain to the ladder-progression analogy relates to learner equity. Lehrer and Schauble (2015) note that conforming to generalised learning models may restrict the landscape of possibilities and deprive students the opportunity of (a) encountering concepts that have traditionally been considered too difficult to learn, but which can be made accessible through appropriate teaching (White & Frederiksen, 1998); or (b) lead teachers to fail to consider that some ideas that are presumed to be self-evident may turn out to be more challenging when encountered from a student's perspective (Sandoval & Millwood, 2005).

Simplification 3: Assessment of progression

The third simplification rests on the previous two simplifications, and states that the stage of a learner's progression towards a goal can be assessed reliably, and reported to learners and teachers in order to support the overall education process (Steedle & Shavelson, 2009). Results from assessments are taken to be a reflection of the concepts grasped by a learner (how far up the ladder they have climbed), and their degree of expertise. Learners are positioned not relative to each other, but are located on the ladder of progression (Corcoran et al., 2009).

Simplification 3 Evidence

The idea that progression through levels can be reliably assessed is a simplification since learners can inconsistently demonstrate a range of abilities that do not support a hierarchy based solely on conceptual difficulty. Hart (1981), as cited in Simons and Porter (2015), shows examples of students who can demonstrate an ability when asked one way, but not another. For example, in the case of the conceptual difficulty of "knowledge of fractions", 90 per cent of students can respond that 5/7 is greater than 3/7, but only 15 per cent can respond that 5/7 is greater than 5/9. This difference exists despite the assumed conceptual commonality of denominator and numerator knowledge that underpins fraction knowledge. Differences have also been found between students' performances on the same skills depending on whether they were assessed by a class test or by an individual interview (Denvir & Brown, 1987, p.106).

The same type of conclusion was reached by proponents of LP assessments. Graf and van Rijn (2016) report that the likelihood of a learner successfully completing a task related to three things: 1) degree of progression through the learning pathway; 2) non-progression related complexity, such as computation (i.e., systematic sources of difficulty that covary with the levels of the progression but which are not specified by the learning pathway through conceptual complexity); and 3) sources of difficulty which are not related to the levels of the progression, such as reading demands. Alonzo and Steedle (2009) observed that students' responses were only 60 per cent consistent within one level of a learning progression, with some of the rest of the variance being explained by features of the items.

This evidence also goes some way to discredit the first simplification: that learners routinely solve "easy questions" on "advanced concepts" and fail "hard questions" on "basic concepts". This means that not all parts of a concept are a single unit, and that teaching and assessment need multiple dimensions to understand variance in performance rather than the unidimensional ladder analogy. Feedback which reflects a learner's inconsistencies would arguably be more useful for planning future educational activities required by educators.

Attempts to make grades from summative assessments in England reflect the trajectories of progression came to the same conclusion in the 1980s: "The larger obstacle appeared in the distribution of performances of students. Each set of criteria of this type had to assume a model of students' progress…and students' performances did not fit these models" (Black & Wiliam, 2002, p.25). Similar concerns have also been voiced more recently over the model of progression implied by "Single Level Tests" (Whetton, 2008) and National Tests (Oates, 2011).

Simplification 4: Big Ideas

A fourth simplification within the LP theory of learning is the generalisation of learning from specific "big ideas" to the breadth of a subject area. The claim here is that learners who gain knowledge and skills from one area of a subject are able to apply these to untaught areas, if the original area is fundamental enough to that subject.

This claim is most relevant for curriculum designers, who may wish learners to achieve a wide range of objectives in a fixed time. The LP approach argues that learners who master the concepts of a specific big idea are better placed to answer new questions from an unrelated sub-area than are learners who master a breadth of areas to a lesser depth. What is and what is not a "big idea" is not obvious *a priori*, but criteria may include that the teaching of the idea should facilitate understanding of current issues, be satisfying to learn, and have cultural significance (Harlen, 2010, p.19). Although the idea of a "big idea" is found elsewhere (Bruner, 1960, p.18), it is something that is hard to falsify, since evidence of "no transfer" could be taken as evidence that the taught idea was not "*big*" enough.

Bruner (1960) advocated a curriculum where topics are revisited at intervals, with different ways of presenting the topic. The theory takes an explicitly Piagetian view that learners go through stages of representation of ideas, from enactive to iconic to symbolic. The challenge of education, according to the Piagetian principles embodied in such "spiral curricula", is to present ideas in ways that correspond to the developmental stage of the learners (Bruner, 1960, p.39; Efland, 1995). When learners are ready for the next stage of representation, educators need to present the same topic correspondingly, which thereby allows a more complex understanding of the topic for the learner. This idea echoes the thinking of LPs, but with an explicit theoretical understanding of knowledge development. For illustrative purposes, the LPs for matter and colour that were used in previous figures are produced within a spiral curriculum framework (Figure 3). Here, the connection between the nodes is unimportant because each node is *qualitatively* distinct from those coming before and after.

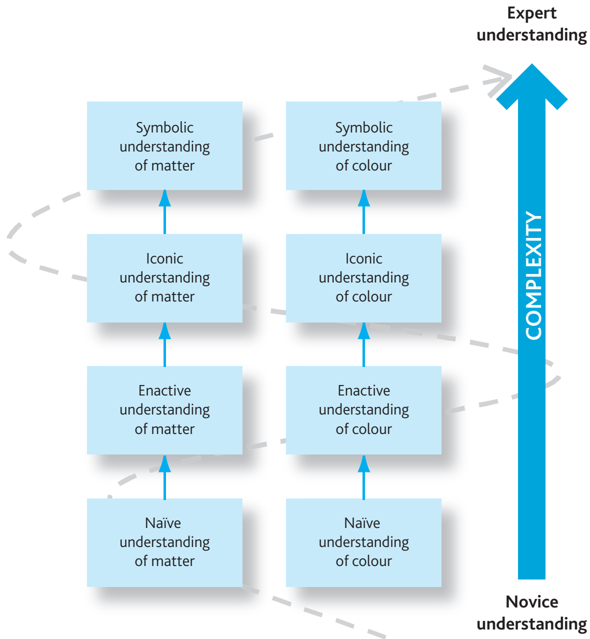


Figure 3: Illustrative Spiral Curricula model of matter and colour (authors' own example)

The concept of "big ideas" in the modern sense arose from studies of experts' thinking (National Research Council, 2007, p.37), where it describes the ways that experts group problems that they have seen and how they identify new problems in relation with these problem groups (Chi, Feltovich, & Glaser, 1981). This does not, however, logically support the idea that reducing the breadth of a subject to increase depth of study improves the transfer of knowledge beyond the types of problems presented. The experts studied to support the existence of "big ideas" had themselves come through a broad curriculum before the study, where the groupings might only emerge with sufficient exposure to a breadth of problems encountered, rather than being directly teachable.

The design of LP curricula is taken to be evidence based, so studying the abilities and concepts of experts, learners, and novices is the method to suggest useful changes to the order of an LP curriculum.

Simplification 4 Evidence

Given the inconsistency of learners' application of knowledge within a subdomain of content, it is unlikely that concepts (skills, knowledge, etc.) should be applied consistently across a subject. There is little evidence in the literature relating to transfer of learning across domains to support a claim that reducing the breadth of a subject will improve transfer (Salomon & Perkins, 1989). Secondly, the role of evidence in building effective curricula predates the debates that have been spurred by the LP approach (e.g., Black & Simon, 1992; Bruner, 1960, p.19; Department of Education and Science/Welsh Office, 1987, p.37).

Conclusion

The findings we have presented suggest that the theory of learning that is inherent to the LP approach is unhelpfully simplistic. This is because the theory does not reflect the inconsistencies and complexities of the actual process of change that learners go through, or how inconsistently they can demonstrate their learning. Therefore, implementation of a theory based on the ladder analogy, replacing other approaches and models of learning, is likely to be counterproductive for learning since learners are never on one rung of a ladder at a time. This is not to say that no learner makes progression, or that simplifications cannot be useful in some contexts (such as for creating a scheme of work from a curriculum), but that the theory described by "Learning Progressions" authors, if implemented with no extra consideration of curriculum and learning, would not lead to positive educational outcomes.

From the perspective of a subject expert who has made the learning journey through a subject, it may seem that the journey was smooth in retrospect, but this is unlikely to have been the case. Learners develop cognitively as well as neurologically, with performance on solving problems being at best inconsistent during intermediate phases of the journey. The highest demonstrated level of performance from a learner might not be maintained across different contexts, and should not be expected to indicate mastery of activities learned previously, since forgetting can occur independently of the order of presentation.

Our consideration of the simplifications in the underlying learning theory of the LP approach suggests that the three aims of the approach (i.e., to support teaching and learning, assessment, and curriculum design) are unlikely to be met.

Firstly, tracking students' progress and identifying actions to support learning (Alonzo & Gearhart, 2006, p.100) is unlikely to support learning if success or failure at a task is taken as being unproblematic evidence of similar success and failure at related tasks. The breadth of possible learning across contexts needs support, and this will not be provided by an LP approach that considers these activities to be on the same rung of the ladder and therefore unimportant. Learners' progress is not ladder-like, but complex, and so learning activities that only consider conceptual difficulty are unlikely to diagnose students' *particular* weaknesses (Briggs, Alonzo, Schwab, & Wilson, 2006). This point may help to explain why it has been observed that the use of LPs to inform teachers' formative assessment practices has not lived up to expectations (Hammer & Sikorski, 2015).

Secondly, the utility of the framework of levels is unlikely to provide a valid way of comparing learner abilities since demonstrations of ability can be inconsistent (Leahy & Wiliam, 2011, p.5). Assuming that success or failure at one task will mean success or failure at related tasks fails to reflect the complexities of abilities that learners have, and assessments that report level-based outcomes will not be any more useful than currently reported outcomes such as grades reflecting atheoretical levels of attainment (Department for Education, 2015).

Generally speaking, the LP theory of learning, although not spelled out in any consistent way by any author, is similar to those that have come before, such as the Piagetian Spiral Curriculum (Bruner, 1960), and Gagné's Learner Sets (e.g., Gagné, 1968). Many of the criticisms laid out here are modified versions of observations of these earlier theories (Phillips & Kelly, 1975), and LP theory would benefit with engaging more with such work.

References

Alonzo, A. C., & Gearhart, M. (2006). Considering Learning Progressions From a Classroom Assessment Perspective. *Measurement: Interdisciplinary Research and Perspectives, 4*(1–2), 99–126.

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education, 93*(3), 389–421.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. New York: Longman.

Australian Council for Educational Research. (2018). The role of learning progressions in global scales. Retrieved June 18, 2019, from [rd] ACER website: https://rd.acer.org//article/the-role-of-learning-progressions-in-global-scales

Bao, L., Hogg, K., & Zollman, D. (2002). Model analysis of fine structures of student models: An example with Newton's third law. *American Journal of Physics, 70*(7), 766–778.

Bernstein, B. (1999). Vertical and Horizontal Discourse: An essay. *British Journal of Sociology of Education, 20*(2), 157–173.

Black, P., & Simon, S. (1992). Progression in learning science. *Research in Science Education, 22*(1), 45–54.

Black, P., & Wiliam, D. (2002). *Standards in Public Examinations*. London: King's College London.

Black, P., Wilson, M., & Yao, S.-Y. (2011). Road Maps for Learning: A Guide to the Navigation of Learning Progressions. *Measurement: Interdisciplinary Research and Perspectives, 9*(2–3), 71–123.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives Book 1: Cognitive Domain* (2nd edition). New York: Longmans, Green.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic Assessment With Ordered Multiple-Choice Items. *Educational Assessment, 11*(1), 33–63.

Bruner, J. S. (1960). *The Process of Education*. Cambridge, Mass: Harvard University Press.

Catley, K., Lehrer, R., & Reiser, B. (2005). *Tracing a Prospective Learning Progression for Developing Understanding of Evolution*. Washington, DC: National Academy of Sciences.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science, 5*(2), 121–152.

Coppola, B. P. (2006). Learning progressions provide a clear map for designing Standards-based assessment [Peer Commentary on: "Implications of Research on Children's Learning for Standards and Assessment: A Proposed Learning Progression for Matter and the Atomic Molecular Theory." by Smith, C. A.; Wiser, M.; Anderson, C. A.; Krajcik, J.] *Measurement: Interdisciplinary Research and Perspectives, 14*(1/2), 111–115.

Corcoran, T., Mosher, F., & Rogat, A. (2009). Learning Progressions in Science: An Evidence-Based Approach to Reform. *CPRE Research Reports*.

Denvir, B., & Brown, M. (1987). The feasibility of class administered diagnostic assessment in primary mathematics. *Educational Research, 29*(2), 95–107.

Department for Education. (2015). *Commission on Assessment Without Levels: final report*. Retrieved from DfE website: https://www.gov.uk/government/publications/commission-on-assessment-without-levels-final-report

Department of Education and Science/Welsh Office. (1987). The National Curriculum 5–16: A consultation document. London: Department of Education and Science/Welsh Office.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching, 46*(6), 606–609.

Duschl, R. (2006). Learning Progressions as Babushkas. *Measurement: Interdisciplinary Research and Perspectives, 4*(1–2), 116–119.

Efland, A. D. (1995). The Spiral and the Lattice: Changes in Cognitive Learning Theory with Implications for Art Education. *Studies in Art Education, 36*(3), 134–153.

Fensham, P. (1994). Progression in school science curriculum: A rational prospect or a chimera? *Research in Science Education, 24*(1), 76–82.

Fischer, K. W. (2008). Dynamic cycles of cognitive and brain development: Measuring growth in mind, brain, and education. In A.M. Barrio, K.W. Fischer, & P.J. Léna (Eds.), *The educated brain: Essays in Neuroeducation* (pp.127–150).

Gagné, R. M. (1968). Presidential address of division 15 learning hierarchies. *Educational Psychologist, 6*(1), 1–9.

Gagné, R. M., & Bassler, O. C. (1963). Study of retention of some topics of elementary nonmetric geometry. *Educational Psychology, 54*(3), 123–131.

Goldstein, H. (1998). *Models for reality: new approaches to the understanding of educational processes*. Professorial Lecture presented at the Institute of Education University of London. Institute of Education, University of London.

Gotwals, A. W., & Songer, N. B. (2013). Validity Evidence for Learning Progression-Based Assessment Items That Fuse Core Disciplinary Ideas and Science Practices. *Journal of Research in Science Teaching, 50*(5), 597–626.

Graf, E. A., & van Rijn, P. W. (2016). Learning Progressions as a Guide for Design: Recommendations Based on Observations from a Mathematics Assessment. In S. Lane, M. Raymond, & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp.165–189). Abingdon, Oxon: Routledge.

Hammer, D., & Sikorski, T.-R. (2015). Implications of Complexity for Research on Learning Progressions. *Science Education, 99(*3), 424–431.

Harlen, W. (Ed.). (2010). *Principles and big ideas of science education*. Retrieved from https://www.ase.org.uk/documents/principles-and-big-ideas-of-science-education/

Hart, K. M. & Chelsea College. CSMS Mathematics Team (1981). *Children's understanding of mathematics: 11–16*. (reprinted 1986). London: John Murray.

Hess, K. K. (2008). *Developing and Using Learning Progressions as a Schema for Measuring Progress*. Dover, New Hampshire: National Center for the Improvement of Educational Assessment.

Kim, E. M., Haberstroh, J., Peters, S., Howell, H., & Nabors Oláh, L. (2017). A Learning Progression for Geometrical Measurement in One, Two, and Three Dimensions. *ETS Research Report Series, 2017*(1), 1–26.

Kim, H., & Care, E. (2018, March 27). Learning progressions: Pathways for 21st century teaching and learning. Retrieved June 18, 2019, from https://www.brookings.edu/blog/education-plus-development/2018/03/27/learning-progressions-pathways-for-21st-century-teaching-and-learning/

Kolb, J. R. (1967). Effects of relating mathematics to science instruction on the acquisition of quantitative science behaviors. *Journal of Research in Science Teaching, 5*(2), 174–182.

Leahy, S., & Wiliam, D. (2011, April). *Devising learning progressions*. Presented at the Symposium on How to Build Learning Progressions; American Educational Research Association Annual Conference, New Orleans.

Lehrer, R., & Schauble, L. (2015). Learning Progressions: The Whole World is NOT a Stage. *Science Education, 99*(3), 432–437.

Masters, G., & Forster, M. (1996). *Progress Maps*. Melbourne, Australia: ACER.

Messick, S. (1984). The Psychology of Educational Measurement. *Journal of Educational Measurement, 21*(3), 215–237.

National Research Council. (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: The National Academies Press.

National Research Council. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: The National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K– 8*. Washington, DC: The National Academies Press.

Oates, T. (2011). Could do better: using international comparisons to refine the National Curriculum in England. *The Curriculum Journal, 22*(2), 121–150.

Paik, S., Song, G., Kim, S., & Ha, M. (2017). Developing a Four-level Learning Progression and Assessment for the Concept of Buoyancy. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(8), 4965–4986.

Phillips, D., & Kelly, M. (1975). Hierarchical Theories of Development in Education and Psychology. *Harvard Educational Review, 45*(3), 351–375.

Salomon, G., & Perkins, D. N. (1989). Rocky Roads to Transfer: Rethinking Mechanism of a Neglected Phenomenon. *Educational Psychologist, 24*(2), 113–142.

Sandoval, W. A., & Millwood, K. A. (2005). The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cognition and Instruction, 23*(1), 23–55.

Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: an examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies, 37*(5), 525–559.

Simons, J., & Porter, N. (2015). *Knowledge and the Curriculum: A collection of essays to accompany E. D. Hirsch's lecture at Policy Exchange | Policy Exchange*. Retrieved from Policy Exchange website: https://policyexchange.org.uk/publication/knowledge-and-the-curriculum-a-collection-of-essays-to-accompany-e-d-hirschs-lecture-at-policy-exchange/

Smith, C. L., & Wiser, M. (2015). On the Importance of Epistemology–Disciplinary Core Concept Interactions in LPs. *Science Education, 99*(3), 417–423.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). FOCUS ARTICLE: Implications of Research on Children's Learning for Standards and Assessment: A Proposed Learning Progression for Matter and the Atomic-Molecular Theory. *Measurement: Interdisciplinary Research and Perspectives, 4*(1–2), 1–98.

Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching, 46*(6), 699–715.

Taber, K. S. (2000). *Teaching chemistry without (too much emphasis on) atoms?* [A Royal Society of Chemistry Teacher Fellowship discussion paper]. Retrieved from British Education Index website: http://www.leeds.ac.uk/educol/documents/00001559.htm

Valverde, G. A., & Schmidt, W. H. (1998). Refocusing U.S. Math and Science Education. *Issues in Science and Technology, 14*(2), 60–66.

Vorst, H. van. (2018). Structuring learning processes by ladders of learning: results from an implementation study. *Chemistry Education Research and Practice, 19*(4), 1081–1095.

Webb, N. L. (1997). *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*. (No. National Institute for Science Education Research Monograph 6). Washington, DC: Council of Chief State School Officers.

Webb, N. L. (1999). *Alignment of Science and Mathematics Standards and Assessments in Four States*. (No. National Institute for Science Education Research Monograph 18). Washington, DC: Council of Chief State School Officers.

Whetton, C. (2008, September). *The Big New Idea: Single Level Tests*. Presented at the International Association for Educational Assessment Conference, Cambridge.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction, 16*(1), 3–118.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*(6), 716–730.

The impact of A Level subject choice and students' background characteristics on Higher Education participation

Carmen Vidal Rodeiro Research Division

Introduction

Researchers investigating progression to higher education (HE) have suggested that student and school characteristics (e.g., gender, prior academic attainment, social background, type of school) are important factors affecting HE participation and the type of HE institution attended (Chowdry, Crawford, Dearden, Goodman, & Vignoles, 2013; Boliver, 2013; Vidal Rodeiro, Sutch, & Zanini, 2015; Montacute & Cullinane, 2018). This could be in part because certain types of qualifications and/or subjects that are good preparation for HE tend to be taken by young people with higher academic attainment, which is related to social background and to the choices available in their schools (Vidal Rodeiro, 2007; Dilnot, 2016; Gill, 2017).

