Research Matters / 41

A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridge.org/about-us/research-matters

The impact of A Level exam scheduling on performance

Tim Gill

To cite this article: Gill, T. (2026). The impact of A Level exam scheduling on performance. *Research Matters: A Cambridge University Press & Assessment publication, 41*, 32–51. https://doi.org/10.17863/CAM.127731

Abstract:
This article examines whether the scheduling of A Level examinations in England influences student performance. Using data from the OCR awarding organisation from 2016 to 2019 and 2023, we analysed the exam results of candidates taking at least three A Levels. Three exam scheduling measures were considered: (i) days since the candidate's previous exam; (ii) days since the candidate's first exam; and (iii) number of previous exams taken by the candidate. Multilevel regression models controlled for concurrent attainment, gender, subject group, centre type, and exam session. Results show no evidence of a cumulative fatigue effect linked to days since the first exam. However, performance was slightly lower when multiple exams occurred on the same day, with gaps of one or more days associated with modest improvements (2 to 3.5 percentage points). The number of previous exams had a small negative effect overall but interacted strongly with concurrent attainment: high-attaining students performed better as the number of exams increased, while low-attaining students performed worse. Overall, the findings suggest that current scheduling practices have limited impact on performance, though reducing same-day exams could offer minor benefits.

The impact of A Level exam scheduling on performance

Tim Gill (Research Division)

Introduction

A Levels are the most common post-16 qualification taken by students in English schools. Most A Level students in England take at least three subjects and most of these have three examined components, giving a total of nine exams, which are all taken at the end of the two-year courses. The A Level summer exam series usually runs for about 45 days, but most students will take all their exams in a considerably shorter number of days. As such, they are required to revise for and take a substantial number of exams in a short period of time. This is referred to in this article as their exam *load*. Because of this, students may have to make trade-offs as to where they concentrate their effort and may experience exam fatigue, particularly towards the end of the series. Thus, there is the potential for students' performance to be affected by their exam schedule.

There is limited previous research looking at the impact of exam scheduling on results. Goulas and Megalokonomou (2018) investigated this topic in the context of exams in Greece for 16-year-olds. They found an overall positive effect on performance of the number of previous exams, which may be due to a "practice" effect of taking more exams. In contrast, the number of days since the first exam was negatively related to performance, suggesting a possible fatigue effect (although this was limited to STEM[1] subjects). They also found a small negative effect of the number of days since the previous exam (for non-STEM subjects only).

Pope and Fillmore (2015) investigated how the number of days between Advanced Placement exams in the US affected performance. They found that students performed significantly better on the second exam when there was a longer gap. This was particularly the case for female and Asian students.

Thus, the limited prior research evidence suggests that there are several different ways in which exam scheduling might affect performance. In this research, we investigated three different mechanisms (as described in Goulas and Megalokonomou, 2018) through which this impact may be felt:

1. The number of days since the previous exam: if there is only a short gap between one exam and the next this leaves little time for revising for the second exam, which may impact performance.

1 Science, Technology, Engineering, and Maths.

2. The number of days since the first exam: towards the end of the exam series, students may experience exam fatigue, which could lead to lower performance.
3. The number of exams already taken: this may have a positive or negative effect on performance. There may be a practice effect whereby students who have taken more exams may perform better. On the other hand, students may experience exam fatigue if they have already taken many exams.

The main purpose of this research was to investigate whether exam scheduling impacts performance on A Levels through these mechanisms.

Data and methods

We used data from the Cambridge OCR awarding organisation (AO).[2] This is one of four AOs which offer A Levels in England. We restricted the data to the results of students who took at least three A Level subjects with OCR. This was because we were interested in exploring the impact of exam scheduling on those most likely to be affected (i.e., those taking many exams). As most students take three A Levels, including those taking fewer than three with OCR would include many students taking at least one A Level with another board (the results of which we did not have access to).[3]

We combined data from several different exam series (2016 to 2019 and 2023) into one dataset. Exams did not take place in 2020 or 2021 due to the COVID-19 pandemic and 2022 was not included as there were some adjustments made in that year (e.g., providing some advanced information on content included in exams) which made it unusual. In the earlier years (2016 to 2018), A Levels were in the process of being reformed. Therefore, the data in these years included some reformed subjects and some yet to be reformed. For the subjects which had not yet been reformed, exams could be taken in a modular fashion (i.e., at different time points throughout the two-year course). For consistency with later years, we only included students from these earlier years who took all their exams in the final exam series.[4]

The data included a variable indicating the date of each exam. For each student, the date of their first exam was recorded, and then for each subsequent exam the difference between the exam date and the first exam date was calculated. This was the "days since first exam" measure. The "days since previous exam" was calculated by simply subtracting the previous exam date from the current exam

2 This data was collected as part of the usual marking and processing of students' examination scripts and has been stored and used in line with Cambridge University Press & Assessment's Data Privacy notice (https://www.cambridge.org/legal/candidate-privacy-notice).