Over the past few years, policy makers and the general public in England have become increasingly concerned about the extent to which different qualifications and subjects prepare young people for careers or further study. Despite policy efforts and claims of "equivalence", multiple studies have identified ways in which the progression of young people differs depending on the qualifications and/or subjects studied, even after controlling for their background characteristics (Smith, Joslin, & Jameson, 2015; Vidal Rodeiro et al., 2015; Hupkau, McNally, Ruiz-Valenzuela, & Ventura, 2017; Dilnot, 2018; Vidal Rodeiro & Williamson, 2018).

In England, the principal measure of academic attainment for 18 year-old pre-university students is the A Level. Although increasing numbers of university entrants hold other types of qualifications (e.g., applied and technical qualifications) alongside or instead of A Levels, in 2015 73 per cent of the 18 year-olds applying to UK HE institutions did so with just A Levels (UCAS, 2016). In recent years, over 80 different subjects have been offered at A Level. Students can decide which and how many of those subjects they wish to study depending on, for example: their career aspirations, their academic ability, the provision at their school/college or the advice given to them. Students aiming for university typically study three or four subjects at A Level.

Choosing A Levels, however, is not straightforward. There is, for example, a disparity in the attitudes of HE admissions staff towards some subjects, which can lead to low A Level take-up in them. There are also prejudices, amongst the general public and other stakeholders, about the value or usefulness of certain subjects for certain areas of HE study. In addition, whilst many HE courses are open to different A Level subjects and combinations of A Level subjects, others require specific A Levels. Furthermore, some institutions have their own lists of "preferred" subjects. The Russell Group (a self-selected group of research intensive and highly selective institutions) had, from 2011 to 2017, published an annual guide to A Level choice known as *Informed Choices* (Russell Group, 2017). In this guide, they advised students to study at least two from a list of "facilitating subjects", which would leave their options open for a variety of courses. However, they acknowledged that this advice would not apply to all students, and those who were definitely intending to study certain specialist courses such as Music would be best served otherwise. Recently, the Russell Group has re-launched their informed choices guide. The new guidance, the *Informed Choices interactive website* (http://www.informedchoices.ac.uk/), supports less advantaged students, who may not always receive the same level of advice as their better-off peers. This new tool allows students to see not only the subjects that are recommended for specific degrees, but also to test combinations of A Levels to see which degrees they open up. Outside the Russell Group, there is less information available to prospective applicants to guide subject choices.

A great deal of the research carried out into progression to HE in England in recent years has focused on the ability of traditional (academic) versus non-traditional (vocational) qualifications to support students' progression (e.g., Hayward & Hoelscher, 2011; Chowdry et al., 2013; Vidal Rodeiro et al., 2015; McCoy & Adamson, 2016; Hupkau et al., 2016). However, little work on the role of A Level subject choice in access to HE (and different types of HE) or on how students' backgrounds interact with A Level choices to influence HE participation has been published to date. Amongst this small body of research, Vidal Rodeiro and Sutch (2013) investigated, using data from UCAS, the proportions of students who held each A Level subject when applying for a place at university. The outcomes of that research provided some evidence about the usefulness of specific A Level subjects or combinations of A Level subjects as currency for university study. In a more recent study, Sutch, Zanini, and Vidal Rodeiro (2016) examined how students' choice of A Level subjects and attainment influenced their HE destinations. The statistical analyses carried out in their research revealed that there was a relationship between A Level subject specialism and the type of university attended. Dilnot (2018) examined the relationship between league table score of university attended and A Level subject choices. She found that holding more "facilitating" A Levels was associated with attending a higher ranked university, even when A Level performance, prior attainment at General Certificate of Secondary Education (GCSE), and school type were accounted for.

As progression to HE continues to be a matter of interest not just from a research point of view but also for students, HE institutions, awarding bodies and policy makers, a better understanding of how A Level subjects are used to access HE (and different types of HE institutions) is important. The main aims of this research were, therefore, to investigate:

1. the proportions of students who hold different A Level subjects (or combinations of A Level subjects) when enrolling for a degree at a HE institution; and

2. how students' backgrounds interact with the choice of A Level subjects to influence the type of HE institution attended.

This article highlights some of the key findings from a wider research project looking at how useful A Level subjects are for gaining admission to HE. The research is described in detail in Vidal Rodeiro (2019).

Data and methodology

This study followed a cohort of Year 13 students in English schools/colleges through the first year of their HE studies using data from two sources. National Pupil Database (NPD) extracts provided information on A Level subjects and attainment, prior attainment (e.g., GCSEs) and students' characteristics such as gender, school type and income-related deprivation. Data from the Higher Education Statistics Agency (HESA), including HE institution and subject of HE course for all full-time first-year undergraduates, was linked to the NPD.

The students in this research were 17 or 18 years old at the beginning of the academic year 2015/16 and achieved at least one A Level, graded A*-E. All A Levels achieved by these students, independently of the year in which students certificated, were included in the analyses. Note that AS levels were not considered. The size of this A Level cohort was 276,703.

Just under 160,000 students in the A Level cohort appeared in the HESA student records for the academic year 2016/17. The A Level students who were not in the HESA data might not have applied to study in a HE institution, they might not have been offered a place at a HE institution, or they might have taken a gap year. For example, in 2017, the HE acceptance rate for A Level students was 89% and for students taking a combination of A Levels and Business and Technology Education Council (BTEC) qualification, it was 87% (UCAS, 2017a). Furthermore, 21,820 students aged 17 or 18 deferred (i.e., applied for a course and then took a year out before going to university) their university entry (UCAS, 2017b). This corresponded to 7.9% of the acceptances in that age group. It is also worth noting at this point that the linking between NPD and HESA data was done by name, date of birth and postcode, so some A Level students might have been lost in the matching process.

Different combinations of A Level subjects were used in the analyses carried out in this research. A Levels were classified as "facilitating" and "non-facilitating" as suggested by the Russell Group in their previous guidance (Russell Group, 2017). In addition, A Levels were classified using content-based groups (e.g., Applied; Expressive; Humanities; Languages; STEM [Science, Technology, Engineering and Mathematics]) as shown in Bramley (2014).

Different classifications of the HE institutions were also used:

- HE institutions were considered in two groups: Russell Group[1] and "Other" universities (newer universities and colleges, which are usually recruiting institutions or universities with former "polytechnic" status). Some analysis focused on Oxford and Cambridge specifically.
- The Complete University Guide (https://www.thecompleteuniversityguide.co.uk/league-tables/rankings) produces the most comprehensive independent rankings of the HE institutions in the UK. In this research, the overall ranking and the rankings by student satisfaction, research quality and graduate prospects were considered to group the HE institutions. Each of the measures was used to divide institutions into three approximately equally sized groups: low, medium or high ranking.

Data on students' characteristics, prior attainment at school (in Year 11) and performance at A Level was obtained from the NPD. In particular:

- *Prior attainment* was measured by the average GCSE and equivalent point score per entry (GCSE grades were converted into points as follows: A*=58; A=52; B=46; C=40; D=34; E=28; F=22; G=16).
- *A Level performance* was measured by the A Level points in the best (up to) three A Levels (A Level grades were converted into points as follows: A*=60; A=50; B=40; C=30; D=20; E=10).
- *Schools* were classified in two groups: independent schools and state-maintained centres (the latter includes academies, comprehensive schools, grammar schools, secondary modern schools, sixth form colleges and further education centres).
- The *income related level of deprivation* that a student experiences was inferred using the Income Deprivation Affecting Children Index (IDACI), which measures the proportion of children in the immediate neighbourhood living in low-income families.

Together with descriptive statistics, which can show the popularity of A Level subjects in relation to HE participation, multilevel logistic regressions were used to study the likelihood of students with different A Level specialisms (two or more A Levels in a subject area) to study in specific HE institutions once their characteristics (e.g., gender, prior/concurrent attainment, previous institution type, socio-economic background) had been accounted for.

Results

Uptake of A Level subjects

The most popular A Levels amongst university students were Mathematics, Psychology, Biology, History, Chemistry and English Literature. However, these subjects were represented in different proportions in HE and, particularly, in different institutions. For example, Mathematics was taken by 31% of the university students, by 48% of the students in Russell Group universities and by 67% of the students in Oxford/Cambridge.

Students in Russell Group institutions and in Oxford/Cambridge in particular, held A Levels in STEM subjects and in Modern Foreign Languages in higher proportions than students in other universities. In particular, around 60% of the students who obtained an A Level in French enrolled in a Russell Group university (almost 10% were at Oxford/Cambridge). Similarly, almost 75% of those with an A Level in Further Mathematics enrolled in an institution in the Russell Group (14% in Oxford/Cambridge).

There was also variation in the uptake of Applied subjects (e.g., Design & Technology, Art & Design, Business Studies, ICT, Media Studies) and Humanities (e.g., Psychology, Sociology) between different types of HE institutions. Overall, lower proportions of students in Russell Group institutions and in Oxford/Cambridge held A Levels in Applied and Humanities subjects than students in other institutions. There was also variation on the popularity of the A Level subjects by the different university rankings. For example, STEM subjects were more popular

1. A full list of universities can be obtained from the HESA website (https://www.hesa.ac.uk/) and the members of the Russell Group can be identified in the group's website (https://www.russellgroup.ac.uk/). The Russell Group includes Oxford and Cambridge universities.

amongst students in institutions of high research quality and high graduation prospects than in institutions with lower rankings in these areas. However, Biology and Chemistry were more popular amongst students in institutions with low student satisfaction than in institutions rated high by their students. The opposite patterns were found in subjects such as Physical Education or Law.

Uptake of individual A Level subjects by degree subject area was also investigated. The subject of study at university was provided in a list of 19 broad degree areas, which related to the principal subject of the student's qualification aim (see https://www.hesa.ac.uk/support/documentation/jacs/hesa-codes for details). Results from these analyses, show, for example, that Mathematics was taken by 99% of the students accepted to pursue a degree in Mathematical Sciences, 29% of the students accepted to subjects allied to Medicine degrees and by 10% of the students accepted to study Languages. On the other hand, Business Studies was taken by only 38% of the students accepted to study a degree in Business and Administrative Studies, and French or Spanish were taken only by 14% and 11%, respectively, of the students enrolled in a Language degree. The Languages degree area includes courses, among others, in Linguistics, Literature, English, American Studies, Celtic Languages, Literature and Culture, Latin, Ancient Greek, Classics or Languages Studies. Therefore, it is possible that a student enrolled in a Language degree without an A Level in French or Spanish.

For more details on the uptake of individual A Level subjects and combinations of individual A Level subjects, overall and broken down by type of HE institution and degree subject area, see Vidal Rodeiro (2019).

Just over 72 per cent of the A Level cohort had three or more A Levels, but this proportion was higher among students who were accepted onto a university course (79 per cent). Note that some students may have held other qualifications in addition to A Levels, such as BTECs, Cambridge Technicals, or Extended Project qualifications.

The number of A Level subjects held by students varied across the different types of HE institutions (see Table 1). For example, students at Oxford/Cambridge held, on average, the highest number of A Levels and students attending low ranking institutions the lowest. Similar patterns were found for A Levels in facilitating subjects. For example, students attending Oxford/Cambridge and Russell Group institutions held, on average, the highest number of facilitating subjects and those attending institutions with a low research quality ranking, or institutions with low graduation prospects, held the lowest.

A Level students were assigned to an A Level specialism using the A Level taxonomy described in Bramley (2014). Students were considered specialists in one area if two or more of their A Levels were in the same subject area. If a student had two A Levels in at least two categories, they were assigned to a "Multi" category; if they did not have at least two A Levels in any single category, they were assigned to a "None" category. More details about how the students were assigned to specialisms are available in Sutch et al. (2016).

Figure 1 below shows that there were higher percentages of students specialising in Humanities, Languages or STEM subjects at A Level in the group of students that enrolled in HE than in the A Level cohort. A similar pattern can be seen for students with multiple specialisms but, for the remaining specialisms (Applied, Expressive, none), the pattern was the opposite.

Just under 45% of students in Russell Group institutions were specialists in STEM. This compares with 54% in Oxford/Cambridge and with only 14% or 19% in low or medium ranked HE institutions,

Table 1: Average number of A Levels and A Levels in facilitating subjects, by type of HE institution (N = number of students in the group*)

		A Levels (average)	*A Levels in facilitating subjects (average)*
A Level cohort (N=276,705)		2.70	1.34
University students (N=159,790)		2.83	1.48
Non-University students (N=116,910)		2.50	1.14
Oxford/Cambridge (N=3,920)		3.52	2.95
Russell Group (N=51,867)		3.15	2.16
Other: Non-Russell Group (N=107,925)		2.68	1.16
Overall HE ranking	Low (N=29,670)	2.53	0.90
	Medium (N=49,830)	2.64	1.06
	High (N=77,565)	3.08	1.99
Student satisfaction ranking	Low (N=41,095)	2.77	1.42
	Medium (N=60,340)	2.84	1.48
	High (N=55,630)	2.88	1.55
Research quality ranking	Low (N=29,680)	2.54	0.87
	Medium (N=57,670)	2.68	1.15
	High (N=69,320)	3.10	2.04
Graduation prospects ranking	Low (N=34,720)	2.55	0.88
	Medium (N=49,985)	2.71	1.19
	High (N=72,360)	3.06	1.99

* Following HESA's statistical disclosure control policy, counts were rounded to the nearest multiple of 5.

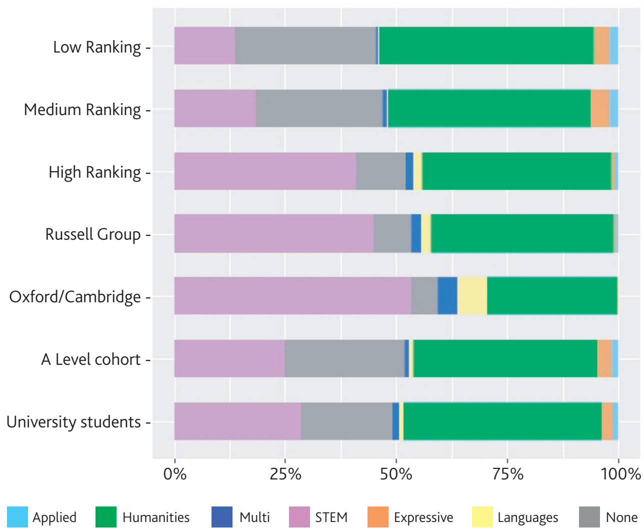


Figure 1: A Level specialists, by type of HE institution

respectively. The percentage of students with multiple specialisms was higher at Oxford/Cambridge than at other institutions, including institutions in the Russell Group, and just over 30% of the students in low-ranking universities did not have an A Level specialism.

The percentage of specialists in Humanities decreased with the increasing ranking of the HE institutions. Conversely, the percentages of specialists in STEM and Language subjects increased with the increasing ranking of the HE institutions (e.g., for STEM, the percentage increased from 14% in low-ranking institutions to 41% in high-ranking ones).

It is worth noting that the above patterns of A Level uptake might be influenced by the type of degrees (and entry requirements) offered at the different types of HE institutions. Students with A Levels in less academic or Applied subjects could be, for example, more attracted to the latter types of degrees and therefore their university choices could be determined by their degree choices.

Factors affecting enrolment in HE

Multilevel logistic regression analyses were carried out in order to look at the relationship between enrolment in HE and A Level specialism, controlling for background variables including performance at A Level and students' characteristics.

Two different sets of regression models were considered: whilst the first set of models looked at enrolment in HE amongst the national A Level cohort (the outcome variable being an indicator of enrolment at any HE institution), the second set of models focused on students who had already enrolled in HE and investigated the likelihood of enrolling at a specific type of HE institution (the outcome variables being: enrolment at an institution of the Russell Group; enrolment at Oxford/Cambridge).

For each set of models, we pursued the following approach. As a first step, a model including only the main effects of the specialism at A Level was considered. The outcomes of this model (Model A) show the effects of each of the different A Level specialisms (STEM, Humanities, Languages, etc.) on the probability of enrolling in HE, controlling for student and school characteristics. To investigate whether some of the background characteristics (in particular, gender and school type) interact with A Level subject specialism to influence the type of HE attended, a model including interaction terms between specialism, and gender and between specialism and school type was also considered (Model B).

Table 2: Enrolment in HE ~ regression analyses

Variable			*Model A*	*Model B*
			Estimate (Standard Error)	*Estimate (Standard Error)*
Intercept			-1.416 (0.054)†	-1.423 (0.054) †
Gender	Male		-0.087 (0.009) †	-0.113 (0.016) †
	[Female]			
Type of school	Independent		-0.736 (0.031) †	-0.644 (0.041) †
	[State]			
Prior attainment			0.020 (0.001) †	0.020 (0.001) †
Number of A Levels	2		0.198 (0.018) †	0.193 (0.018) †
	3		0.255 (0.021) †	0.248 (0.021) †
	4		0.474 (0.030) †	0.468 (0.031) †
	5+		0.364 (0.070) †	0.363 (0.070) †
	[1]			
Number of A Levels in facilitating subjects	1		0.065 (0.012) †	0.064 (0.012) †
	2		0.090 (0.016) †	0.089 (0.016) †
	3		0.140 (0.021) †	0.147 (0.021) †
	4+		0.070 (0.043)	0.089 (0.043)
	[0]			
A Level performance			0.024 (0.000) †	0.024 (0.000) †
A Level specialism	Applied		-0.060 (0.036)	0.005 (0.061)
	Expressive		-0.298 (0.026) †	-0.424 (0.032) †
	Humanities		0.165 (0.013) †	0.190 (0.017) †
	Languages		0.121 (0.048) †	0.113 (0.069)
	Multi		0.066 (0.045)	0.150 (0.066) †
	STEM		0.146 (0.018) †	0.070 (0.023) †
	[None]			
Type of school * A Level specialism	Independent	Applied		-0.230 (0.110) †
		Expressive		0.379 (0.085) †
		Humanities		-0.065 (0.037)
		Languages		-0.135 (0.097)
		Multi		-0.623 (0.117) †
		STEM		-0.207 (0.039) †
		[None]		
Gender * A Level specialism	Male	Applied		-0.046 (0.072)
		Expressive		0.344 (0.055) †
		Humanities		-0.057 (0.021) †
		Languages		0.069 (0.097)
		Multi		0.006 (0.084)
		STEM		0.171 (0.024) †
		[None]		

† Statistically significant at the 0.05 level.

Enrolment at any HE institution

Table 2 shows the effects of the A Level specialism on the probability of enrolling at any HE institution, after taking into account students' background characteristics such as their gender, prior attainment, prior institution, their A Level uptake (number of subjects) and their A Level performance.

Model A in Table 2 shows that the A Level specialism was a significant predictor of attending HE, even after controlling for students' characteristics and taking into account school effects. In particular, students who specialised in Expressive subjects were significantly less likely to enrol in HE than students with no specialism. On the contrary, students specialising in Humanities, Languages, STEM and those with a multiple specialism were significantly more likely to enrol at a HE institution than students with no specialism. Figure 2 shows that specialists in Humanities at A Level had the highest probabilities of attending HE, followed by STEM specialists. Students specialising in Expressive A Levels were the least likely to enrol in HE.

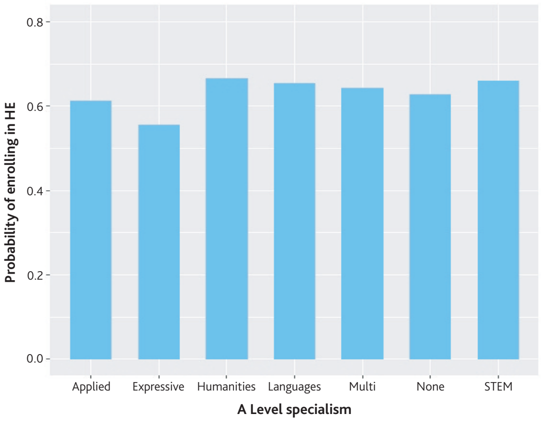


Figure 2: Probability of enrolling in HE, by A Level specialism (the calculated probabilities are for female students, attending a state school, taking three A Levels, one in a facilitating subject, and having average prior attainment and average A Level performance)

A brief summary of the effects of the background variables included in Model A on enrolment in HE, is provided below.

Gender: Male students were significantly less likely to enrol in HE than female students with the same attainment at GCSE and uptake/performance at A Level.

Type of school: Students with similar prior attainment and A Level uptake/performance from independent schools were less likely to enrol in HE in the year following completion of Key Stage 5 than students from state schools. Note that previous research (e.g., Crawford & Cribb, 2012) showed that gap-year takers were more likely to come from families of higher socio-economic status, including having university-educated parents and higher household incomes. Also, they were more likely to come from schools with relatively few students on free school meals and higher average academic performance, or from independent schools. Crawford and Cribb (2012) showed, for example, that nearly 20 per cent of gap year takers come from independent schools.

Prior attainment: Prior attainment at school was a significant predictor of enrolment at a HE institution. In particular, the probability of enrolment in HE increased with increasing prior attainment.

Number of A Levels: The probability of attending a HE institution increased significantly with the number of A Levels achieved. Students having two A Levels were slightly more likely to be in HE than those with just one, after controlling for all other variables. This likelihood increased further for students having three and four or more A Levels.

Number of A Levels in facilitating subjects: As above, the likelihood of enrolling in HE increased significantly for students having one, two or three A Levels in these subjects (compared to students with none). The effect of having four or more was not significantly different to the effect of having just three.

A Level performance: Overall achievement at A Level was a significant predictor of enrolment in a HE institution. In particular, the higher the average A Level score, the higher the probability of enrolment, suggesting that A Levels are good preparation for university.

Note that the *level of deprivation* was missing for around 40,000 students (approximately 15 per cent of the A Level cohort). Furthermore, there was high collinearity between missing level of deprivation and type of school (75 per cent of the students with missing data were in independent schools). An alternative model with the level of deprivation included was fitted. However, the effect of the different school types and, more importantly in this research, the effect of the A Level specialism were very similar to those in Model A. As a result, the level of deprivation was not considered in the rest of this article.

Model B, also shown in Table 2, investigated whether gender and type of school interact with A Level specialism to influence enrolment in HE.

Regarding gender, Model B shows that the interaction between gender and A Level specialism was significantly associated with enrolment in HE. Table 3 shows how the probabilities of enrolling in HE by students with each of the A Level specialisms varied by gender. For example, male students specialising in STEM and Expressive subjects were more likely than female students specialising in the same areas to enrol in HE. On the contrary, female students were more likely to enrol in HE than male students if they were specialists in Applied, Humanities

Table 3: Enrolment in HE ~ probability for students with each A Level specialism, by gender*and by type of school**

Background characteristics		*A Level specialism*						
		Applied	*Expressive*	*Humanities*	*Languages*	*Multi*	*STEM*	*None*
Gender	Female	0.63	0.52	0.67	0.65	0.66	0.64	0.63
	Male	0.59	0.58	0.63	0.64	0.64	0.66	0.60
Type of school	Independent	0.41	0.46	0.50	0.46	0.35	0.43	0.47
	State	0.63	0.52	0.67	0.65	0.66	0.64	0.63

* These probabilities are for students in state schools, who achieved three A Levels (one in a facilitating subject) and with average attainment at Key Stage 4 and at A Level. Note that, although the probabilities are slightly different, the patterns (in terms of differences between male and female students) were the same for students in independent schools.

** These probabilities are for female students, who achieved three A Levels (one in a facilitating subject) and with average attainment at Key Stage 4 and at A Level. Note that, although the probabilities are slightly different, the patterns (in terms of the differences between students in independent and state schools) were the same for male students.

or Language A Level subjects. Female students were also more likely to enrol in HE if they had multiple specialisms or did not specialise at all.

Regarding type of school, Model B also shows that the interaction between type of school and A Level specialism was significantly associated with enrolment in HE. In particular, Table 3 shows how the probabilities of enrolling in HE by students with each of the A Level specialisms varied by type of school. Although students in independent schools had a lower probability of enrolling in HE overall, the differences between these probabilities varied by A Level specialism: the smallest difference was between students specialising in Expressive A Level subjects (followed by those with no specialism) and the highest difference was between students with multiple specialisms or a specialist in STEM.

Enrolment at different types of HE institutions

Regression models similar to the ones reported in Table 2 were fitted for enrolment at an institution of the Russell Group and enrolment at Oxford/Cambridge. The results are briefly described below and full details of the regression models are available in Table 4. Note that the analyses reported in this section were restricted to students with three or more A Levels.

A Level specialism was a significant predictor of attending a university of the Russell Group, even after controlling for students' characteristics and school effects such as the type of secondary school. Similarly to the results for enrolling in HE institutions in general (described in the previous section), students specialised in Expressive subjects were significantly less likely to enrol at a Russell Group institution than students with no specialism. On the contrary, students specialising in Applied subjects, Humanities, Languages, STEM and those with a multiple specialism were more likely to enrol at a Russell Group institution than students with no specialism. In particular, after accounting for other student and school characteristics, specialists in Languages at A Level had the highest probabilities of attending institutions in the Russell Group, followed by Humanities specialists. Students specialising in Expressive A Levels were the least likely to enrol at HE in Russell Group institutions (see Figure 3[a]).

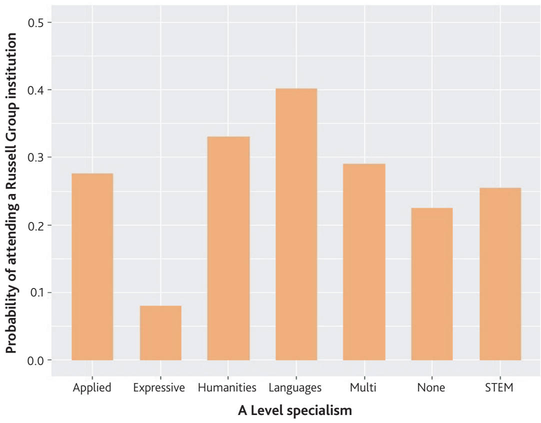
The A Level specialism effect was slightly different for students enrolling at Oxford/Cambridge. In particular, students with no specialism were more likely to enrol in Oxford/Cambridge than students specialising in STEM or Expressive subjects, and more likely than students with multiple specialisms. Figure 3(b) shows that specialists in Languages at A Level had the highest probabilities of attending

Table 4: Enrolment at different types of HE institutions ~ regression analyses*

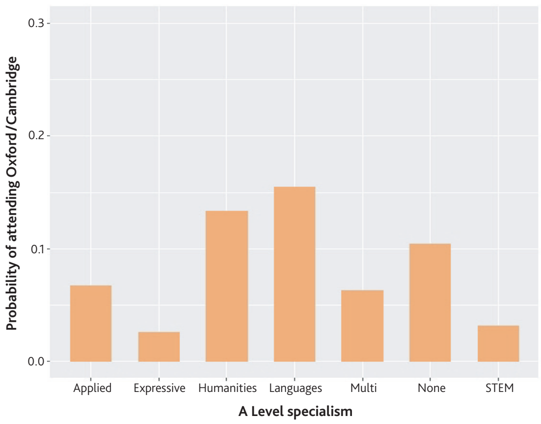
Variable			*Russell Group*		*Oxford/Cambridge*	
			Model A	*Model B*	*Model A*	*Model B*
			Estimate (Standard Error)	*Estimate (Standard Error)*	*Estimate (Standard Error)*	*Estimate (Standard Error)*
Intercept			-11.256 (0.119) †	-11.196 (0.123) †	-35.021 (0.638) †	-34.589 (0.639) †
Gender	Male [Female]		0.120 (0.018) †	-0.021 (0.057)	0.225 (0.047) †	-0.417 (0.228)
Type of school	Independent [State]		0.568 (0.036) †	0.561 (0.076) †	-0.096 (0.057)	-0.370 (0.208)
Number of A Levels	4		0.017 (0.031)	0.023 (0.031)	0.198 (0.062) †	0.177 (0.062) †
	5+		0.212 (0.103) †	0.221 (0.103) †	0.219 (0.114)	0.222 (0.114)
	[3]					
A Level specialism	Applied		0.276 (0.079) †	0.324 (0.124) †	-0.483 (0.941)	-1.086 (1.578)
	Expressive		-1.203 (0.089) †	-1.217 (0.110) †	-1.488 (0.679) †	-1.673 (0.895)
	Humanities		0.532 (0.030) †	0.445 (0.041) †	0.281 (0.106) †	0.023 (0.140)
	Languages		0.840 (0.081) †	0.981 (0.107) †	0.455 (0.143) †	0.103 (0.205)
	Multi		0.347 (0.072) †	0.342 (0.095) †	-0.548 (0.153) †	-0.977 (0.215) †
	STEM		0.165 (0.034) †	0.130 (0.045) †	-1.277 (0.112) †	-1.617 (0.145) †
	[None]					
Type of school * A Level specialism	Independent	Applied		-0.724 (0.276) †		-0.199 (4.877)
		Expressive		-0.294 (0.225)		-0.065 (1.443)
		Humanities		0.211 (0.079) †		0.114 (0.221)
		Languages		-0.301 (0.178)		0.256 (0.283)
		Multi		0.284 (0.260)		0.442 (0.312)
		STEM		-0.242 (0.081) †		0.373 (0.215)
		[None]				
Gender * A Level specialism	Male	Applied		0.081 (0.159)		0.919 (2.046)
		Expressive		0.191 (0.198)		0.730 (1.423)
		Humanities		0.147 (0.062) †		0.622 (0.240) †
		Languages		-0.234 (0.170)		1.025 (0.306) †
		Multi		-0.037 (0.136)		0.648 (0.311) †
		STEM		0.154 (0.062) †		0.673 (0.235) †
		[None]				

* The effects of *prior attainment, A Level performance* and the *number of A Levels in facilitating subjects* are not shown in this table as they are very similar to those reported in Table 2. Full results from the regression analyses can be found in Vidal Rodeiro (2019).

† Statistically significant at the 0.05 level.



(a) Russell Group



(b) Oxford/Cambridge

Figure 3: Probability of enrolling at different types of institution, by A Level specialism. The calculated probabilities in Figure 3(a) are for female students, attending a state school, taking three A Levels (two in facilitating subjects), and having average prior attainment and average A Level performance. The calculated probabilities in Figure 3(b) are for female students, attending a state school, taking three A Levels (two in facilitating subjects), and having prior attainment at the 90% percentile and A Level performance at the 90% percentile.

Oxford/Cambridge, followed by Humanities specialists and students with no specialism. Students specialising in Expressive A Levels were the least likely to enrol at Oxford/Cambridge, followed by those with a specialism in STEM or Applied subjects.

As in the previous section, a brief summary of the effects of the background variables included in Model A on enrolment at different types of HE institutions is provided below.

Gender: Contrary to the effect on enrolment in HE, male students were significantly more likely than female students with the same prior attainment and same background characteristics to enrol at institutions in the Russell Group or at Oxford/Cambridge.

Type of school: Students with similar prior attainment and A Level uptake/performance from independent schools were more likely to enrol at institutions in the Russell Group. Note that the effect of school type on enrolling at Russell Group institutions (generally prestigious and highly ranked) was the opposite to the effect of school type on attending HE in general. There was, however, not an effect of school type on the probability of attending Oxford/Cambridge.

Number of A Levels: The probability of attending a more prestigious group of universities, such as the Russell Group or Oxford/Cambridge, increased significantly with the number of A Levels achieved. In particular, students having four A Levels were slightly more likely to be in this type of HE institution than those with just three (the baseline in these analyses), after controlling for all other student characteristics. This likelihood increased for students having five or more A Levels. However, having four A Levels did not have a significant effect on the probability of attending a Russell Group institution, relative to having three.

The effect of *prior attainment*, *A Level performance* and the *number of A Levels in facilitating subjects* on the probability of enrolment at Russell Group institutions or at Oxford/Cambridge was the same as the effect on enrolment in HE.

In this section, we also investigated whether gender and type of school interact with A Level specialism to influence the type of HE institution attended. A summary of the results (in the form of probabilities of enrolment) is given in Table 5.

The probabilities of enrolling in different types of HE institutions by students with each of the A Level specialisms varied by gender. For example, the probability of attending an institution in the Russell Group and the probability of attending Oxford/Cambridge for a specialist in STEM was very similar for males and females. However, female students were more likely to enrol in Russell Group institutions than male students if they specialised in Language subjects at A Level, and the opposite was found for enrolment at Oxford/Cambridge.

Regarding type of school, its interaction with A Level specialism was also significantly associated with the type of HE institution attended. In particular, Table 5 shows that, all else being equal, STEM specialists were more likely to attend Russell Group institutions if they took their A Levels at an independent school than if they did so at a state school. However, for these students, the probability of attending Oxford/Cambridge did not vary by the type of school attended. The smallest difference in the probabilities of enrolment at Russell Group institutions between students from independent and state schools was between specialists in Expressive A Level subjects (followed by those with an specialism in Applied subjects) and the highest difference was between students with no specialism or specialists in Humanities. Differences in the probabilities of enrolment in Oxford/Cambridge between both groups of students were fairly small for all specialisms.

Conclusions and discussion

The process of application and admission to universities in the UK places a relatively strong weight on the type of A Level subjects achieved by students. As a result, A Level choice is a key factor influencing progression from secondary education to HE. This research aimed to provide quantitative evidence to show how different A Level subjects (and combinations of A Level subjects) are used by students to access HE and, in particular, different types of HE institutions. The key results are discussed below.

Uptake of A Level subjects

A Level subjects were represented in different proportions in HE and particularly in the different institution types, suggesting that subject choice is associated with the type of HE institution attended. The relationships observed in this research could result from a mixture of different factors. For example: subject requirements for certain degree courses; usefulness of certain subjects for certain areas of degree study;

Table 5: Enrolment at different types of HE institutions ~ probability for students with each A Level specialism, by gender* and by type of school**

Type of HE institution	*Background characteristics*		*A Level specialism*						
			Applied	*Expressive*	*Humanities*	*Languages*	*Multi*	*STEM*	*None*
Russell Group	Gender	Female	0.30	0.08	0.32	0.45	0.30	0.26	0.23
		Male	0.31	0.10	0.35	0.39	0.29	0.28	0.23
	Type of school	Independent	0.26	0.11	0.51	0.51	0.50	0.32	0.35
		State	0.30	0.08	0.32	0.45	0.30	0.26	0.23
Oxford/Cambridge	Gender	Female	0.05	0.03	0.14	0.15	0.06	0.03	0.14
		Male	0.08	0.04	0.17	0.25	0.07	0.04	0.10
	Type of school	Independent	0.03	0.02	0.11	0.14	0.06	0.03	0.10
		State	0.05	0.03	0.14	0.15	0.06	0.03	0.14

* These probabilities are for students in state schools, who achieved three A Levels (two in facilitating subjects) and with average attainment at Key Stage 4 and at A Level. Note that, although the probabilities are slightly different, the patterns (in terms of differences between male and female students) were the same for students in independent schools.

** These probabilities are for female students, who achieved three A Levels (two in facilitating subjects) and with average attainment at Key Stage 4 and at A Level. Note that, although the probabilities are slightly different, the patterns (in terms of the differences between students in independent and state schools) were the same for male students.

different prevalence of degree subject areas at different institutions; or different levels of selectivity at different HE institution types.

The number of A Level subjects held by students varied across the different types of HE institutions considered in this research. For example, students at Oxford/Cambridge held the highest number of A Level subjects and students attending low-ranking institutions the lowest. Similar patterns, also reported by Dilnot (2018), were found for A Levels in facilitating subjects.

The current study supports previous research (e.g., Vidal Rodeiro et al., 2015) showing that students with more academic backgrounds were more likely to go to universities in the Russell Group and those with more applied or vocational backgrounds were more likely to study in other types of universities. In fact, just under half of the students in Russell Group institutions were specialists in STEM and the percentages of specialists in STEM and Language subjects increased with the increasing ranking of the HE institutions. For Oxford/Cambridge this percentage was around 55 per cent. It should be noted that one reason for this could be that STEM degrees courses are more common in high-ranking and prestigious HE institutions and the more applied/vocational degrees are overrepresented in other types of HE institutions. Students with A Levels in less academic or Applied subjects could be, for example, more attracted to the latter types of degrees and therefore their university choices are determined by their degree choices.

Factors affecting enrolment in HE

The regression analyses carried out in this research revealed that there was a relationship between A Level subject specialism and the type of university attended, and that this association holds even after controlling for other variables, such as attainment and type of school attended. In particular, students specialising in Expressive subjects were significantly less likely to enrol in HE, and to attend an institution in the Russell Group, than students with no specialism were. Conversely, students with Humanities, Languages or STEM specialisms and those with a multiple specialism were significantly more likely to enrol at a HE institution than students with no specialism. The A Level specialism effect was slightly different for students enrolling at Oxford/Cambridge. For example, students with no specialism were more likely to enrol in in these universities than students specialising in STEM or Expressive subjects. This supports the view that careful choice of subjects post-16 is crucial to avoid students inadvertently closing their options down prematurely.

Across all the models fitted in this work, a common result, consistent with previous research (Vidal Rodeiro & Sutch, 2013; Dilnot, 2018) emerged: the probability of attending any HE institution increased significantly with the number of A Levels achieved and with the number of A Levels in facilitating subjects. This suggests that studying A Levels in facilitating subjects may be a sensible choice for students wanting to attend prestigious and high-ranking HE institutions.

The regression analyses also showed that male students were significantly less likely than female students with the same prior attainment and same background characteristics to enrol in HE. However, if they enrolled at all, male students were significantly more likely than female students to attend institutions in the Russell Group and, in particular, Oxford/Cambridge.

As expected, and in line with previous research (e.g., HEFCE, 2003; Smith & Naylor, 2005; Crawford, 2014; Vidal Rodeiro & Zanini, 2015), A Level performance was strongly associated with participation in HE and with attendance at specific types of HE institutions. Specifically, the higher the average A Level score, the higher the probability of enrolment in HE overall and, for those who enrol, the probability of attending more prestigious institutions. Similarly, performance at Key Stage 4 was found to be an important factor for university entry, even after taking into account the performance at A Level.

Students in independent schools with the same A Level specialism and the same A Level performance were less likely to enrol in HE immediately after completing their A Levels than students in state-maintained schools. However, the probability of attending prestigious and high-ranking institutions, such those in the Russell Group, was higher for them compared to similar students in state-maintained schools. This is important from a widening participation point of view, as it supports other research findings (e.g., Sutton Trust, 2011; Chowdry et al., 2013; Sullivan, Parsons, Wiggins, Heath, & Green, 2014; Montacute & Cullinane, 2018) in providing evidence that young people from state, rather than independent, schools continue to be underrepresented at high-status universities.

However, in contrast, there was not an effect of school type (independent vs. state) on the probability of attending Oxford/Cambridge.

Further regression analyses showed that, when prior schooling and other background characteristics were accounted for, the likelihood of enrolling in HE by students with each of the A Level specialisms varied, indeed, by gender and type of school. For example, male students specialising in STEM and Expressive subjects were more likely than female students specialising in the same areas to enrol in HE. On the contrary, female students with any other specialism, or no specialism at all, were more likely to enrol in HE than male students. Although these patterns were fairly similar for the likelihood of enrolling in different types of HE institutions, there were some differences. In particular, female students were less likely to enrol in Oxford/Cambridge than male students if they specialised in Language subjects at A Level or if they had multiple specialisms. Regarding type of school, its interaction with A Level specialism was also significantly associated with HE enrolment. For example, STEM specialists were more likely to attend Russell Group institutions if they took their A Levels in an independent school than if they did so in a state school. However, for this group of students, the probability of attending Oxford/Cambridge did not vary by the type of school.