3 Ideally, we would have used data from all AOs, so that we could be certain that we had data on all exams taken by each student. However, the all-AO data available (e.g., National Pupil Database or GRADE data provided by the Department for Education) only includes grades or marks for the overall qualification. For our analysis we needed marks on individual components.

4 This was mainly students who took only exams in reformed subjects, but also included some students who took exams in pre-reform subjects, as long as they took all their exams in the final exam series.

date. Finally, a running total of the number of previous exams was calculated. An example of the data for one candidate is shown in Table 1. This shows one line per exam, ordered by exam date. This candidate's first exam was on 16 May. For this exam, the "days since previous exam" measure was set to Missing because this was their first exam. The "days since first exam" was 0. The second exam (23 May) was 7 days after the previous exam and 7 days after their first exam. They had taken one exam previous to this. Their final exam was on 21 June. This was one day after their previous exam, 36 days after their first exam, and prior to this they had taken 8 exams. The final column shows the marks achieved on each exam (as a percentage of maximum marks on the paper).

Table 1: Example student data

Subject	Component	Exam date	AM or PM	Days since previous exam	Days since first exam	No. of previous exams	Mark (as a % of max. marks)
Classical Civilisation	01	16/05	PM	.	0	0	84.0
Classical Civilisation	02	23/05	PM	7	7	1	82.7
Classical Civilisation	03	06/06	PM	14	21	2	88.0
Maths	01	06/06	PM	0	21	3	55.0
Biology	01	07/06	AM	1	22	4	52.0
Maths	02	13/06	PM	6	28	5	58.0
Biology	02	16/06	AM	3	31	6	50.0
Maths	03	20/06	PM	4	35	7	49.3
Biology	03	21/06	AM	1	36	8	69.6

This student also had two exams scheduled at the same time (PM on 6 June). The procedure when this happens is for the student to take one of these exams at the correct time and then the other exam in the other session on the same day (while being kept apart from other students).[5] As it was not possible to know which of these exams was sat first, we did not know the correct values for the "days since previous exam" or the "number of previous exams" variables. The way we dealt with this issue was different for the descriptive analysis and for the regression analysis. Which exam came first was not important for the descriptive analysis, so we chose one of the exams at random to be categorised as the first exam on the day (as shown in Table 1). However, for the regression analysis, we needed to know the order of the exams, and therefore we excluded all instances of exams scheduled at the same time.[6]

Descriptive analysis

The first part of the analysis was descriptive, focusing on the exam load on students. This includes the number of exams taken, the total time spent on exams, and the average amount of time between exams. Examples of common schedules and the most compressed schedules are also included. This analysis was undertaken at the candidate level (i.e., one row per candidate).

5 See page 24 of https://www.jcq.org.uk/wp-content/uploads/sites/2/2025/10/Print-version-JCQ-Instructions-for-conducting-examinations-2025_6_FINAL.pdf

6 The number of exams excluded was 5 433, approximately 5 per cent of the total.

Regression analysis

For the main analysis, that is, to investigate whether performance is affected by exam scheduling, regression models were fitted, predicting candidate performance (on individual exams) from the measures of exam scheduling and several contextual variables. This analysis was undertaken at the candidate exam level (i.e., one row per exam, per candidate).