As discussed above, this research showed that a clear relationship between A Level specialism and the type of HE institution attended exists and that this relationship varied by gender and school type. However, the multilevel logistic regression, as any regression technique, can only ascertain relationships, but never be sure about the underlying causal mechanism. Therefore, caution must be taken when interpreting the results of the regression analyses presented in this work. Furthermore, the existing literature on transitions from school to different education and employment destinations (e.g., Chowdry et al., 2013; Boliver, 2013; Smith et al., 2015; Hupkau et al., 2017; Montacute & Cullinane, 2018) shows that there is a very complex set of factors that can influence progression and, therefore, HE participation. Although the most relevant factors identified in the literature were accounted for in this research, others (e.g., aspirations, career goals) can be difficult or impossible to measure and cannot be included in quantitative research studies such as this one.

The above results confirm that, although careful choice of A Level subjects/specialisms is crucial for enrolling in HE and, in particular, for enrolling in specific HE institutions, background characteristics such as gender and school type are still part of the explanation for differential participation in HE in the UK. While the access gap between students from different backgrounds has narrowed somewhat in recent years due to widening participation activities, the gap in the most selective institutions remains (Boliver, Crawford, Powell, & Craige, 2017). Contextualising admissions (i.e., taking into account a student's background when making decisions) might be one way to make progress towards narrowing the gap. Some HE institutions had already changed their admissions requirements for state school students and for students from disadvantaged backgrounds (Ogg, Zimdars, & Heath, 2009; Boliver et al., 2017) and continue with this practice. However, there is still scope to improve the use of contextual data in the admission processes to widen access and to reduce the differences in participation between students with different backgrounds, particularly at prestigious and highly selective institutions.

References

Boliver, V. (2013). How fair is access to more prestigious UK universities? *The British Journal of Sociology, 64*, 344–364.

Boliver, V., Crawford, C., Powell, M., & Craige, W. (2017). *Admissions in context: The use of contextual information by leading universities*. London: The Sutton Trust.

Bramley, T. (2014). Multivariate representations of subject difficulty. *Research Matters: A Cambridge Assessment Publication, 18*, 42–48.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 176*, 431–457.

Crawford, C. (2014). *The link between secondary school characteristics and university participation and outcomes*. London: Department for Education.

Crawford, C., & Cribb, J. (2012). *Gap year takers: update, trends and long-term outcomes*. Research Report DFE-RR252. London: Department for Education.

Dilnot, C. (2016). How does the choice of A-level subjects vary with students' socio-economic status in English state schools? *British Educational Research Journal, 42*, 1081–1106.

Dilnot, C. (2018). The relationship between A-level subject choice and league table score of university attended: the 'facilitating', the 'less suitable' and the counter-intuitive. *Oxford Review of Education, 44*(1), 118–137.

Gill, T. (2017). *Provision of GCE A level subjects 2016*. Statistics Report Series No. 117. Cambridge: Cambridge Assessment.

Hayward, G., & Hoelscher, M. (2011). The use of large-scale administrative data sets to monitor progression from vocational education and training into higher education in the UK: possibilities and methodological challenges. *Research in Comparative and International Education, 6*(3), 316–329.

HEFCE. (2003). *Schooling effects on higher education achievement*. Bristol: Higher Education Funding Council for England.

Hupkau, C., McNally, S., Ruiz-Valenzuela, J., & Ventura, G. (2017). Post-compulsory education in England: Choices and implications. *National Institute Economic Review, 240*(1), R42–R57.

McCoy, T., & Adamson, D. (2016). Building a house on sand? Are vocational education and training qualifications an (in)equitable stepping-stone to success in higher education? In G. Steventon, D. Cureton and L. Clouder (Eds.), *Student Attainment in Higher Education: Issues, controversies and debates* (pp.161–173). Abingdon: Routledge.

Montacute, R., & Cullinane, C. (2018). *Access to advantage: the influence of schools and place on admissions to top universities*. London: Sutton Trust.

Ogg, T., Zimdars, A., & Heath, A. (2009). Schooling effects on degree performance: a comparison of the predictive validity of aptitude testing and secondary school grades at Oxford University. *British Educational Research Journal, 35*(5). 781–807.

Russell Group. (2017). *Informed Choices: a Russell Group guide to making decisions about post-16 education*. London: Russell Group.

Smith, S., Joslin, H., & Jameson, J. (2015). *Progression of College Students in England to Higher Education* (BIS Research Paper Number 239). London: Department for Business, Innovation and Skills.

Smith, J., & Naylor, R. A. (2005). Schooling effects on subsequent university performance: evidence for the UK university population. *Economics of Education Review, 24*, 549–562.

Sullivan, A., Parsons, S., Wiggins, R., Heath, A., & Green, F. (2014). Social origins, school type and higher education destinations. *Oxford Review of Education, 40*, 739–763.

Sutch, T., Zanini, N., & Vidal Rodeiro, C.L. (2016). The effect of specialism and attainment in secondary school on the choice of Higher Education institution and field of study. *Research Matters: A Cambridge Assessment Publication, 21*, 2–11.

Sutton Trust. (2011). *Degrees of success: University chances by individual school*. London: Sutton Trust.

UCAS. (2016). *Applicants and acceptances by groups of applicants 2015*. Cheltenham: Universities and Colleges Admissions Service.

UCAS. (2017a). *End of Cycle Report 2017: qualifications and competition*. Cheltenham: Universities and Colleges Admissions Service.

UCAS. (2017b). *End of Cycle 2017 Data Resources: acceptances by deferred and age*. Cheltenham: Universities and Colleges Admissions Service.

Vidal Rodeiro, C.L. (2007). *A level subject choice in England: patterns of uptake and factors affecting subject preferences*. Research Report. Cambridge: Cambridge Assessment.

Vidal Rodeiro, C.L. (2019). *Popularity of A level subjects among university students*. Cambridge Assessment Research Report. Cambridge: Cambridge Assessment.

Vidal Rodeiro, C.L., & Sutch, T. (2013). *Popularity of A level subjects among UK university students*. Statistical Report Series no.52. Cambridge: Cambridge Assessment.

Vidal Rodeiro, C.L., Sutch, T., & Zanini, N. (2015). Progressing to Higher Education in the UK: the effect of prior learning on institution and field of study. *Research Matters: A Cambridge Assessment Publication, 20*, 13–21.

Vidal Rodeiro, C.L., & Williamson, J. (2018). Meaningful destinations: using national data to investigate how different education pathways support young people's progression in England. *Research Papers in Education*.

Vidal Rodeiro, C.L., & Zanini, N. (2015). *The role of the A* grade at A level as a predictor of university performance in the United Kingdom*. Oxford Review of Education, 41(5), 647–670.

Studying English and Mathematics at Level 2 post-16: issues and challenges

Jo Ireland Research Division

Introduction

Alison Wolf stated in her *Review of Vocational Education* (2011) that despite General Certificate of Secondary Education (GCSE) Mathematics and English being key to employment and education prospects, less than 50 per cent of students achieved grades A*-C in both qualifications. The review recommended that English and Mathematics should form a required component of post-16 study programmes for those without GCSE grades A*-C in these subjects, working either directly towards GCSE or other qualifications which provide "significant progress" towards GCSE success.

This recommendation proved to be the catalyst for a number of revisions to Government policy. This article looks at these post-16 policy changes and the subsequent challenges and issues faced by students, teachers and providers of English and Mathematics at Level 2. It covers GCSE resits and Functional Skills. In addition, it aims to shed light on what support is needed by students and teachers, and whether those support needs differ according to the qualification in question.

The majority of published commentary in this area relates to learners in Further Education (FE) colleges who are resitting GCSE English and/or Mathematics. There is little work on those in other settings (e.g., sixth form colleges), or those taking other qualifications in English and Mathematics post-16. There is also more research relating to Mathematics than to English, although many of the findings may apply to both subjects.

Policy changes and effects

In 2014, the Government introduced the requirement that students in England aged 16–18 who have not achieved at least a grade C/4[1] in GCSE English or Mathematics should continue to study these subjects as part of their programme of study (Education and Skills Funding Agency, 2019). Furthermore, from 2015, it was decided that those students who have achieved a "near pass" (grade D/3) in English[2] and/or Mathematics must study a GCSE course. Those who achieved grade E/2 or below must study for a GCSE or an approved stepping-stone qualification (Education and Skills Funding Agency, 2019).

For the academic year 2019/20, students with a GCSE grade 2 or below in English and/or Mathematics can study for a Level 2 Functional Skills qualification or a GCSE grade 9 to 4. If these students achieve a pass grade for Level 2 Functional Skills, then they are no longer required to work towards the GCSE. Those with a GCSE grade 3 are still required to study for GCSE grade 9 to 4 (Education and Skills Funding Agency, 2019).

These Government requirements have led to significant increases in entries for students aged 17 and older taking GCSE English Language and GCSE Mathematics since 2014 (FFT Education Datalab, 2018). In 2014/15, around half of all post-16 students who did not achieve a pass grade in English or Mathematics were enrolled in FE colleges (Department for Education, 2016). Interestingly, sixth form colleges and sixth forms in schools achieve higher pass rates compared to FE colleges, possibly due to their higher entrance criteria and differences in teacher characteristics (Higton et al., 2017).

Although much of the research in this area concentrates on FE colleges, there are other settings to consider. (Creese, Litster, & Mallows, n.d.) addressed the impact of the Government requirements on

1. In 2017, the 9-1 grading scale was introduced for GCSEs in England, replacing the A*-G scale. Grade 4 is equivalent to grade C (Ofqual, 2017).
2. Students achieving C/4 or higher in English Literature are not required to study English post-16 (Education and Skills Funding Agency, 2019).

different settings including FE colleges, private training providers and work-based learning, adult and community learning, and prisons. They found that staff in organisations delivering apprenticeships or work-based learning qualifications may be confident when teaching Functional Skills, but have little to no experience of GCSEs.

Vidal Rodeiro (2018) reported that the resit policy has been criticised by stakeholders and other commentators. The figures for 2018 showed that most students[3] who retook GCSE English and/or Mathematics did not improve their grade (Sezen, 2018). Vidal Rodeiro (2018) showed that many students failed to achieve the required grade by the time they left compulsory education, even if they retook multiple times. In fact, the probability of improving their grade decreased with the number of resit attempts. There are concerns that repeated GCSE resits are demotivating and can impact students' mental health (Belgutay, 2018).

Students

This section looks at the challenges faced by students and considers student backgrounds, motivation and support needs.

Student backgrounds

There is evidence that students from disadvantaged backgrounds are disproportionately represented in the cohort who have not achieved the required grades for GCSE (Jerrim, Greany, & Perera, 2018; Maughan, Smith, Mitchell, Horrocks, & Taylor, 2016; Skills Commission, 2016). Furthermore, disadvantaged students are less likely to achieve C/4 grade GCSEs in English and Mathematics post-16 than non-disadvantaged students (Belgutay, 2017; Impetus, 2016).

Those students who have special educational needs or disabilities (SEND) or speak English as an additional language (EAL) may also be overrepresented. Students will have differing cultural, socio-economic and geographic backgrounds. There may also be other considerations faced by this group, such as personal maturity, aspirations (Skills Commission, 2016), vocational confusion, described as "not having identified their skills and not knowing what is involved in specific careers," (Williams, Hadjivassiliou, Marvell, Green, & Newton, 2017) and complications such as caring responsibilities, homelessness, or any number of adverse life events (Higton et al., 2017).

Student motivation, attitudes and emotions

There are concerns that the mandatory requirement to study English and Mathematics creates resentment and demotivates students (Education & Training Foundation, 2014a). Students affected by this policy:

- tend to be disaffected by their prior learning experiences;
- are more likely to see the compulsory study as a result of their "failure" at GCSE;
- hold negative beliefs about their ability (e.g., that they cannot achieve in the subject); and
- demonstrate an unwillingness to engage.

(Creese, Litster, & Mallows, n.d.; ETF, 2014a, 2015; Higton et al., 2017).

As a consequence, one key finding from the literature is that providers recognise the need to restore students' self-confidence and their confidence in the education system in order to give them the best chance of achieving in English and/or Mathematics (Williams et al., 2017). Demotivation is likely to be most acutely felt by those students with D/3 at GCSE, who are required to resit GCSE without the opportunity to build confidence with stepping-stone qualifications (Williams et al., 2017).

Bellamy (2017) describes a particular tension when it comes to GCSE resits post-16. On the one hand, students have embarked on their chosen progression path and expect to be treated as young adults. On the other, English and/or Mathematics is a compulsory requirement, which is undeniably linked with the school context and the negative associations that those students who have left school may have. Robey and Jones (2015) found that post-16 resit students were likely to say they had not enjoyed English and/or Mathematics at school, so disengagement and demotivation are likely to stem from the requirement to continue with them.

When FE students talked about Mathematics, Bellamy (2017) found that they often referred to the set that they were placed in at school, which suggests that long term personal beliefs about ability can be formed while at school. In a review carried out for the Department for Education, Professor Adrian Smith related this to a wider societal negative view of Mathematics in particular (Department for Education, 2017). Larger class sizes at school and the perceived "unavailability" of the teacher were cited by students as barriers to achievement at GCSE, while learners also described embarrassment at having to ask for support in front of peers (Robey & Jones, 2015).

When studying English and/or Mathematics post-16, negative emotions can be further compounded by repetition of subject content, which can demotivate students who found the concepts difficult first time round, whilst also serving to emphasise where they have previously failed (ETF, 2015). Although it appears to be less commonly used post-16, the traditional transmission teaching style associated with the school context can cause students to feel like they are "back at school".

Higton et al. (2017) noted that although the majority struggled with negative emotions around English and Mathematics, some students took a more positive approach. This was mainly observed among those who recognised that gaining the GCSE is required for their future progression or desired career paths.

Student support needs

A number of supportive strategies which improved outcomes for students were identified in the literature (Anderson & Peart, 2016; Curee, n.d.; ETF, 2014a; Robey, Woodhouse, & Downes, 2016). These included:

- a range of ways to access support, for example: additional one-to-one tuition or drop-in sessions which allow students to ask for help on specific areas of difficulty;
- building relationships with teachers in which students feel respected and are treated more like adults to increase motivation;
- small teaching groups to encourage a supportive learning environment and increase access to support;
- regular assessments and feedback to enable learners to chart their progress;
- practical and interactive teaching methods;

3. Only 27.7 per cent of 18-year-olds retaking English language achieved 4/C and above, while 14.3 per cent of 18-year-olds retaking Mathematics achieved 4/C and above (Ofqual, 2018).

- real-life contexts and an understanding of why tasks or activities undertaken are relevant and necessary; and
- assessment strategies, for example: understanding mark schemes, advice on structuring answers and revision guidance.

Students who speak English as a second language may struggle with Level 2 English, as might be expected, but also Mathematics, since a certain level of English is required to understand the Mathematics content (Higton et al., 2017; Robey et al., 2016). Students with SEND or mental health issues may also have experienced difficulties with English and Mathematics at school and need extra support at post-16. This support can take the form of "official" allowances for those with a statement of additional needs, such as extra time in assessments and scribes. Less formal support can also be offered such as additional one-to-one support, sessions on tackling exam anxiety, or tailored learning pathways.

Teaching

Approaches to teaching need to take into account the support needs of post-16 learners. This section looks at motivating students, determining students' skills and support needs, and delivery strategies.

Motivating students and building confidence

ETF (The Education and Training Foundation) stated that "building learner confidence and challenging maladaptive beliefs" is vital for those studying at post-16. They listed mentoring, one-to-one support, clear progression routes and tailored provision as ways to build learner confidence (ETF, 2014a). Some providers begin the process of improving engagement before students are enrolled with them. When describing how they publicised their offering to prospective students, many providers presented their programme as an alternative to the academic progression route, emphasised how the teaching differed from school approaches, and characterised their offering as a second chance (Williams et al., 2017).

There are indications that levels of engagement improve if students see gaining the GCSE as a means to achieving their desired career or progressing to university (Higton et al., 2017; Robey & Jones, 2015). For those following vocational courses, presenting English or Mathematics in a vocational or real-life context heightens the perceived relevance and subsequent engagement (this is explored further below).

Some providers recognise the impact of students' negative beliefs and encourage positive mental attitudes and growth mindsets, reframing the situation as "haven't achieved the required grade yet," instead of focusing on "failure" (ETF, 2014a; Higton et al., 2017). There is a great deal of research into mindsets, and how to encourage students to change from a fixed to a growth mindset. ETF (2014a) cited Carole Dweck's work and summarised the ways that teachers can promote growth mindset:

- By teaching students about the new science of brain plasticity and the new view of talent and giftedness as dynamic attributes that can be developed.
- Through the portrayal of challenges, effort, and mistakes as highly valued.
- Through process praise and feedback.

(Dweck, 2008, pp.9–13)

Anderson and Peart (2016) interviewed students enrolled on a fast-track GCSE course at an FE college and found that teacher support and peer support both contributed to these students' motivation. They suggested that peer cooperation and group interaction can increase positive perceptions of learning and educational achievement. Learners also cited smaller class sizes as increasing the support available from teachers compared with previous class sizes at school.

Stepping-stone qualifications can be a way to build students' confidence in preparation for GCSE. Functional Skills is the most commonly used qualification for this purpose (although not designed as such).

Diagnostic assessment

Some providers find a GCSE grade alone to be an unreliable indicator of ability (ETF, 2014a) and may seek students' exam *marks* data in order to establish their position within a grade band. Some colleges therefore request students' prior results data from schools (Higton et al., 2017). However, given that colleges rely on schools' willingness and ability to provide the information, this situation can lead to difficulties with the timeliness and quality of the data available. As Robey and Jones (2015) pointed out, delays can have a knock-on effect on learners, such as being placed in the wrong group or not having any additional needs identified early enough.

Previous formal assessments may not provide the full picture of students' skill levels, and in some cases the length of time between taking GCSEs and the start of the new academic year can further reduce knowledge and skills (Higton et al., 2017). In addition, some students may not have previously taken GCSEs while some students may possess good subject knowledge and skills but struggle with exam strategy, test anxiety or a lack of preparation skills. Diagnostic assessments are therefore a vital part of the post-16 picture.

When a new intake of students who did not achieve the required grades in English and/or Mathematics begin post-16 study, providers' first step is usually to ascertain their current level of knowledge and skills in order to organise classes and make decisions about which qualification(s) to target (ETF, 2014a). Maughan et al. (2016) found evidence that initial assessment improved outcomes for both English and Mathematics.

Higton et al. (2017) stated that the aims of diagnostic testing are usually to identify students':

- current level;
- existing knowledge of topics;
- proficiency with routine tasks;
- gaps in knowledge; and
- additional support needs.

Providers should be mindful of how they conduct their diagnostic testing however, because students who are already at risk of being demotivated may find that an initial raft of tests confirms their belief that English or Mathematics at post-16 will simply be a repeat of the experience they encountered at school (ETF, 2014a).

The information gleaned from diagnostic assessment is then used to decide which qualification(s) students will work towards, organise levelled classes if applicable, and prioritise topics and concepts for groups. Porter (2015) reported that successful institutions divided students into two groups – borderline (D/3 and some E/2 grades) and those further behind (F/1 and below) – in order to provide targeted teaching. As Higton et al. (2017) noted however, the resources available to the provider determines to what extent they are able to tailor their provision in this way.

An important outcome of diagnostic testing is to establish the topics and/or concepts that are well understood, in order to avoid unnecessary coverage. ETF (2014a) noted two benefits to this: one is preventing disengagement of students through repetition, and the other is not using valuable time to cover content that has already been mastered. They also emphasised the importance of Mathematics teaching being based on the current level of students. Students arrive at post-16 with at least some prior knowledge, and although the failure to achieve a desired grade might lead providers to "go back to basics", trying to cover a whole course may be repeating content that students have already mastered.

Higton et al. (2017) noted that results from diagnostic tests can be used formatively and some teachers incorporate formative assessment in lessons through progression tests and collaborative exercises such as peer marking.

One common strategy adopted by providers is to identify those students who are likely to achieve a pass grade with an early resit. This has the double benefit of motivating students (they can stop studying the subject if they pass) and, assuming they achieve the required grade, reduces the class sizes, allowing support to be focused on those that need it most (Higton et al., 2017; Porter, 2015).

Most providers use diagnostic testing to group students by level (Higton et al., 2017). However, if course structure or resources do not allow for this, and even when grouping is used, there is likely to be a range of students' abilities across a teaching group and teachers will therefore need differentiated tasks. Two approaches observed by Higton et al. (2017) were schemes of work with flexibility for teachers to adapt as necessary, and the setting of individual targets for students via software and online resources.

Delivery strategies

In post-16 settings, successful teaching strategies are likely to use a range of learning activities, such as group discussion or paired working – rather than approaches generally used in school, such as "explanation-example-exercises" (ETF, 2014a). A range of effective and less effective approaches were identified in the literature (Curee, n.d.; Higton et al., 2017; Maughan et al., 2016). These are collated in Tables 1 and 2.

It is important to note that effective strategies may differ between English and Mathematics. These tables reflect the research conducted into the approaches with measured outcomes, and this is the reason why classroom discussion, for instance, is not listed for English – it is unlikely that English is taught anywhere without classroom discussion, so its effectiveness has not been called into question.

There is increasing use of technology such as online learning, social media, and interactive apps on tablets or smartphones which give instant feedback (ETF, 2014a). Benefits of technology can be increased access to support, such as email contact with teachers, and streamlining activities, for example homework can be submitted online. Reviewing research into the use and benefits (or drawbacks) of technology was beyond the remit of this paper, but it may be that a large part of the value stems from engagement of learners.

Table 1: English delivery strategies

More effective	*Less effective*
Sustained support over time	Withdrawing students from mainstream lessons/ "catch up" study
Literacy interventions embedded in other curriculum areas	Mixed classes by vocational areas (e.g.,plumbers and hairdressers learning together).
Multi-strand approaches	
Peer-mediated support	
Whole language approaches	

Table 2: Mathematics delivery strategies

More effective	*Less effective*
Connectionist teaching methods[4]	Transmission teaching methods
Multi-strand approaches	Withdrawing students from mainstream lessons/ "catch up" study
Peer learning and support	Mixed classes by vocational areas (e.g., plumbers and hairdressers learning together).
Independent learning	
Interactive tasks	
Rotation of tasks/short tasks	
Use of technology	
Use of realistic contexts	
Classroom discussion	
Content embedded in vocational learning	

Contextualising

Williams et al. (2017) identified a variety of approaches to the delivery of English and Mathematics within FE institutions:

- English and Mathematics embedded in vocational areas and delivered by the same tutors;
- students grouped by subject areas and skills levels; and
- English and Mathematics taught separately from vocational classes by specialist teachers.

There is general agreement on the advantages of embedding English and Mathematics in vocational or real-life contexts (ETF, 2014a). Dalby and Noyes (2016) argued that Mathematics integrated with vocational learning encouraged students to change their beliefs about the irrelevance of the subject formed through their experience of GCSE. Higton et al. (2017) compared centralised and dispersed English and Mathematics departments and found that there were benefits and

4. Connectionist methods focus on dialogue-based teaching and encouraging students to make connections between Mathematical concepts while "reasoning out" challenges. Askew, Rhodes, Brown, Wiliam, and Johnson (1997) contrast connectionist methods with transmission and discovery teaching orientations.

drawbacks to each approach, with provider characteristics driving the decision to employ one or the other.

Embedding Mathematics content in vocational areas can help to clarify the relevance of abstract concepts and in such a way that students may not think of the activity as "Maths". Emphasising the importance of maintaining engagement, ETF (2014a) advised that the Mathematics content should not be prioritised over discussion in classroom settings.

Maughan et al. (2016) found some robust evidence for the positive effects of the use of realistic contexts in Mathematics and discussed Realistic Mathematics Education (RME) as one example of an intervention which emphasises contexts that interest students along with practical work and discussion. The RME approach was shown to improve understanding and the ability to solve mathematics problems (Searle & Barmby, 2012). ETF (2014a) outlined the benefits of real-life contexts for teaching Mathematics to vocational students, as increasing engagement, improving understanding and retention of information. They cautioned that contexts should be authentic and not contrived in order to maintain engagement.

Contexts can be used to teach Mathematics to vocational students in a number of ways:

- A realistic problem for students to solve, using skills they have already acquired.
- A realistic problem for students to solve in order to motivate and facilitate the learning of new skills.
- A realistic context to enable the students to see the point of the mathematics they are learning.
- A realistic context to help students make sense of abstract mathematics.
- A pseudo-context which looks as though it refers to real-life at first sight but does not.

(MEI, n.d., p.10)

Some providers were able to relate English and Mathematics to particular vocational courses, for example Higton et al. (2017) referred to a "Mathematics for plumbers" course. This is in line with the finding that English and Mathematics content tailored for certain vocational areas may be more effective than generic groups.

Teacher backgrounds

In 2015, Hayward and Homer found marked differences between those teaching Mathematics in sixth form colleges and FE colleges. Of those they surveyed, sixth form college Mathematics teachers tended to be younger, held higher Mathematics qualifications, mainly taught A Level Mathematics and were more likely to receive Continuing Professional Development (CPD) on a regular basis. In contrast, those teaching numeracy/functional skills in FE colleges were more likely to work part-time and hold qualifications in a non-Mathematical subject. Almost half of FE college teachers did not have an A Level or equivalent in Mathematics (Hayward & Homer, 2015). This broadly appears to be the same scenario for English, according to figures from Moss, Duncan, Harmey, and Muňoz-Chereau (2018).

Increased numbers of students of English and Mathematics at this level, fuelled by policy changes, have created a greater demand for teachers (Creese et al., n.d.). Coupled with prevailing teacher recruitment problems, this has resulted in reliance upon inexperienced teachers who may have had little time to develop the necessary knowledge and skills (Department for Education, 2017). Providers have needed to develop strategies for coping with this situation, including both recruiting new staff and ensuring existing staff are trained and supported (Creese et al., n.d.)

Strategy suggestions have come from a number of sources. Porter (2015) recommended that FE colleges should recruit English and Mathematics specialist teachers solely for the delivery of GCSE retakes and from there to build college-wide expertise through shared resources and skills. However, this would not provide the contextualisation of content discussed above. She also advised colleges to make the most of enhancement programmes and training bursaries, and to consider joining with other providers to share resources and expertise.

ETF reported the findings of the 2013 Commission on Adult Vocational Teaching and Learning, which characterised a problem with English and Mathematics in FE colleges as follows: "Specialist teachers have subject knowledge but lack vocational context, while vocational teachers are unable to embed literacy and numeracy" (ETF, 2015, p.14). ETF also commissioned the (2014b) *Strategic Consultation: Mathematics and English*, which brought to light some teachers' lack of confidence in their own ability to deliver English and Mathematics at Level 2.

Teacher support

In order to provide support to students, teachers themselves need development and support. A number of suggestions have been made to address those needs. In terms of learning content, ETF made the following recommendations for teacher development:

- A focus on aspirations for specific subgroups of students to enable in-depth exploration of evidence regarding how staff and student learning connects.
- Activities that help teachers develop an understanding of the underpinning rationale for Level 2 Mathematics and English teaching approaches side by side with their use.
- The provision/development of tools and resources to secure consistency.
- Structures that ensure work based professional learning about teaching is sustained over time.

(ETF, 2014b, p.13)

ETF grouped teachers' responses by theme on what types of support would help them deliver English and Mathematics up to Level 2:

- Contextualising/embedding English and Mathematics
- Development of teaching skills, particularly GCSE
- How to identify support needs of students with complex needs
- Integration between providers
- Staff forums for those delivering functional skills.

(ETF, 2014b, p.14)

In addition, specific support needs relating to English included the development of interactive and stimulating resources, and support for literacy teaching in the same range as the National Centre for Excellence in the Teaching of Maths. Mathematics teachers expressed a need for support in developing collaborative learning and discussion, which implies that delivery may largely still rely on transmission methods (ETF, 2014b).

Functional Skills

Much of the literature in this field concentrates on GCSE at post-16, therefore this section focuses on Functional Skills specifically. Functional Skills qualifications in England were introduced to replace key and basic skills qualifications. First teaching was available in 2010. Functional Skills qualifications are used for apprenticeships, as free-standing qualifications for 14–19-year-olds and for adult education (Ofqual, 2015). Although not designed to be stepping-stone qualifications, some providers use Functional Skills as such and they are listed as approved stepping-stone qualifications which meet the conditions of funding (Education and Skills Funding Agency, 2019). There is no formal progression pathway from Functional Skills to GCSE. Following redevelopment, revised Functional Skills qualifications have been taught from September 2019. The redevelopment focused on making Functional Skills qualifications more relevant for employment, and comparison with other qualifications easier, although there is still no formal progression to GCSE (Department for Education, 2018).

Student motivation and support needs

Robey et al. (2016) conducted focus groups with Functional Skills students in a variety of settings and found that like those retaking a GCSE post-16, many Functional Skills students had previous negative experiences of English and/or Mathematics. Therefore, much of the previous section relating to student engagement and motivation will apply to Functional Skills as well as to GCSE. ETF (2015) surveyed teachers and found that they believed "non-GCSE qualifications, and particularly Functional Skills, unlock motivation, give confidence and provide the tools to allow learners to fulfil their aspirations." This was confirmed by Robey et al. (2016), who found that learners themselves reported increased self-confidence and self-esteem through studying Functional Skills.

Enrolment on Functional Skills courses

The majority of providers sampled by Higton et al. (2017) routed students with E or lower in English and/or Mathematics onto Functional Skills courses, a pattern also reported by Williams et al. (2017). The reasons given for this decision were: to build up basic knowledge; concerns about students not achieving required grades at GCSE; and the effects of not reaching the required grades on student confidence and provider performance. However, some FE colleges preferred to enrol grade E students onto GCSE courses. The main reason given for this was to enable learners to concentrate on the GCSE, sometimes over two years, with the possibility of multiple resits to progress towards the required grade. Other reasons given were concerns over the changes to Functional Skills and the effectiveness of the qualification as a stepping-stone to GCSE.

Finally, Higton et al. (2017) reported that the requirement for those students achieving grade D/3 to study towards GCSE has resulted in organisational changes in FE colleges. Where previously two-thirds of students studied Functional Skills and one-third GCSE, that has now reversed.

Approaches to delivery

As with GCSE, alternatives to transmission teaching methods appear to be more effective for those studying Functional Skills. In focus groups, Functional Skills students mentioned their preference for smaller class sizes for the increased access to tutor support. They also found a mixture of classroom activities beneficial, including practical tasks and opportunities to discuss learning with their classmates (Robey et al., 2016).

Contextualising

For Functional Skills, contextualisation also appears to be a key issue. Students reported significant use of contexts in the teaching of Functional Skills, which encouraged them to use their skills in everyday life (Robey et al., 2016). For Mathematics, this could include comparing the cost of shopping or working out how much paint would be needed to decorate a room. For English, activities included writing a letter or an application for a job. These tasks are grounded in real life, and their relevance to the world of work or everyday life is clear to students.

Despite the emphasis on the benefits of context in the literature (ETF, 2014b; Maughan et al., 2016), there were mixed views on the use of context in assessment from students participating in focus groups. In both English and Mathematics, some found contexts useful while others found the context obscured the objective of the question (Robey et al., 2016).

Problems with context arose particularly for Mathematics, where learners described their time being taken up with reading the question or having to block out the words in order to access the numbers (Robey et al., 2016). Equally, contexts seen as irrelevant could be a distraction from the question objective. For Mathematics, the use of heavily contextualised questions can be an impediment, since English skills are required to access the questions. See Crisp, Johnson, and Constantinou (2018) for an exploration of how context, along with other features, can affect question quality.

Functional Skills and GCSEs

Views on Functional Skills versus GCSEs vary across students and teachers. Robey et al. (2016) reported that Functional Skills students themselves recognised the differences in content and delivery between Functional Skills and GCSEs. Functional Skills qualifications were considered by students to be relevant and useful, although some expressed the belief that GCSE was "better" and saw Functional Skills as part of their progression to GCSE.

Some teachers reported that it was easier to gain enthusiasm from students for Functional Skills as they believed they had already "failed" at GCSE and were averse to retaking it (Williams et al., 2017). ETF (2015) found that some FE teachers disliked the characterisation of Functional Skills as a stepping-stone to GCSE. The reasons for this were that the design and purpose of Functional Skills relate to preparation for the workplace and that practitioners felt progression should not be seen as the only or main outcome.

Addressing the challenges

It is clear that the biggest issues facing post-16 students and teachers of Level 2 English and Mathematics are student motivation and engagement, whichever qualification is being studied. Aside from students' attitudes, there is a complex set of challenges including varied support needs, students' differing levels of subject knowledge, identification of effective teaching strategies and the availability of provider resources and lesson time.

Building students' confidence can be achieved through challenging negative beliefs, focusing on effort rather than attainment, encouraging growth mindsets, using stepping-stone qualifications, and offering regular feedback and clear progression paths. Increased confidence is likely to improve motivation.

Access to appropriate support is vital for all students, especially those with additional needs. Generally, there is little difference in the support needs of learners of GCSE or Level 2 Functional Skills English and Mathematics at post-16. Levels of support required can be identified through diagnostic testing, but equally important is how support can improve rapport between teachers and students, encouraging dialogue and enabling students to seek help when needed.

As well as improving engagement, real-life contexts are likely to help students grasp concepts and retain information. Delivering content through real-life contexts may prove more difficult when teaching GCSE, whereas Functional Skills qualifications are already designed to develop life skills. For GCSE, if using a more applied approach, teachers must enable students to translate their learning into the more abstract concepts they are likely to encounter in the final assessment.

Students value an understanding of what is expected of them in exams, and how to access marks. Developing this understanding also serves to increase confidence since exam anxiety can be a longstanding problem, particularly for those who have not achieved the required grade in previous exams.

Improving engagement and motivation of students can also be achieved through deployment of stimulating learning activities and tasks. As shown in Tables 1 and 2, teachers can use a range of teaching methods to maximise the effectiveness of their provision. In addition, teachers may need differentiation strategies, given that classes may consist of students working at different levels. However, research shows that teachers would like support with development of resources and implementing engaging teaching strategies.

One of the main challenges faced by teachers is deciding what content and skills to target when confronted with short timescales. Initial diagnostic testing is clearly important for planning learning pathways, but formative assessment and feedback are also valuable for both teachers and students in order to maintain engagement and to track progress.

There is great diversity in the professional backgrounds of those who deliver English and Mathematics at Level 2 post-16, whether Functional Skills or GCSE, in FE colleges or other settings. Thus, the support needs of teachers are likely to depend on the qualifications of the teacher and the setting in which they are working. English and Mathematics specialists are likely to need support with integrating content into vocational contexts and with ways to motivate and engage disaffected learners. Vocational teachers who are not English and Mathematics specialists may need support to allow them to develop a clear understanding of the rationale behind the qualifications they deliver and confidence in their own knowledge in order to offer targeted support to students.

Despite the wealth of information they collated on successful approaches and recommended actions, Higton et al. (2017) warned that there is no single "best" approach to the challenges experienced by those involved in post-16 English and Mathematics. As demonstrated, post-16 students arrive from diverse backgrounds with mixed attitudes and abilities; they frequently need additional individual support and tailored provision. Providers are also diverse and much of the support offered to students and teachers is reliant on the policies and approaches of the provider.

References

Anderson, N., & Peart, S. (2016). Back on track: exploring how a further education college re-motivates learners to re-sit previously failed qualifications at GCSE. *Research in Post-Compulsory Education, 21*(3), 196–213.

Askew, M., Rhodes, V., Brown, M., Wiliam, D., & Johnson, D. (1997). *Effective Teachers of Numeracy: Report of a study carried out for the Teacher Training Agency*. London: King's College, University of London. http://musicmathsmagic.com/page2/files/EffectiveTeachersofNumeracy.pdf

Belgutay, J. (November, 2017). Most disadvantaged students fall behind peers in resit success. *TES*. Retrieved from https://www.tes.com/news/most-disadvantaged-students-fall-behind-peers-resit-success

Belgutay, J. (August, 2018). GCSE Resits: Students forced to sit exams 9 times. *TES*. Retrieved from https://www.tes.com/news/gcse-resits-students-forced-sit-exams-9-times

Bellamy, A. (2017, March). *Forced GCSE mathematics resits: Students' voices*. Paper presented at the British Society for Research into Learning Mathematics, Birkbeck College, University of London.

Creese, B., Litster, J., & Mallows, D. (n.d.). *A Strategic Guide for the delivery of GCSE English and maths to the 16–19 cohort*. Retrieved from Education and Training Foundation website: https://www.excellencegateway.org.uk/content/etf2019%20#

Crisp, V., Johnson, M., & Constantinou, F. (2018). A question of quality: Conceptualisations of quality in the context of educational test questions. *Research in Education, 0*(0), 1–24.

Curee. (n.d.). *What are the most effective forms of maths study and assessment post 16 (for different groups of learners)?* Retrieved from Society for Education & Training website: https://set.et-foundation.co.uk/digital-assets/qtlsmap/Resources/18/Effective_forms_of_maths_study_post_16%20(c).pdf

Dalby, D., & Noyes, A. (2016). Locating mathematics within post-16 vocational education in England. *Journal of Vocational Education & Training, 68*(1), 70–86.