The regressions fitted were multilevel models as these account for the hierarchical nature of the data. Specifically, three-level models were fitted, with results nested in centres, and centres nested in components. The clustering within centres accounted for the fact that results within a centre tended to be more similar than results between centres. The clustering within components accounted for the fact that centre results within a component tended to be more similar than results between components.[7]

The general form of the model was as follows:

$$y_{ijk} = \beta_0 + \beta_1 X_{ijk} + \beta_2 W_{jk} + \beta_3 Z_k + \nu_{0k} + \mu_{0jk} + e_{0ijk}$$

where y_{ijk} is the mark percentage for exam *i* in centre *j* and component *k*, X_{ijk} is the set of independent variables at exam level, W_{jk} is the set of independent variables at centre level, Z_k is the set of independent variables at component level, β_1 to β_3 are the sets of regression coefficients, v_{0k} is a component random effect, μ_{0jk} is a random effect of centre within component, and e_{0ijk} is the residual difference between actual and predicted mark.

Performance was measured by the mark achieved on the paper as a percentage of maximum mark. This meant that performance on components of different length (in terms of maximum mark) could be directly compared.

The independent variables of interest were the measures of exam scheduling. These were as follows:

- Days since previous exam:[8] This was converted from a continuous to a categorical variable. This was to allow for the effect of this measure in the regressions to be non-linear. We expect that the longer the time between exams, the better the performance is likely to be, because students will have more time to recover and to revise for the next exam. However, this benefit may tend to tail off as the number of days increases and this effect would not be captured by a continuous variable. The following categories were used: 0, 1, 2, 3, 4, 5, 6, 7, 8 or more, First exam. The "First exam" category was added because we wanted to include the first exam for each student in the analysis and without this category the "days since previous exam" measure would not have been applicable for these exams.

7 The Intraclass Correlation Coefficients (which measure the proportion of the total variance which is attributed to differences between clusters) were 0.20 for clustering within centres and 0.21 for clustering within components. These figures indicate that clustering was present and therefore a multilevel model was justified.

8 This measure takes no account of whether the exams were in the morning or the afternoon, so the true amount of time may be less or more than the full days stated. For example, if the number of days was equal to 1, this could actually be somewhat less than 24 hours if the previous exam was in the afternoon and the current exam in the morning.

- Days since first exam: A continuous measure, ranging from 0 to 45.
- Number of previous exams taken: As above, this variable was converted to a categorical variable, to allow for non-linear effects. Increasing the number of exams taken previously may be associated with improved performance to begin with (practice effect) but worse performance towards the end of the exams (fatigue effect). The categories were: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or more.

It was not possible to include all exam scheduling variables in the same model, due to a high level of correlation (0.79) between the "days since first exam" variable and the "number of previous exams" variable. Therefore, two separate sets of models were fitted for different combinations of the exam scheduling variables, as follows:

1. "Days since first exam" and "days since previous exam".
2. "Number of previous exams" only.

The contextual variables in the models were as follows:

- **Exam series** – June 2016, 2017, 2018, 2019 and 2023.
- **Candidate gender** – Female or male, as recorded in the database.
- **Candidate concurrent attainment** – As measured by the Instant Summary of Achievement Without Grades (ISAWG) and recorded in the database. This measure (as described in Benton, 2017) is a summary, on a single scale, of the achievement of students taking OCR qualifications, no matter which qualifications were taken. A higher score indicates higher levels of attainment. Every student taking at least one OCR component will have an ISAWG score. This is a series-specific measure and therefore ISAWG scores cannot be directly compared across different series. As we were combining data from several different series, we needed to account for this in the models. To do this, we included an interaction term between the ISAWG and the series.
- **Centre type** – Centres were classified into one of: Comprehensive School, Independent School, Further Education (FE) College, Sixth Form College, Selective School, Other.
- **Subject group** – Each component was allocated to one of the following subject groupings: Creative, English, Humanities / Social Sciences, Information Technology, Maths, Professional, Sciences, Other. A list of which subjects were included in each subject group is shown in the Appendix.

For the regression analysis, continuous variables were centred around either their overall median value (days since first exam, days since previous exam, and number of previous exams) or their overall mean value (ISAWG).

Results

Descriptive statistics on the exam load

In this analysis, we examined the exam load for students taking A Levels. Table 2 presents descriptive statistics on the number of exams taken by each student in the OCR data, overall and broken down by exam year.

Note that there were far fewer students in the years 2016 to 2018 because we only included students taking all of their A Level exams at the end of the course.