Department for Education. (2016). *Level 1 and 2 attainment in English and maths by students aged 16–18: academic year 2014/15*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/525119/English_and_maths_SFR_2016_FINAL.pdf

Department for Education. (2017). Report of Professor Sir Adrian Smith's review of post-16 mathematics. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/630488/AS_review_report.pdf

Department for Education. (2018). *Functional Skills Reform*. Retrieved from https://api.excellencegateway.org.uk/resource/etf:2946

Dweck, C. S. (2008). *Mindsets and Math/Science Achievement*. Retrieved from Growth Mindset Maths website: http://www.growthmindsetmaths.com/uploads/2/3/7/7/23776169/mindset_and_math_science_achievement_-_nov_2013.pdf

Education and Skills Funding Agency. (2019). *16 to 19 funding: maths and English condition of funding*. Retrieved from https://www.gov.uk/guidance/16-to-19-funding-maths-and-english-condition-of-funding

ETF. (2014a). *Effective Practices in Post-16 Vocational Maths*. Retrieved from Education & Training Foundation website: https://www.et-foundation.co.uk/wp-content/uploads/2014/12/Effective-Practices-in-Post-16-Vocational-Maths-v4-0.pdf

ETF. (2014b). *Strategic Consultation: Mathematics and English*. Retrieved from Education & Training Foundation website: https://www.excellencegateway.org.uk/content/etf2706

ETF. (2015). *Making maths and English work for all*. Retrieved from FE Week website: https://feweek.co.uk/wp-content/uploads/2015/03/Report.pdf

FFT Education Datalab (2018) *GCSE Results 2018: Four Things to Look out for on Thursday*. https://ffteducationdatalab.org.uk/2018/08/gcse-results-day-2018-four-things-to-look-out-for-on-thursday/

Hayward, G., & Homer, M. (2015). *Profile of the Mathematics Teaching Workforce in the College Sector in England*. Retrieved from Gatsby Charitable Foundation website: http://www.gatsby.org.uk/uploads/education/reports/pdf/profile-of-the-maths-teaching-workforce.pdf

Higton, J., Archer, R., Dalby, D., Robinson, S., Birkin, G., Stutz, A., . . . Duckworth, V. (2017). *Effective practice in the delivery and teaching of English and Mathematics to 16–18 Year Olds*. Retrieved from University College London, Institute of Education website: http://dera.ioe.ac.uk/30578/1/English_and_Mathematics_to_16-18_year_olds.pdf

Impetus–The Private Equity Foundation. (2016). *The road most travelled? The 16–19 journey through education and training*. Retrieved from https://impetus-pef.org.uk/assets/publications/Report/Impetus-PEF-Position-Paper-Final-draft-Web-individual-pages.pdf

Jerrim, J., Greany, T., & Perera, N. (2018). *Educational disadvantage: How does England compare?* Retrieved from Education Policy Institute website: https://epi.org.uk/wp-content/uploads/2018/04/Educational-Disadvantage-England-EPI-IOE.pdf

Maughan, S., Smith, J., Mitchell, T., Horrocks, N., & Taylor, A. (2016). *Improving Level 2 English and Maths Outcomes for 16 to 18 Year Olds: Literature Review*. Retrieved from Education Endownment Foundation website: https://educationendowmentfoundation.org.uk/public/files/Presentations/Publications/16-18_Literature_Review.pdf

MEI. (n.d.). *GCSE Mathematics retake for vocational students: GCSE teaching from 2015*. Retrieved from Mathematics Education Innovation website: http://mei.org.uk/files/pdf/GCSE-vocational-students.pdf

Moss, G., Duncan, S., Harmey, S., & Muňoz-Chereau, B. (2018). *Current Practice in Using a System of Phonics with Post-16 Learners*. Retrieved from Education & Training Foundation website: https://www.et-foundation.co.uk/wp-content/uploads/2018/06/ETF18002_PHONICS_REPORT_80PP_FINAL_AW.pdf

Nye, Philip (2018, August 20). GCSE results 2018: Four things to look out for on Thursday [Web log post]. Retrieved from https://ffteducationdatalab.org.uk/2018/08/gcse-results-day-2018-four-things-to-look-out-for-on-thursday/

Ofqual. (2015). *Improving Functional Skills Qualifications.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/398441/2015-01-27-improving-functional-skills-qualifications.pdf

Ofqual. (2017). *Grading new GCSEs*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719124/Grading_new_GCSEs25.6.2018.pdf

Ofqual. (2018). *Guide to GCSE results for England, 2018*. Retrieved from https://www.gov.uk/government/news/guide-to-gcse-results-for-england-2018

Porter, N. (2015). *Crossing the Line: Improving success rates among students retaking English and maths GCSEs*. Retrieved from Policy Exchange website: https://policyexchange.org.uk/wp-content/uploads/2016/09/crossing-the-line.pdf

Robey, C., & Jones, E. (2015). *Engaging Learners in GCSE Maths and English.* Retrieved from Learning and Work Institute website: https://www.learningandwork.org.uk/wp-content/uploads/2017/01/Engaging-learners-in-GCSE-maths-and-English.pdf

Robey, C., Woodhouse, J., & Downes, S. (2016). *Functional Skills Reform Consultation: Findings of the learner focus groups*. Retrieved from Education & Training Foundation website: https://www.et-foundation.co.uk/wp-content/uploads/2016/12/Learner-focus-groups-report-FINAL.pdf

Searle, J., & Barmby, P. (2012). *Evaluation Report on the Realistic Mathematics Education Pilot Project at Manchester Metropolitan University*. Retrieved from Mathematics Education Innovation website: https://mei.org.uk/files/pdf/RME_Evaluation_final_report.pdf

Sezen, C. (August, 2018). 'GCSE resit data shows we need a policy rethink', *TES*. Retrieved from https://www.tes.com/news/gcse-resit-data-shows-we-need-policy-rethink

Skills Commission. (2016). *Spotlight On... young people with below average academic attainment and the skills sector*. Retrieved from Policy Connect website: https://www.policyconnect.org.uk/research/spotlight-young-people-below-average-academic-attainment-and-skills-sector

Vidal Rodeiro, C. (2018). Which students benefit from retaking Mathematics and English GCSEs post-16? *Research Matters: A Cambridge Assessment Publication, 25*, 20–28.

Williams, J., Hadjivassiliou, K., Marvell, R., Green, M., & Newton, B. (2017). *Effective curriculum practice at below Level 2 for 16/17 year olds*. Retrieved from Department for Education: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/628908/Effective_provision_for_low_attainers.pdf

Wolf, A. (2011). *Review of Vocational Education – the Wolf Report*. Retrieved from Department for Business, Innovation & Skills and Department for Education website: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf

Methods used by teachers to predict final A Level grades for their students

Tim Gill Research Division

Introduction

Prior to 2015, there was a requirement for teachers in centres in England to submit an estimated grade to the awarding organisation (AO) for all students undertaking Advanced (A) Level qualifications. This information was used as part of the evidence base for grading and for reviews of marking (Cambridge Assessment, 2013). Estimated grades are no longer collected by the AOs, but they still serve a number of purposes. Firstly, teachers are required to provide them as part of the university application process[1]. Secondly, estimated grades may be produced at several different points during an A Level course to monitor student progress, or serve as a motivational tool (Martinez, 2001). Finally, they may be used within the centre for teacher accountability purposes.

1. University admissions tutors use them to assess students' potential so that they can decide whether to make an "offer" of grades that the student needs to achieve to secure entry onto a course.

However, there is evidence that predicting A Level grades accurately is a task that teachers find difficult (see Gill & Rushton, 2011; Gill & Chang, 2013; Gill & Benton, 2015). This lack of accuracy may impact negatively on teachers' perceptions of the quality of marking for a qualification. The purpose of the research presented in this article was to understand more about how teachers go about the process of making grade predictions for their students, in order to help them make more accurate predictions. If teachers are able to make more accurate predictions, then this may increase their confidence in the reliability of marking.

This research was a replication of a previous study undertaken by Cambridge Assessment (Child & Wilson, 2015) which used a survey and interviews to investigate how teachers of A Level qualifications made predicted grades for their students. A further aim of that study was to calculate the accuracy of those predictions, using data collected from the survey. Since the original work was undertaken, there have been some significant reforms to A Level qualifications (see Ofqual, 2016) which are likely to have had an impact on how grades are estimated and on the accuracy of these predictions.

Context: Reforms to A Levels

One of the most important changes brought about by the reforms relates to the connection between A Levels and another qualification: Advanced Subsidiary (AS) Levels. Prior to the reforms, an AS Level counted as half of an A Level (in the same subject), and was assessed at the end of the first year of a two-year A Level course. It was also available as a stand-alone qualification for students who did not want to go on to take a full A Level. The reforms led to AS Levels being "decoupled" from the A Level, meaning that currently they do not count towards an A Level. Centres, therefore, have the choice of no longer offering the AS Level for their A Level students, or getting students to take the AS Level in Year 12 and then the A Level in Year 13 (which means they will be reassessed on some of the same content as the AS Level). The structure of the new qualifications is such that co-teaching of the AS and A Level is (theoretically) possible, so that students in the same Year 12 class in a subject can be planning to take the AS Level only, the A Level only, or both qualifications. However, Vitello and Williamson (2018) found that only just over half of the heads of department they surveyed thought that AS Levels were *actually* co-teachable. There is also evidence of significant falls in both uptake and provision of reformed AS Level subjects after they had been decoupled (Vitello & Williamson, 2018).

The reforms are important in the context of making predictions of A Level grades because it means that some centres no longer teach AS Levels in the subject, and are therefore not able to use that information to help them make predictions. According to the previous study (Child & Wilson, 2015), the AS Level grade was an important source of information for estimating A Level grades. Some centres still offer AS Levels to their A Level students, which may be partly because AS Level exams are useful practice for taking A Level exams. Even so, the decoupling may have had an impact on the way in which results of AS Levels are used to help make predictions.

Previous research

There has been little previous research which looked at how teachers go about the process of making grade predictions in centres in England, apart from the original study on which the present research was based (Child & Wilson, 2015). The analysis of responses to that questionnaire found that teachers tended to combine data from several different sources to make their predictions. AS Level grades were used by 94 per cent of the respondents and were generally thought to be the best predictor of A Level grades. Other commonly used sources of information included observations of the quality of work or of student commitment and performance in coursework and mock exams.

There is some previous research investigating the *accuracy* of grade predictions in England. Several reports from Cambridge Assessment (Gill & Rushton, 2011; Gill & Chang, 2013; Gill & Benton, 2015) compared the A Level forecast grade which was submitted to the AO with the final grade achieved. The percentage of accurate predictions varied from 55% in the 2011 report to 43% in the 2015 report. Inaccurate predictions were much more likely to be optimistic (varying from 33% in 2011 to 43% in 2015) than pessimistic (12% in 2011 and 14% in 2015). Similar results were found in research undertaken by Universities and Colleges Admissions Service (UCAS, 2013) on behalf of the Department for Business, Innovation and Skills, which compared the predicted grades sent to UCAS as part of the university application process with the final grade. They found that grades were accurately predicted 42% of the time, with 48% of predictions being optimistic. In the study by Child and Wilson (2015) predictions were more likely to be optimistic than pessimistic, and some respondents revealed that this was deliberate, to provide motivation for students.

Hopkin (2011) found that just using the AS Level grade to predict the A Level grade produced more accurate results than the teacher predictions for the AO (see Gill & Chang, 2013; Gill & Benton, 2015). However, AS Level grades were still only accurate around 55 per cent of the time. There was no tendency (in contrast to the teacher predictions) for inaccurate predictions to be more optimistic than pessimistic. However, the reports are not entirely comparable with one another because of the use of different datasets (teacher predictions were for all students taking A Levels from a specific AO, whereas the AS Level predictions were based on data from all AOs, but restricted to students taking at least three A Levels).

Wyness (2016) investigated the accuracy of predicted grades for university applications. She found that only 16 per cent of applicants were predicted the same points score from their best three A Level grades as they actually achieved. Almost all of the remaining applicants (75 per cent) were over-predicted (i.e., achieved lower grades than predicted). She also found that lower ability applicants were more likely than higher ability applicants to be over-predicted. One possible explanation for this tendency to over-predict (particularly for lower ability students) is that teachers are using the predicted grade as an aspirational target for students to aim for, so that for students who they feel are on the borderline of two grades the teacher will tend to choose the higher grade.

The most up-to-date data on the accuracy of predictions comes from UCAS (2017). This included a comparison of the accuracy of predictions sent to UCAS for reformed subjects (first tranche only) and for non-reformed subjects. Throughout the period investigated (2012–2017) the accuracy was worse for the reformed subjects (including during the pre-reform period). Post-reforms (i.e., 2017 only) the gap between reformed and non-reformed subjects was slightly wider, which might suggest that predicting A Level results is harder post-reform. However, the difference between the pre-reform and post-reform gap was only very small (exact figures were not available).

The main aim of the present study was to gather up-to-date information on grade predictions made in post-reform A Level subjects, in particular, the methods used by teachers to make grade predictions and the accuracy of the predictions. It was also hoped that more people would respond to the questionnaire than in the previous study, which would allow for more robust conclusions to be drawn. The subjects investigated in the current research were part of the first tranche of A Levels to be reformed, with first results in 2017. This meant that the centres contacted as part of the research were all teaching qualifications where the AS Levels had been decoupled from A Levels.

Methods

The methods for this research replicated those of the previous investigation by Child and Wilson (2015), by using a questionnaire sent to a large number of centres, followed by more in-depth interviews with teachers.

A Level subjects

We selected three A Level subjects offered by the Oxford, Cambridge and RSA (OCR) awarding organisation for this research. Two of these were the subjects that were used in the previous research (Chemistry and English Literature), so that direct comparisons could be made between these subjects pre- and post-reform. A third subject, Psychology, was also included because this is a very popular A Level which differs from the other subjects in that it does not include any non-exam assessment. This may have an impact on the way in which predictions were made.

Questionnaire

The first part of the data collection consisted of a questionnaire, to be filled in by centres offering OCR A Levels in either Chemistry, English Literature or Psychology.

Questionnaire design

The questions and structure of the survey were very similar to that used in the previous research project (Child & Wilson, 2015). There were two main sections:

- *Estimated grades for your students:* We asked participants to give their grade predictions for all their students who were completing their A Levels in 2018. They were also asked to give a ranking of where they believed students would reside within each grade (e.g., first, second).
- *How you decide estimated grades for your students:* We asked participants to say how important different sources of information were (from a list of options) and whether they asked anyone for advice in making their estimated grading decisions. We then asked them to describe in as much detail as possible their procedure for making estimated grading decisions. Finally, they were asked if there was any other information or support (not currently available) that would be useful.

Table 1: Number of centres invited, by subject

Subject	*Centres*
Chemistry	1,186
English Literature	508
Psychology	210
All	**1,904**

Participants

We recruited participants from several different lists of contacts provided by OCR. Where possible, we contacted the subject teacher directly. However, this was not always possible and for the majority of centres we used a general email address instead, with a request included to forward the email to a relevant teacher. The total number of centres contacted in each subject is shown in Table 1.

In May 2018, we sent an email to each centre to invite them to take part in the questionnaire. We advised participants that in order to complete the questionnaire they would need to know the estimated grades of their A Level students and that the questionnaire should take around 15 minutes to complete. We provided a link to the online questionnaire.

There were 54 respondents who completed the whole questionnaire, 38 for Chemistry, 8 for English Literature and 8 for Psychology. This was quite a low overall response rate (2.8 per cent). However, it is worth noting that a much larger number started it (went as far as to put in their name and their centre number), but did not finish. It seems likely that these people at least intended to complete the questionnaire. It was not clear if their decision not to continue was because they were concerned about revealing the candidate numbers of their students or because of the length of the task of putting in the predicted grades for all their students.

Interview

We asked questionnaire respondents if they were interested in taking part in a follow-up interview, after A Level results had been issued (in August 2018). The interview schedule had two sections:

- *How you decide and use predicted grades:* We asked interviewees about the relations between the different sources of information they said that they used in deciding predicted grades. This section included a specific question about the impact of the decoupling of AS Levels.
- *Questions on specific candidates:* For each interview, we identified three students who were of particular interest: one who achieved the same grade as their predicted grade; one who performed above predictions; and one who performed below predictions. We also asked if they had made any requests for a review of marking for any of their students.

Each interview took between 15 and 30 minutes to complete and was conducted by telephone. Each interviewee received a £20 book token as a thank you. In total, 45 of the 54 survey participants volunteered to take part in the follow-up interview, 32 for Chemistry, 7 for English and 6 for Psychology. Of these, we selected two for interview in each subject (six in total).

Results

Quantitative analysis of questionnaire responses

Comparison between predicted and actual grades

We asked respondents to list all students predicted to get each grade in turn (from A* to U), using their candidate number. This was so that after results had been issued we could compare this with the final grade for each student. We also asked respondents to rank students within each predicted grade; this enabled a complete ranking of all students in a centre to be generated based on their grade and their predicted position. We then compared this with the actual ranking (generated from total A Level marks).

There were a few issues with the recording of the predicted grades and ranking of students by respondents, perhaps because they did not understand the task they were being asked to do. Where the grades or rankings were duplicated or unclear, it was necessary to exclude the data. The remaining data was then merged with the final grade for each student. After doing this, there were 741 grade predictions, from 48 centres, with the breakdown shown in Table 2.

Table 2: Subject breakdown of matched grades

Subject	*Centres*	*Predicted grades matched to final grade*
Chemistry	33	524
English Literature	8	125
Psychology	7	92
Total	**48**	**741**

Table 3 presents a cross-tabulation of predicted and actual grades (all subjects together). This shows that for each predicted grade (apart from a grade U, which was only predicted for six students), the most common actual grade was the same grade (between 40 per cent and 50 per cent). For predicted grades A* to B the next most common actual grade was one grade below (i.e., over-prediction by one grade). However, for grades C to E the next most common actual grade was one grade above (i.e., under-prediction by one grade).

Table 3: Predicted vs actual grades (all subjects)

Predicted grade	*Students*	*Actual grade*						
		*A**	*A*	*B*	*C*	*D*	*E*	*U*
A*	105	**49.5**	41.0	9.5				
A	182	13.2	**42.9**	35.2	7.1	1.6		
B	205	1.0	20.5	**45.4**	26.3	5.9	1.0	
C	138	0.7	2.2	23.2	**40.6**	21.7	9.4	2.2
D	68			5.9	25.0	**47.1**	20.6	1.5
E	37		2.7	2.7	10.8	29.7	**45.9**	8.1
U	6					16.7	66.7	**16.7**

Figure 1 presents the accuracy of the predicted grades in each subject. It shows the percentage with exact agreement between predicted and actual grade, and the percentages which were optimistic or pessimistic and by how many grades. The numbers at the top of the bars show the absolute numbers of predictions in each category. This shows that for Chemistry and English Literature almost 50% of the predictions were accurate, but for Psychology only just over 30% were accurate. For Psychology around 45% were one grade optimistic, compared with about 25% for Chemistry and English Literature.

Tables 4 and 5 compare the accuracy of predictions from the current research with those from the previous analysis (Child & Wilson, 2015), for Chemistry and English Literature. This shows that in both subjects the levels of accuracy were much lower in the present study. In Chemistry, grade predictions in the present study were more likely to be either optimistic or pessimistic than in the previous study. In English Literature, predictions in the present study were more likely to be pessimistic, and were less likely to be optimistic.

Figure 2 presents a comparison of the predicted and actual rankings for each subject (all centres combined).

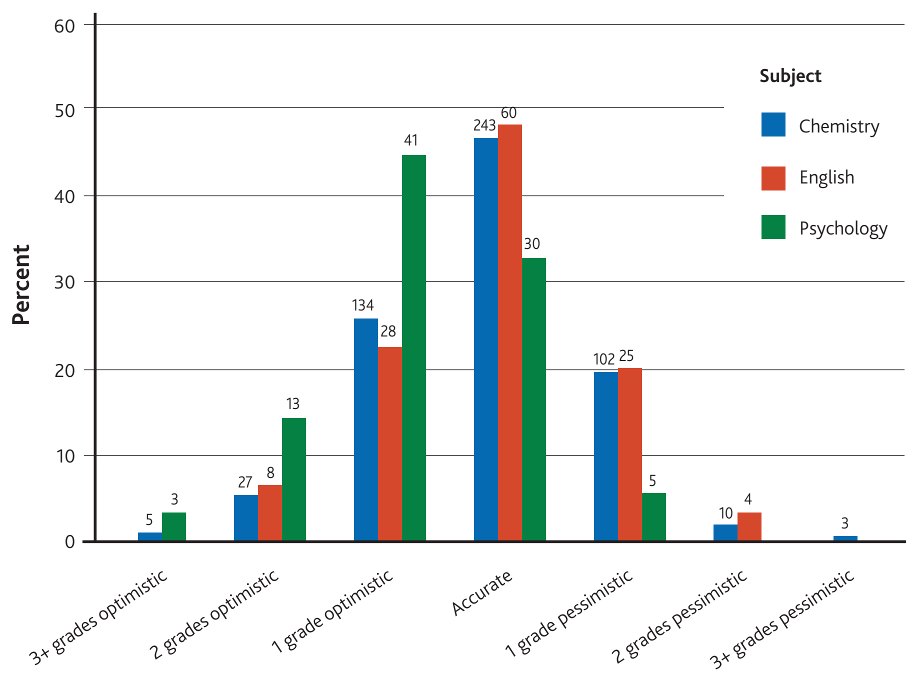


Figure 1:
Breakdown of accuracy of predicted grade, by subject

Table 4: Comparison of predicted grade accuracy from current and previous study (Chemistry)

	N	*Accurate (%)*	*Optimistic (%)*	*Pessimistic (%)*
Previous study	106	54.7	26.4	18.9
Current study	524	46.4	31.8	21.4

Table 5: Comparison of predicted grade accuracy from current and previous study (English Literature)

	N	*Accurate (%)*	*Optimistic (%)*	*Pessimistic (%)*
Previous study	133	57.1	30.1	12.8
Current study	125	48.0	28.8	23.2

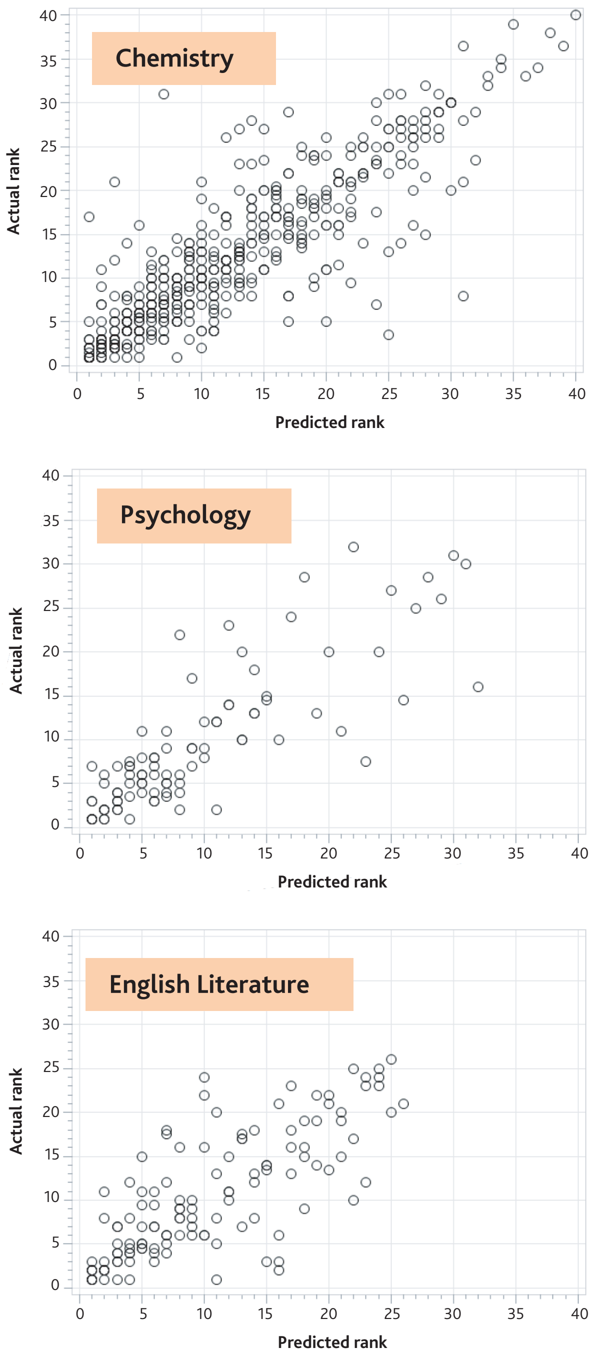


Figure 2: Predicted v actual rankings, by subject

The correlations were 0.87 for Chemistry, 0.76 for English Literature and 0.83 for Psychology. These results compare favourably with those reported in a meta-analysis into the accuracy of teachers' judgements (Machts, Kaiser, Schmidt, & Moller, 2016), which found average correlations of 0.61 between teachers' judgements of their students' cognitive abilities and their actual academic achievement. It is worth noting that the actual rankings were based on an overall mark for the qualification, and some students were very close in terms of marks. Therefore, it was not surprising that some teachers found it difficult to correctly rank these students.

How teachers decide the estimated grades for their students.

The first question in this section of the questionnaire asked how important different sources of information (as listed in the questionnaire) were in helping respondents make grade predictions. For each source of information they were given four possible options (*Very important; Somewhat important; Little importance;* and *Not used*).

There are a variety of different sources of information available to teachers to help them make estimates, which can be split into three categories, as outlined below:

1) *Statistical information:* A commonly used statistical method for tracking A Level students is the Advanced Level Information System (ALIS), which is provided by the Centre for Evaluation and Monitoring (CEM) at Durham University[2]. ALIS is an adaptive baseline test, which is usually taken at the start of Year 12.
The results are used (alongside the results of General Certificate of Secondary Education (GCSE) exams, which are taken at the end of Year 11) to provide information on the level a student is working at, and a predicted AS or A Level grade (including an estimate of the probability of achieving each grade). Other similar packages are also available, including Active Learning Practice For Schools (ALPS[3]), the Cognitive Abilities Test (CAT)[4] and Fischer Family Trust[5].

2) *Assessment performance:* Performance in previous assessments is usually a strong predictor of performance in a later assessment. Therefore, teachers are likely to use the GCSE or AS Level grades (either in the same subject as the A Level, or as an overall average grade) to help them predict A Level performance for students. Other assessments within the course may also prove useful, such as coursework, practical endorsements and formative assessments.

3) *In-class judgements:* More qualitative factors may also be used by teachers, such as the perceived motivation of students, their interest in the subject and the day-to-day quality of their work.

Figure 3 presents the results of the responses to this question, which suggest that the most important factors overall were the students' performance in mock (practice) A Level exams and observations of their quality of work and commitment. Each of these factors had more than 80% of respondents saying that they were very important or somewhat important. Despite the decoupling of AS Levels from A Levels for these subjects, over 60% of respondents said that AS Level performance in the same subject was very important or somewhat important. Two other factors ("AS Level mock examinations" and "other formative

2. https://www.cem.org/alis
3. https://alps.education/about-us/
4. https://www.gl-assessment.co.uk/products/cognitive-abilities-test-cat4/
5. https://fft.org.uk/fft/target-setting/

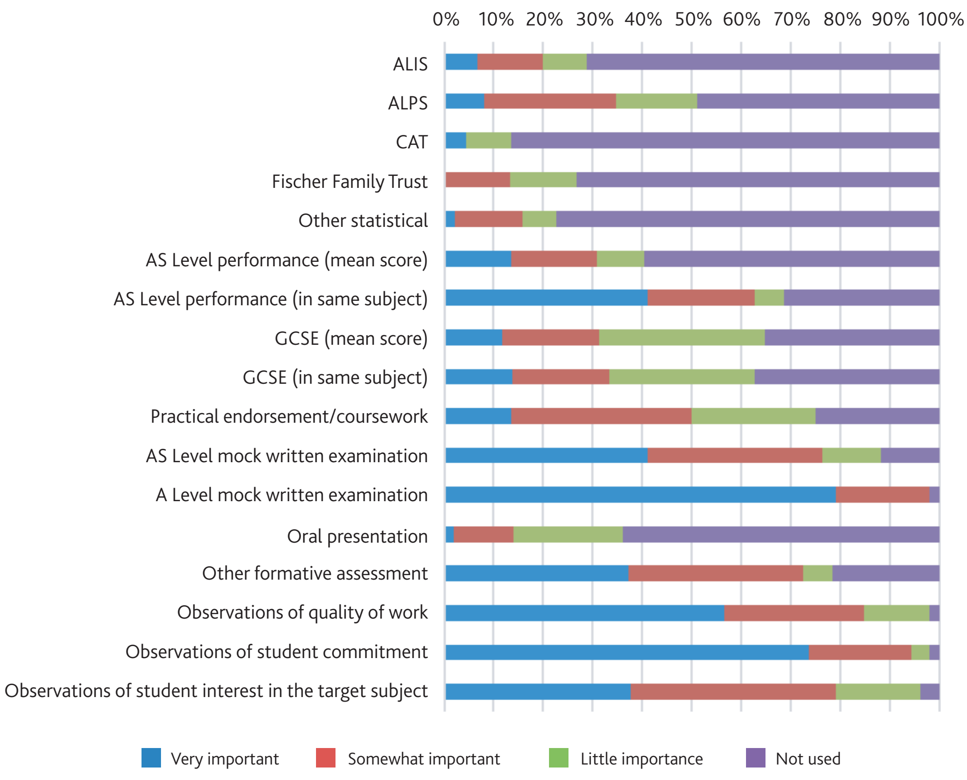


Figure 3: Importance of different sources of information in helping make grade predictions

assessments") were very or somewhat important for over 70% of respondents. Relatively few respondents said that they found statistical information (e.g., ALPS, ALIS) important. Another factor which was of little importance for most respondents was oral presentation. GCSE performance in the same subject or overall was only important to around a third of respondents.

Most respondents listed multiple sources of information in their answer to this question. The overall mean number of sources which were deemed to be very important was 4.3, and the mean number which was either very or somewhat important was 7.8. However, there was a significant variation in the number of different sources reported, varying from one to nine for very important and one to fifteen for very or somewhat important.

Respondents were asked to list any other sources of information that they used to help them estimate grades. There were 19 responses to this question. The most popular responses were around the use of end of topic/unit/chapter tests and the results of homework. Other responses included additional observations of students in the class, particularly in terms of their resilience/mental health and how well they responded to feedback from teachers about their assessments. Finally, one respondent mentioned that they used information on the progress of similar students from previous years.

The next question asked whether respondents asked for help in making predictions from other people or organisations. Again, they were given a list of possible sources of help, but were also able to add to this. Figure 4 presents the percentage of all respondents (n=54) saying they received advice from each of the different people. There were seven questionnaire respondents who did not give an answer to this question.

By far the most popular was other teachers in the same subject, which was selected by 74% of respondents. Head of department (30%) and teachers of a different subject (22%) were the next most popular selections. Very few respondents selected members of SLT, students, parents or the AO.

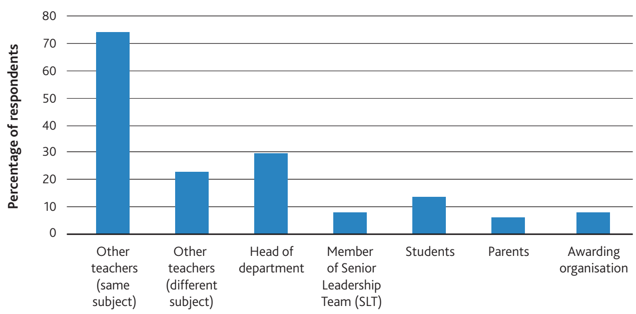


Figure 4: Sources of advice in making grade predictions

Qualitative analysis of questionnaire responses and interviews

There were a number of questions in the questionnaire and the interviews which required a more qualitative analysis. In the questionnaire, respondents were asked to explain in as much detail as possible how they go about the process of estimating the grade for a typical student. There were seven people who did not write anything for this question, leaving 47 responses in total. A number of related questions were then asked during the interviews to try and elicit further information about how predictions were made.

Sources of information for making predictions

Most respondents mentioned that they used several different sources of information, suggesting that it was a combination of factors which are taken into account. Two respondents explicitly noted that this was necessary because the individual sources of information were not reliable on their own. There were four main sources mentioned by respondents.

1) Results of internal tests/mocks

There was a fairly consistent message from respondents in terms of how they made their predictions. The most common sources of information mentioned were the results of end of topic or unit tests and the results of mock exams. For some, this was then combined with information on the achievement of similar students in the previous year:

> *There are end of chapter assessments all the way through AS and A-Level. These, along with mock exams in January and past papers in April/May provide an overall picture of what level they are working at* (Chemistry, questionnaire).

> *In-class assessments, mock examinations and in-class performance are of the utmost importance. I use my years of experience with many examination classes to judge pupils' ability against past pupils and the grades that they managed to achieve* (Chemistry, questionnaire).

How well students responded to feedback following tests or mock exams was also important to some respondents:

> *...The main starting point is how students are doing on assessments, in particular, on weekly timed essays on each of the two main papers since the start of 2018; and then how much they are improving their scores on these week on week following teacher feedback* (English Literature, questionnaire).

Mock exam results were rated more important than in-class tests or homework by some respondents, because they are the closest the students get to a real exam.

> *We try and complete full exam papers, mark them using the mark scheme and then use the grade boundaries from the exam board to set our predicted grades… because the closer we are to how they are going to be examined at the end, the better the quality of our predictions.. if they've been very good at their homework, but…they've sat looking at their books and they've got help from their friends and they've gone on the internet… that's not reflective of how they're going to be tested when we get to the end of A Level* (Chemistry, interview).

In the interview, participants were asked about when during the course they undertook mock exams and end of topic tests, and also what the sources of these were (e.g., past papers, textbooks, online resources). All interviewees said that mocks were undertaken in Year 13, with dates varying from January to April, and these were usually the final opportunity to change the predicted grade. Mocks were also undertaken in Year 12, generally at the end of the year. However, two interviewees noted that these did not affect the final predicted grade. Most said that they used past papers for mocks, usually from the previous year because they are still secure. However, one issue with the Year 13 mocks was that having longer exams meant it was difficult for the teachers to find time for a full mock of all the exam papers. For one interviewee (Psychology), this meant "cobbling together" a mock from various previous papers, focusing on the long answer questions.

The sources of the end of topic tests included past papers, specimen papers, textbook exemplars and then various online resources such as OCR Exam Builder[6], *Kerboodle*[7] and *Doddle*[8]. Four of the interviewees said that they converted the results of these tests into grades for their students. There were two methods for doing this: either using the grade boundaries from previous years (converted to a percentage of total mark) or using a straight percentage conversion (90 per cent = A*, 80 per cent = A etc.).

2) Student characteristics

Most respondents to the questionnaire made it clear how important it was to consider the characteristics of the students themselves when making predictions. This can either be what was described as "commitment", "attitude" or "work ethic", or it could be in terms of students' ability to cope with the stress of exams. As these factors are not something that can easily be tested, this was often combined with the teacher's own professional judgement:

> *The estimated grade is based mainly on the AS grade achieved, with adjustments made on the professional judgement of the teachers of that student as to their commitment, ability and willingness to improve* (Chemistry, questionnaire).

> *…My grade predictions are informed by my assessment of students' resilience and capacity to cope with exam pressure. Often, with this exercise, it's a candidate's mental health which has dictated both the grade prediction and the position in the rank order* (English Literature, questionnaire).

3) Verbal discussions

Some questionnaire respondents also used verbal discussions to get an impression of students' true understanding of topics:

> *Frequent marking of homework allows progression to be tracked and verbal discussions with the pupil allow me to gauge true understanding* (Chemistry, questionnaire).

> *More and more in Year 13 the quality of verbal responses influences predictions as I can judge their ability to apply existing knowledge to new situations – which is after all what it all boils down to* (English Literature, questionnaire).

4) Statistical information

There was little mention in questionnaire responses of the use of statistical information, such as ALIS, in helping to make predictions. Where it was mentioned, it tended to be used as a starting point only, which could then be adjusted as students progress.

> *Obviously prior attainment is an indicator but, unless a student's work matches their ALPS predicted grade, I would never predict, say, an A simply because that's what ALPS says they should get or because they did well in English at GCSE.*

> *ALIS is a starting point for where a student should be. The main evidence has been in class summative tests using exam papers* (Chemistry, questionnaire).

In the interviews, both English Literature teachers mentioned that the statistical information could be misleading:

> *…That's the starting point, we look at the ALPS but then we look at the exceptions, because it may be that they're very good at English and not so good at other subjects and this might have a negative impact on what the ALPS target grade is* (English Literature, interview).

In contrast, another interviewee (for Psychology) revealed that in the future they were planning to rely only on the ALIS predicted grade, because it seemed to be more accurate than their predictions were.

Role of AS grades and the impact of decoupling.

As seen in Figure 3, over 60 per cent of respondents said that AS Level grades in the same subject were "somewhat" or "very" important in

6. https://exambuilder.ocr.org.uk/marketing/about-us/
7. https://global.oup.com/education/secondary/kerboodle/?region=uk
8. https://www.doddlelearn.co.uk/

helping to make grade predictions. However, in their responses to the open-ended questions, there were only a few references made to AS grades amongst Chemistry respondents, even by those who said it was an important factor. However, almost all of the Psychology respondents who said AS grade was important mentioned them when explaining how they make predicted grades. This suggests that, amongst those who answered this question, AS Level grade was more important for Psychology than for Chemistry. One Psychology respondent said that the removal of AS Level grades will make the predictions much harder.

> *This will become a much more difficult process when students are not sitting real AS examinations (from this Summer in my school) and in truth I have no idea at this point how I will choose predictions of my current Year 12s come September/October*
>
> (Psychology, questionnaire).

Interviewees were asked specifically about the effect of decoupling on making predictions. Only one of the centres still offered the AS Level in the subject, and this was only done on very rare occasions. Three interviewees said that the centre offered the AS Level for the first year or two following reform, but were no longer offering it, or were stopping this year.

Several interviewees agreed that they were less confident about their predictions following the decoupling, because in the pre-reform situation they would have had the results for 70 per cent of the course (AS exam plus coursework) when making their predictions. Additionally, one mentioned that the AS grade was used to inform predictions at the start of Year 13 because it was the "most real life exam they are going to sit." Any internal exam, however stringent, will not have the same high-stakes nature. However, one interviewee (English Literature) thought that decoupling was not going to make predictions harder, because some students coast during the first year, so their AS grade can be an underestimate anyway.

The final question in the questionnaire asked whether there was any further information or support that the respondents would find useful in making predictions. There were 12 responses to this question, which covered a range of different issues:

- Two respondents asked for more past papers but recognised that this was difficult with a new specification.
- Several mentions were made of A Level grade boundaries, and what each boundary should mean in terms of skills developed.
- Two teachers made reference to mark schemes (or guidelines on marking) and how they are applied.
- One respondent requested data on the performance on each question for students achieving different grades, to see "where each grade is typically gaining or losing marks."
- There was one request to provide more accurate sample scripts, with one per grade, per paper for each exam series. This respondent did not believe that the A*/A exemplars were written by students or written under exam conditions.

Reflections on results

Following the publication of results, the interviewees were asked to reflect on the performance of their students compared to the predictions. In terms of their overall perceptions, all but one of the interviewees thought that students had mostly performed to expectations. The one exception to this (Psychology) said that the students predicted to receive low grades tended to do better than expected and the students predicted an A* mainly achieved an A.

Interviewees were also asked about the performance of specific students and, if possible, to give reasons for why they performed as they did. For each interview, three students were identified: one who achieved the same grade as their predicted grade, one who performed above predictions and one who performed below predictions. For those whose final grade matched their predicted grade, there were only two reasons which were mentioned by the interviewees. Firstly, the students in question had performed consistently at that grade throughout the course, or throughout Year 13. Secondly, the statistical prediction (ALIS or ALPS) for those specific students had been correct.

For students whose predicted grade was higher than their actual grade, there were several reasons given by teachers for the inaccurate prediction. Some of these related to events occurring after predictions had been made. For example, personal circumstances, or the student prioritising Science subjects (over English Literature) because they wanted to do Physics at university. Other reasons related more to unexpected performance on the day, such as poor performance on one paper, and being only a few marks below the grade boundary. One interviewee talked more generally about under-performing students rather than describing a specific example. For him, there was a certain amount of looking back with hindsight:

> *I think our grade boundaries have been a bit too generous, so we've sort of got students predicted into a 'C' where actually they've ended up with a 'D' and if we looked back at what they did when we predicted them a 'C', it probably was a 'D'* (Chemistry, interview).

This interviewee also suggested another reason, which came from his perspective as the head of department:

> *Based solely on January mock exams we were fairly close to actual outcomes, but when you looked at the final predicted grade which had a bit of teacher input, so if you like, he's been doing well for the last three months, so he got a 'C' in the exam, but actually I think he might be a bit better than that, they tended to be a bit inflated*
>
> (Chemistry, interview).

It is worth noting that this interviewee was under the mistaken belief that they were still required to send in predicted grades to the AO and that these might be used in determining a final grade for a student who missed an exam. This led him to say that they were "more positive, rather than negative" in their predictions. Finally, there was one interviewee who could not explain why the prediction was wrong.