Table 2: Statistics on the number of A Level exams taken

Year	Students	Mean	Median	Standard deviation	Minimum	Maximum
2016	127	11.9	12	3.0	6	18
2017	824	9.7	9	1.9	6	18
2018	1 424	9.5	9	1.8	6	20
2019	3 943	9.2	9	1.1	6	18
2023	4 429	9.4	9	1.3	6	18
All	10 747	9.4	9	1.4	6	20

The overall mean number of exams taken by students was 9.4, and the median was 9. Most students in the data took three A Levels and most A Levels consist of three examined components. The mean number of exams decreased between 2016 and 2019 but marginally increased in 2023. The maximum number of exams taken by a student was 18 in all years apart from 2018 where it was 20.[9] The minimum number was 6 in every year.[10]

Table 3 shows descriptive statistics on the amount of time (in hours) that students spent on their A Level exams, by exam series and overall.

Table 3: Statistics on the total amount of time spent on A Level exams

Year	Students	Mean	Median	Standard deviation	Minimum	Maximum
2016	127	19.11	19.50	4.16	9.25	28.50
2017	824	18.10	17.25	2.52	10.00	29.50
2018	1 424	17.78	17.25	2.67	11.00	32.00
2019	3 943	17.33	17.25	1.72	11.50	33.00
2023	4 429	17.68	17.50	2.04	12.00	31.75
All	10 747	17.61	17.25	2.12	9.25	33.00

On average, students spent over 17 hours on A Level exams. The shortest amount of time was just over 9 hours and the longest was 33 hours. The average amount of time spent decreased between 2016 and 2019 but slightly increased in 2023. This follows the same pattern as the average number of exams over time (see Table 2).

Table 4 and Figure 1 present statistics on the number of days between consecutive exams and Table 5 and Figure 2 show statistics on the number of days between the first and last exams for each student. These tables and figures give an indication of how compressed exam schedules tended to be and how long they tended to last.

9 These were students taking four or five A Levels with OCR. In each case, one of the subjects was Maths which allowed students to take multiple exams (more than the minimum necessary to obtain the qualification) and use their highest marks to contribute to the final grade.

10 These were students taking three A Levels, all of which had only two examined components.

Table 4: Statistics on the number of days between consecutive exams

Year	Students	Mean	Standard deviation	Minimum	Maximum
2016	127	3.32	0.89	1.60	6.20
2017	824	3.06	1.12	1.88	7.80
2018	1 424	2.35	0.74	1.25	6.17
2019	3 943	3.05	0.85	1.22	5.20
2023	4 429	3.41	0.83	1.70	5.86
All	10 747	3.11	0.92	1.22	7.80