For students who achieved a higher grade than predicted, there were again a number of different reasons suggested. One interviewee for English Literature had not predicted a higher grade (A*) for the student because they felt their "simplistic" writing style would hold them back. In contrast, the other English teacher said that their student wrote very fluently, but was worried about a slight "superficiality" in her writing, which was "disguised by the quality of her prose." Another reason given was that the student in question worked incredibly hard towards the end of the course, motivated by a university offer. One of the interviewees for Psychology did not predict A* grades as a rule and seemed to think that the new A Levels meant that it was harder to get an A*. Finally, there were two interviewees who could not explain why their predictions were wrong.

The final question asked the interviewees whether the results for 2018 would affect their approach to making predictions in the future. Almost all of them said that they would alter their approach to some degree. Two interviewees mentioned changing their approach for specific grades following under- or over-prediction at that level:

> *Where we've really had a problem this year…is the C/D and the D/E borderline, I think we've been too generous there…so we reviewed that for this year and we're gonna [sic] sort of concentrate on making sure [the grade boundaries] are higher* (Chemistry, interview).

> *We're going to…hold back on predicting A stars because we think that's where we got caught out the most this year, is the number of A stars we predicted and didn't get* (Psychology, interview).

One interviewee said that in the future they will rely on ALIS grades only to make predictions, as this year these were more accurate than the centre predictions. Other future changes in the approach to making predictions included more rigorously keeping data from exam questions, and reviewing the use of Kerboodle and using more past papers instead.

Discussion

Teachers' predictions of their students' A Level grades were, on the whole, fairly accurate. Across all three subjects, 44.9% of predictions were correct and 90.1% were within one grade. Where they were not correct, the predictions tended to be optimistic (35.0%) rather than pessimistic (20.1%). This pattern fits in with previous research (Child & Wilson, 2015; Gill & Benton, 2015). However, it is worth noting that the level of complete agreement was lower in this study than in the previous analysis undertaken by Child and Wilson (2015) three years ago, which found 54.7% of grades were correctly predicted in Chemistry and 57.1% in English Literature. The equivalent figures in the current study were 46.4% and 48% respectively. It is not known why the accuracy was lower, but it may be due to the decoupling of AS Levels, meaning that for many centres they did not have the AS Level result to help in the prediction. It may also be because teachers were still getting to know a relatively new specification.

Breaking the analysis down by subject, it was found that the accuracy of predictions was very similar in Chemistry and English Literature, but was considerably lower in Psychology, where only 32.6 per cent of predictions were correct. Over-predictions were also much more likely in Psychology than in the other two subjects. It is not clear why accuracy was so much lower in Psychology, but it may be related to the lack of non-exam assessment, which potentially made predictions more difficult.

In terms of the accuracy of the rankings data, there was a reasonably good association between the predicted and actual ranking of students, with correlations of between 0.76 and 0.87. This was despite the fact that, according to respondents to the previous survey (Child & Wilson, 2015), they did not usually rank students in making their predictions, so this process may have been a new experience for them. However, it may be that the respondents were, in fact, undertaking a ranking of sorts, even if they were not aware of it. According to Laming, "there is no absolute judgement. All judgements are comparisons of one thing with another" (Laming, 2004, p.9). In the context of making grade predictions, this might mean that teachers were comparing students in their current cohort with an internalised standard of what, for instance, an "A grade" student looks like, or with "A grade" students from previous years.

The findings from the questionnaire and interviews were quite consistent in terms of which factors were most important in helping teachers make their predictions. These were mock exams, in-class tests, and student characteristics such as attitude and how they cope with the stress of exams. Mock exams were seen as particularly important because with the removal of AS Levels in many centres, these were the closest experience that students had to a real exam. Examples were given of students who performed at a particular grade on in-class tests or homework but were given a lower predicted grade because of poorer performance in mocks. However, there was still concern from some respondents about the lack of high stakes for mock exams, which might mean that students did not always treat them as seriously as the real exams. Statistical predictions such as ALIS or ALPS were used as a starting point for predictions by some teachers, but always with the belief that the prediction could be over-written if the evidence of in-class work or mocks pointed to a different grade.

The interviewees revealed that mock exams and in-class tests were usually based on questions taken from past papers, or specimen papers. Requests were made for more of these to be made available and, by one interviewee, for an OCR specific source of questions. Therefore, it is important for OCR to continue to ensure the quality of these resources, in terms of alignment with content coverage and topics in the actual exam paper, and providing guidance for teachers to help them make reliable marking judgements. For example, the extra resources requested by respondents were mainly related to understanding how to assess students work better, either by improved understanding of the mark scheme or better understanding of what performance at each grade should look like.

It was not surprising that most of the interviewees said that the decoupling of AS Levels made them less confident about the accuracy of their predictions, particularly as all but one of them had stopped offering the AS Level in their subject. Prior to reforms, the AS Level was worth half of the A Level and was, as one interviewee put it "the most real life exam they are going to sit" prior to taking the A Level exam(s). It is worth noting that the predictions in both Chemistry and English Literature were less accurate than when the same analysis was undertaken prior to the decoupling (Child & Wilson, 2015). We cannot know for certain that the decoupling caused this fall in accuracy, but it was likely to have had a negative impact. It was unfortunate that none of the interviewees were in centres which continued to offer the AS Level, so that the effect of the decoupling in these centres could have been investigated, in terms of the way this information was used in making predictions.

There were many different reasons given for why students did not perform as expected. Some of these related to factors which influenced performance after predictions were made, and therefore there was no way that the teachers could have predicted these. This worked in both ways, with better than expected performance from a student who worked very hard at the end of the course, and worse than expected performance by students who did little work before the exam. Other reasons acknowledged the fact that students sometimes just perform unexpectedly (well or badly) on the day, and there were some instances where the teacher was at a loss to explain their students' performances. This highlights that grade predictions are inherently unreliable, because of the multitude of factors which are beyond the teacher's control.
In considering why predictions were inaccurate, it is worth noting that there was no mention made of poor marking by OCR. Indeed,

one interviewee commented that changes to marks after a review of marking were rare because the marking was generally accurate in the first place. Even the respondent who felt that, overall, their students had not performed to expectations thought that this was an error on their part and talked about the need for them to improve predictions.

Most of the interviewees said that they would change their approach to making predictions in the future, following the results for their students in 2018. Some were making quite radical changes, such as relying on ALIS predictions only. Others were focused on changes to predictions at particular grades. We were encouraged to see that teachers were continually adjusting their approach in an effort to improve their predictions. However, we hope that these decisions were not knee-jerk reactions to inaccurate predictions. There are many reasons why student performance can be unexpected in any one year (e.g., marking reliability, exam coverage of curriculum, personal circumstances) and therefore, it would be wise for centres to look at the accuracy of their predictions over several years, rather than just the previous year. Furthermore, it may be that predictions were influenced by the mistaken assumption that results within a centre should be reasonably stable year-on-year. Previous research (Crawford & Benton, 2017) has shown that some centres can experience high levels of volatility in (GCSE) results between consecutive years. It may be that results in a centre which differ from predictions may be partly due to this natural variation in year-on-year results. A further area of research would be to look at whether predictions are less accurate when performance in a centre in a particular year is unexpected compared to previous years.

Finally, it is worth noting one limitation with the results presented in this article. Responses to the questionnaire were more likely to come from smaller centres. For each subject, the mean value of centre "size" (number of students finishing the A Level in 2018) was higher in the whole cohort (19.7 for Chemistry, 18.6 for English Literature and 25.8 for Psychology) than in the sample of respondents (16.0, 15.6 and 13.1 respectively). Furthermore, the maximum centre size amongst the sample data was only 40 for Chemistry (compared with 423 amongst all centres), 26 for English Literature (compared with 180) and 32 for Psychology (compared with 378).

The most likely reason for this pattern was because of the size of the task for a larger centre (listing predicted grades for all students). This issue was highlighted by an email from one teacher stating that they would not consider completing the questionnaire because they had 260 students sitting the exam[9]. This may have been an issue if, as one teacher commented, it was easier for smaller centres to make predictions because they know individual students better.

In hindsight, it probably would have been better to reorder the questionnaire so that the questions about how the respondents made their predictions came before the questions about the predicted grades for each student. As previously noted, there was a large number of people who started the filling in the questionnaire, but did not get past this section. Changing the order may have led to more responses to the questions about methods, particularly from larger centres. Another way to encourage responses from larger centres would have been to say that it was only necessary to enter the predicted grades for one or two classes, rather than for all A Level students.

In conclusion, the research reported in this article found that teachers were reasonably accurate in their predictions of final A Level grades for their students, and were more likely to be optimistic than pessimistic. Prediction is clearly a complex process, involving the weighing up of different factors, which has seemingly become more difficult since A Levels were reformed. Therefore, it is of added importance for AOs to provide the best available support to teachers.

9. The teacher was encouraged to complete the questionnaire anyway, referring to one or two classes only, not the whole cohort.

References

Cambridge Assessment. (2013). *A simple guide: Assuring OCR's marking accuracy*. Retrieved 24th August 2018 from http://www.cambridgeassessment.org.uk/Images/143058-a-simple-guide-to-marking-and-grading.pdf

Child, S., & Wilson, F. (2015). *An investigation of A level teachers' methods when estimating student grades*. Cambridge Assessment internal report. Cambridge, UK: Cambridge Assessment.

Crawford, C., & Benton, T. (2017). *Volatility happens: Understanding variation in schools' GCSE results*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Gill, T., & Benton, T. (2015). *The accuracy of forecast grades for OCR A levels in June 2014. Statistics Report Series No. 90*. Cambridge, UK: Cambridge Assessment.

Gill, T., & Chang, Y. (2013). *The accuracy of forecast grades for OCR A levels in June 2012. Statistics Report Series No. 64*. Cambridge, UK: Cambridge Assessment.

Gill, T., & Rushton, N. (2011). *The accuracy of forecast grades for OCR A levels. Statistics Report Series No. 26*. Cambridge, UK: Cambridge Assessment.

Hopkin, R. (2011). *Predicting A level grades using AS level grades. Statistics Report Series No. 29*. Cambridge, UK: Cambridge Assessment.

Laming, D. (2004). *Human Judgment: The Eye of the Beholder*. London, UK: Thomson.

Machts, N., Kaiser, J., Schmidt, F. T. C., & Moller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85–103.

Martinez, P. (2001). *Great expectations: setting targets for students*. London, UK: Learning and Skills Development Agency.

Ofqual. (2016). *Your qualification, Our regulation: GCSE, AS and A level reforms in England*. Coventry, UK: Office of Qualifications and Examinations Regulation.

UCAS. (2013). *Investigating the accuracy of predicted A level grades as part of the 2010 UCAS admission process*. London, UK: Department for Business, Innovation and Skills.

UCAS. (2017). *End of cycle report 2017: Qualifications and competition*. Cheltenham, UK: UCAS.

Vitello, S., & Williamson, J. (2018). *AS level trends 2017*. Cambridge Assessment Research Report, Cambridge, UK: Cambridge Assessment.

Wyness, G. (2016). *Predicted grades: accuracy and impact: A report for University and College Union*. London, UK: University and College Union.

Research News

David Beauchamp Research Division

Conference presentations

European Association of Curriculum Studies (EuroACS)

The European Conference of Curriculum Studies conference took place in June 2019 at Maynooth University, Ireland, and gathered researchers from the European Association of Curriculum Studies research network. The themes discussed during the conference were the origins, trajectories and practices of Curriculum. Sinéad Fitzsimons from the Research Division presented work co-authored with colleagues Martin Johnson and Tori Coleman: *Curriculum Progressions in Education in Emergencies: Challenges, opportunities and theoretical underpinnings.*

Global Perspectives: re-imagining education

This event took place in June 2019 at the University of Worcester, UK, and focused on themes related to the impact of globalisation and internationalisation on sustainable education. An adapted version of the paper above was presented by Tori Coleman: *Developing a learning framework for Education in Emergency contexts.*

Journal of Vocational Education and Training (JVET)

The 13th Journal of Vocational Education & Training conference took place in Oxford in June 2019 and gathered researchers around the theme of *Researching Vocational Education and Training*. Martin Johnson and Sylvia Vitello from the Research Division presented the following papers:

Martin Johnson, Research Division and Neil Mercer, Faculty of Education, University of Cambridge: *Learning to think alike: Using Sociocultural Discourse Analysis to explore examiners' standardised professional discourse*.

Sylvia Vitello and Carmen Vidal Rodeiro, Research Division: *How do vocational qualifications fit into students' programmes of study following recent governmental reforms to 14–18 education?*

Council of Chief State School Officers (CSSO) National Conference on Student Achievement (NCSA)

This annual conference was created in 1971 and draws attendees from various areas of assessment, including federal agencies, test publishers, educational consultants, university faculty and organisations supporting technology, education and business. This year's theme was: *Measure What Matters, and Create Accountability for Equity*. Thanos Patelis (HumRRO), Stuart Shaw (Cambridge Assessment International Education), Russ Keglovits (Nevada DoE), and Allison Timberlake (Georgia DoE) presented a paper: *The role of context on college and career readiness indicators in accountability systems.*

Applied Human Factors and Ergonomics Conference

This conference took place in July 2019 in Washington, USA. Topics that were discussed included human performance, root cause analysis, automation error, and cognitive and statistical modelling of human error. Irenka Suto from the Research Division presented a paper (co-authored with Joanne Ireland): *'To err is human' but it's time to go deeper. An analysis of human and system level challenges in the educational testing industry.*

IAFOR/The European Conference on Education

Held in July 2019 in London, the European Conference on Education was dedicated to the topic *Independence & Interdependence*. This conference is organised into thematic streams such as Teaching & Learning, Educational Structures, Community & Society, Language & Culture, Psychology, Mind & Brain, Innovation & Technology. Emma Walland from the Research Division presented a paper based on her research with colleague Ellie Darlington: *Impact of A Level reform on post-16 provision*.

European Association for Learning and Instruction (EARLI)

This biannual conference was hosted at RWTH Aachen University, Germany, in August 2019. The researchers were given the opportunity to present ideas on the topic *Thinking Tomorrow's Education: Learning from the past, in the present and for the future*. Filio Constantinou, from the Research Division, shared her work: *Why are some subjects less popular than others? Extending the debate*.

European Conference of Educational Research

This conference was held at the University of Hamburg, Germany, in September 2019, and focused on *Education in an Era of Risk – the Role of Educational Research for the Future*. Joanna Williamson from the Research Division presented a paper based on her research with colleague Simon Child: *How can mark scheme design support reliable and valid school-based assessment?*

Cambridge Schools Conference

The triannual Cambridge Schools Conference allows teachers from various countries to gather and discuss common challenges in education and share knowledge and new concepts in order to achieve professional development. The theme for the conference that took place in Cambridge, UK, in September, was *Evaluating impact: how effective is our school and classroom practice?* Researchers from Cambridge Assessment International Education presented two papers:

Judith Roberts, Stuart Shaw, and Sarah Nelson, Cambridge Assessment International Education: *The Cambridge International Curriculum Impact Framework*.

Stuart Shaw, Cambridge Assessment International Education: *Measuring academic language proficiency - towards a new scale?*

Second International Textbook Summit

In June 2019, together with the Icelandic Ministry of Education and Culture, Cambridge Assessment convened the second International Textbook Summit. With eight nations attending, it comprised a major contribution to the continued growth of interest in the form and function of learning materials, and the markets in which they exist. The addition of consideration of funding and sustainability of supply - summarised as 'market conditions' - was extremely important. While the first summit in England, held in 2018, considered quality and function, issues relating to the structure of funding and supply in different national settings have not previously been explored sufficiently, and are almost entirely absent from the dwindling body of textbook research. The second summit corrected this international omission, and dealt with key public policy matters of private-state relationships and sustainability of supply. A third summit for 2020 is being planned.

Centre for Evaluation and Monitoring joins the Cambridge family

In June 2019 the Centre for Evaluation and Monitoring (CEM) was jointly purchased by Cambridge Assessment and Cambridge University Press. Formerly part of Durham University, CEM is one of the largest and longest established research groups providing formative assessments for children of all ages and is used by education professionals in over 70 countries around the world. A not-for-profit organisation, CEM joins the Cambridge family following the successful conclusion of a consultation with employees. The CEM team will remain in the North East of England but work closely with new colleagues in Cambridge and around the world.

Publications

The following reports and articles have been published since *Research Matters*, Issue 27:

Child, S., & Shaw, S. (2019). A purpose-led approach towards the development of competency frameworks. *Journal of Further & Higher Education*. doi: 10.1080/0309877X.2019.1669773.

Constantinou, F. (2019). The construct of language competence over time: using high-stakes tests to gain insight into the history of L1 education in England. *Language and Education*, 1–15. Available online at https://www.tandfonline.com/doi/pdf/10.1080/09500782.2019.1597106?needAccess=true

Crisp, V., Johnson, M. & Constantinou, F. (2019). A question of quality: Conceptualisations of quality in the context of educational test questions. *Research in Education*, *105*(1), 18-41. https://journals.sagepub.com/doi/10.1177/0034523717752203

Greatorex, J., Rushton, N., Coleman, T., Darlington, E., & Elliott, G. (2019). *Towards a method for comparing curricula*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment. Available online at https://www.cambridgeassessment.org.uk/Images/549208towards-a-method-for-comparing-curricula.pdf

Johnson, M., & Mercer, N. (2019). Using sociocultural discourse analysis to analyse professional discourse. *Learning, Culture and Social Interaction*, *21*, 267–277. Available online at https://www.sciencedirect.com/science/article/pii/S2210656118302940

Johnson, M., & Rushton, N. (2019). A culture of question writing: Professional examination question writers' practices. *Educational Research, 61*(2), 197–213. Available online at https://www.tandfonline.com/doi/full/10.1080/00131881.2019.1600378

Vidal Rodeiro, C.L. (2019). *Popularity of A Level subjects among university students*. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.

Further information on all journal papers and book chapters can be found on our website: www.cambridgeassessment.org.uk/our-research/all-published-resources/journal-papers-and-book-chapters/

Reports of research carried out by the Research Division for Cambridge Assessment and exam boards, or externally funded research carried out for third parties, including the regulators in the UK and many ministries overseas, are also available from our website: www.cambridgeassessment.org.uk/our-research/all-published-resources/research-reports/

Data Bytes

Data Bytes is a series of data graphics from Cambridge Assessment's Research Division, designed to bring the latest trends and research in educational assessment to a wide audience. Topics are often chosen to coincide with contemporary news or recent Cambridge Assessment research outputs. All our *Data Bytes* can be found on our website: www.cambridgeassessment.org.uk/our-research/data-bytes

The following *Data Bytes* have been published since *Research Matters*, Issue 27:

- May 2019: *Summarising Parliamentary Questions*
- July 2019: *Students' favourite subjects around the world*

Sharing our research

Cambridge Assessment is home to the largest research capacity of its kind in Europe. In addition to our research publications, we provide a wealth of resources on our website including::

- *Insights* – a platform for sharing our views and research on the big education topics that impact assessment around the globe at www.cambridgeassessment.org.uk/insights/
- *Research Bytes* – short presentations and commentary based on recent conference presentations at www.youtube.com/user/CambridgeAssessment1
- *#CamEdLive* – our online live debates – at www.youtube.com/user/CambridgeAssessment1
- Podcasts at www.youtube.com/user/CambridgeAssessment1
- Blogs at www.cambridgeassessment.org.uk/blogs

You can also learn more about our recent activities from Facebook, LinkedIn, Twitter and Instagram.

Finally, a reminder that all previous issues of *Research Matters* are available to download, in full and as PDFs of individuals articles, from our website: www.cambridgeassessment.org.uk/research-matters

Contents / *Issue 28* / Autumn 2019

Cambridge Assessment
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA
United Kingdom

+44(0)1223 553985
researchprogrammes@cambridgeassessment.org.uk
www.cambridgeassessment.org.uk

2465294696

ISSN: 1755–6031