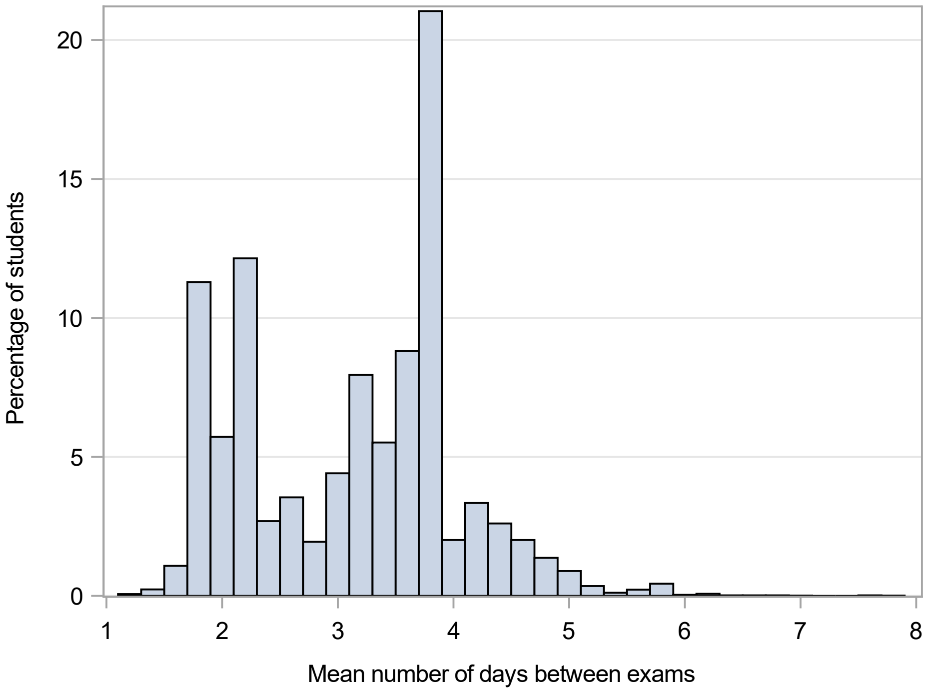

Figure 1: Distribution of mean number of days between consecutive exams

Table 5: Statistics on number of days between first and last exam

Year	Students	Mean	Standard deviation	Minimum	Maximum
2016	127	34.8	8.0	8.0	42.0
2017	824	26.6	11.3	13.0	45.0
2018	1 424	20.5	9.4	10.0	43.0
2019	3 943	24.8	6.6	11.0	42.0
2023	4 429	28.1	5.9	14.0	42.0
All	10 747	25.8	7.7	8.0	45.0

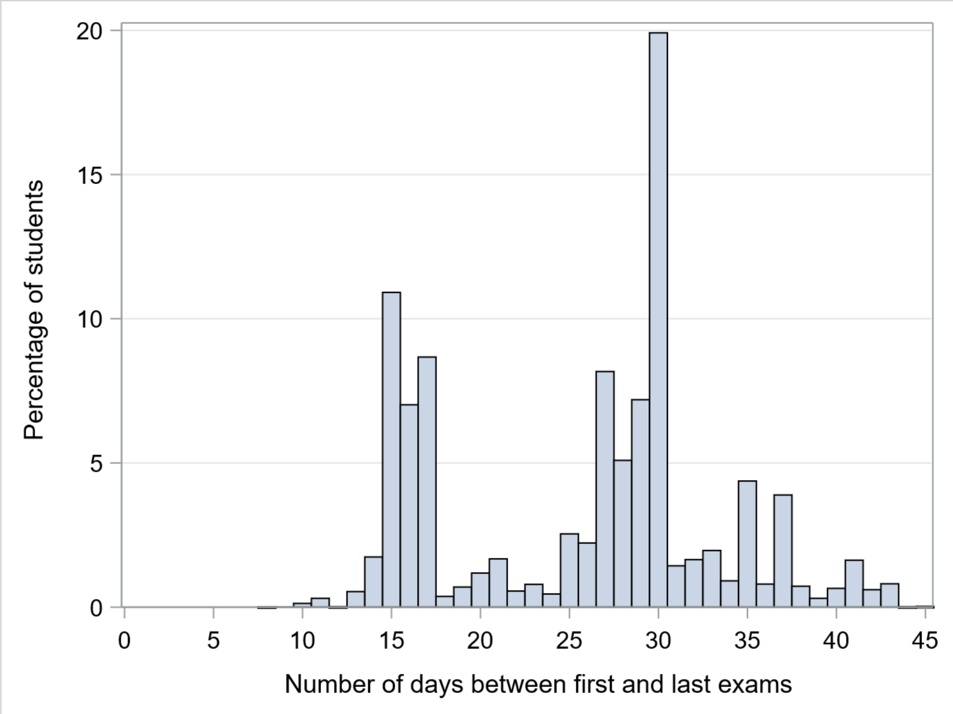

Figure 2: Distribution of number of days between first and last exams

On average, students had around three days between consecutive exams. There were some differences between years, with a mean of 2.35 in 2018 and a mean of 3.41 in 2023. The mean number of days between first and last exam was just below 26 days. Again, the lowest mean was for 2018 (20.5 days), with a considerably higher mean in 2023 (28.1 days). This suggests that students' exam schedules were much more compressed in 2018 than in 2023. This was at least partly due to changes implemented following the COVID-19 pandemic, whereby exams from the same qualification were scheduled further apart so that students had more time for revision (Department for Education, 2023).

The overall distributions (see Figures 1 and 2) for both these measures show two separate peaks, at around two days between consecutive exams (15 to 17 days between first and last exam) and at around 3.8 days between consecutive exams (30 days between first and last exam). This was mainly due to the half-term holiday in England, when schools were closed and no exams were taken. A significant minority of students had no exams before half-term and therefore only had around 15 to 17 days between the end of the half-term holiday and the end of the exam period (i.e., around two days between exams).

From the minimum values in Tables 4 and 5, it is clear that there were some students with very compressed schedules. We looked into these schedules in more detail. Table 6 shows the five most compressed schedules in terms of mean number of days between exams. This table shows the year in which the schedule took place, the number of students with that specific schedule, the mean number of days between consecutive exams ("Mean days"), the number of exams, the number of days between first and last exam ("Total number of days"), the total exam time and the subjects taken. For example, Table 6 shows that there were three students who took 10 exams (total time of over 19 hours) in 11 days in 2019 (an average gap of 1.22 days between exams). These students took Latin, Maths, and Further Maths. There were also three students who took 13 exams in 15 days, with a total time of 25 hours.

Table 6: Five most compressed exam schedules (lowest mean days between exams)

Year	Students	Mean days	Number of exams	Total number of days	Total exam time	Subjects
2019	3	**1.22**	10	11	19 hrs 10 mins	Latin, Maths, Further Maths
2018	1	**1.25**	9	10	16 hrs 30 mins	Geography, Psychology, English Language and Literature
2019	3	**1.25**	13	15	25 hrs	Chemistry, Biology, Maths, Latin
2019	2	**1.31**	14	17	25 hrs 30 mins	Chemistry, Biology, Maths, Further Maths
2019	4	**1.33**	13	16	24 hrs	Chemistry, Biology, Maths, Further Maths
2018	1	**1.33**	13	16	25 hrs	Chemistry, Biology, Physics, Latin

We also looked at the most compressed schedule in terms of the least number of days to complete all exams. Using this measure, there were five students whose exam schedule was 10 days, in which they took either 9 or 8 exams.

Although the examples in Table 6 revealed some very compressed schedules, these were only for very few students. To put this in more context, we also looked at the most common exam schedule in each year. In all years apart from 2016, the most common schedule was very similar, requiring 9 exams (of around 17 hours total time) in between 15 and 17 days. The subject combinations taken by these students were either Biology, Chemistry, and Maths, or Biology, Chemistry, and Physics.

Regression results

Table 7 presents the results of the first set of regression models, which included the "days since first exam" and "days since previous exam" variables as predictors of performance.

In model 1 these two exam scheduling variables were included on their own and in model 2 contextual variables were added. Model 3 adds the statistically significant interactions between the exam scheduling variables and the contextual variables.

The results of model 1 show a statistically significant but very small negative effect (-0.10) of the number of days since the first exam. However, after including the contextual variables[11] (model 2), this effect disappears completely.

For the "days since previous exam" variable there were significant and positive effects in model 1 for all categories apart from 7 days and 8 or more days. In model 2, all categories were significant and positive. The reference category for this variable was 0 days, so these positive effects indicate that students performed better in exams taken at least a day after their previous exam than they did when the exam was on the same day as the previous exam.

11 Note that centre type did not have a statistically significant effect, so was removed from the models.

However, the size of the effect was very small in each case (between 1.65 and 3.51 per cent of maximum marks). To illustrate the size of the effect, Figure 3 shows the predicted mark percentages for the different levels of this variable, for a typical student.[12]

Table 7: Regression results (models including "days since first exam", and "days since previous exam" variables)

Fixed effect		Model 1 (n=95 129)	Model 2 (n=95 094)	Model 3 (n=95 094)
Intercept		61.63 (0.68)*	65.80 (2.17)*	66.11 (2.15)*
Days since first exam		-0.10 (0.01)*	0.00 (0.01)	0.01 (0.01)
Days since previous exam	0			
	1	2.29 (0.36)*	1.93 (0.22)*	1.38 (0.31)*
	2	1.25 (0.38)*	2.35 (0.23)*	2.30 (0.33)*
	3	2.78 (0.39)*	2.01 (0.24)*	1.67 (0.33)*
	4	3.20 (0.40)*	3.51 (0.24)*	2.68 (0.35)*
	5	1.87 (0.43)*	2.60 (0.26)*	2.71 (0.37)*
	6	3.74 (0.50)*	2.72 (0.30)*	2.88 (0.44)*
	7	-0.04 (0.70)	1.65 (0.42)*	1.61 (0.55)*
	8 or more	0.80 (0.41)	2.74 (0.25)*	2.13 (0.36)*
	First exam	1.97 (0.44)*	2.92 (0.27)*	2.36 (0.36)*
Gender	Female			
	Male		0.50 (0.08)*	-0.35 (0.37)
ISAWG			20.16 (0.43)*	17.70 (0.51)*
Subject group	Creative			
	English		7.50 (3.60)*	7.37 (3.56)*
	Humanities / Social Sciences		-0.30 (2.32)	-0.28 (2.29)
	Information Technology		-11.55 (4.15)*	-11.74 (4.09)*
	Maths		-14.23 (2.43)*	-14.06 (2.40)*
	Other		0.32 (2.60)	0.29 (2.56)
	Professional		2.40 (2.94)	2.07 (2.90)
	Sciences		-11.86 (2.46)*	-11.81 (2.43)*
Series	2016			
	2017		-1.71 (0.50)*	-1.80 (0.50)*
	2018		3.37 (0.51)*	3.34 (0.50)*
	2019		15.45 (0.51)*	15.41 (0.51)*
	2023		2.59 (0.51)*	2.47 (0.51)*
ISAWG*Series	2016			
	2017		2.47 (0.57)*	2.44 (0.46)*
	2018		2.06 (0.45)*	2.12 (0.45)*
	2019		4.87 (0.44)*	5.13 (0.44)*
	2023		6.82 (0.44)*	6.98 (0.44)*

12 For this exemplification and for others that follow, a "typical" student is one in the reference category for all applicable categorical variables (e.g., female, 2016 series, creative subject group) and with a value equal to the mean for all applicable continuous variables.

Fixed effect		Model 1 (n=95 129)	Model 2 (n=95 094)	Model 3 (n=95 094)
Days since previous exam*Gender	0			
	1			1.05 (0.39)*
	2			0.41 (0.41)
	3			0.75 (0.42)
	4			1.68 (0.44)*
	5			0.10 (0.47)
	6			-0.02 (0.55)
	7			-0.37 (0.68)
	8 or more			1.08 (0.45)*
	First exam			0.93 (0.43)*
Days since previous exam*ISAWG	0			
	1			3.81 (0.29)*
	2			2.95 (0.30)*
	3			2.78 (0.31)*
	4			2.83 (0.32)*
	5			2.74 (0.35)*
	6			1.79 (0.40)*
	7			-0.67 (0.52)
	8 or more			1.50 (0.33)*
	First exam			-1.91 (0.32)*

Note: Standard errors in parentheses. * $p < 0.05$.

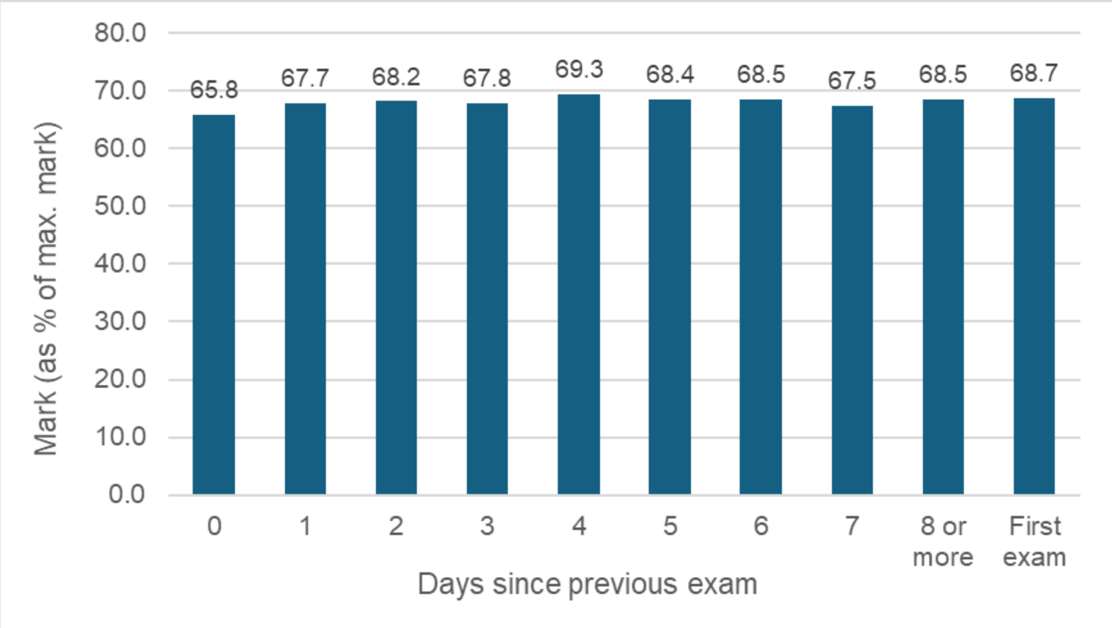

Figure 3: Predicted mark percentage for different levels of "days since previous exam" variable (model 2)

Figure 3 shows, for example, that when the days since the previous exam was 0 the predicted mark percentage was 65.8. With a break of 4 days the prediction was 69.3, an increase of 3.5 percentage points.

The results of model 3 (Table 7) show two significant interactions with the "days since previous exam" variable, for gender and for ISAWG score.

For the interaction with gender there were significant positive effects for 1 day, 4 days, 8 or more days, and First exam. This means that the positive effect of these numbers of days (compared with 0 days) was larger for males than females.

Figure 4 illustrates this effect for a "typical" student, showing the predicted mark percentage for males and females separately. The effects, however, were very small. For example, the largest difference compared to 0 days between exams for males was 4.4 percentage points for a gap of 4 days. For the same time between exams for females the difference was 2.7 percentage points.

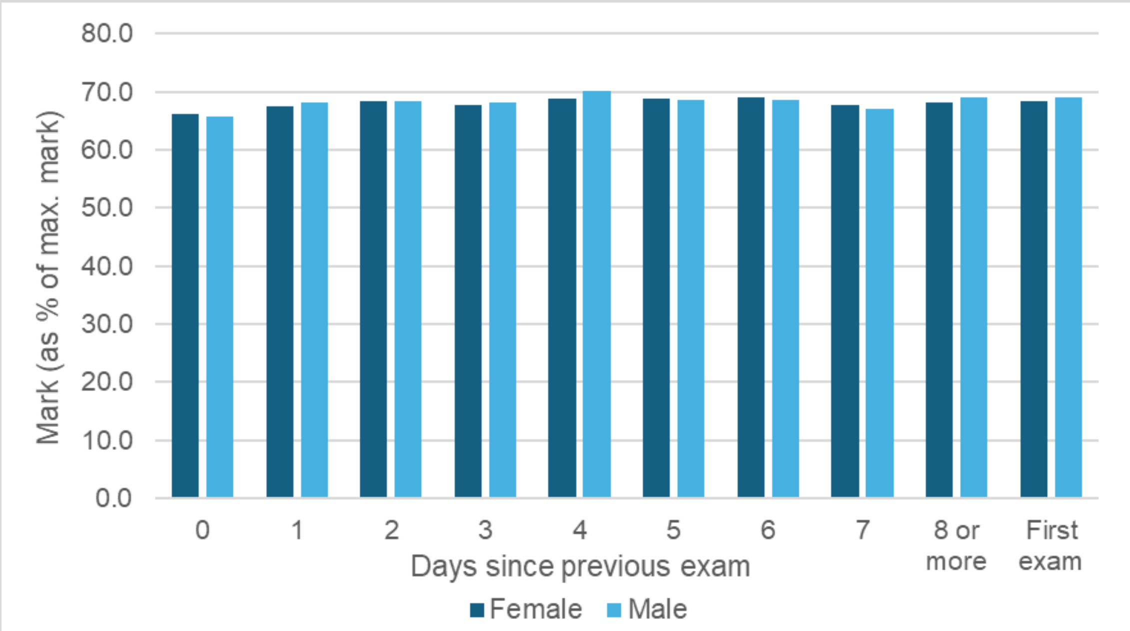

Figure 4: Predicted mark percentage by gender and "days since previous exam" variable (model 3)

The interaction effect with ISAWG score was positive (and significant) for all categories apart from 7 days (negative, but not significant) and First exam (negative and significant). This means that the positive effect (that is, better performance) of more days since the previous exam (compared with 0 days) was greater for those with higher ISAWG scores. However, the largest effect was for 1 day and for any more days than this the effects (although still positive) were lower. The interaction is illustrated in Figure 5, which shows the predicted mark percentage (for a typical student) for different numbers of days since the previous exam across a range of ISAWG scores. This figure shows that, for example, the largest difference in predicted mark percentage compared to 0 days for a student with an ISAWG score of 1.75 was for 4 days (5.4 percentage points). In contrast, for students with an ISAWG score of -0.25 the difference for 4 days was negative (-0.3 percentage points).

Table 8 presents the results of the second set of models, which included the "number of previous exams" variable. In model 4 the exam scheduling variable was included on its own, model 5 added contextual variables,[13] and model 6 added the statistically significant interactions.

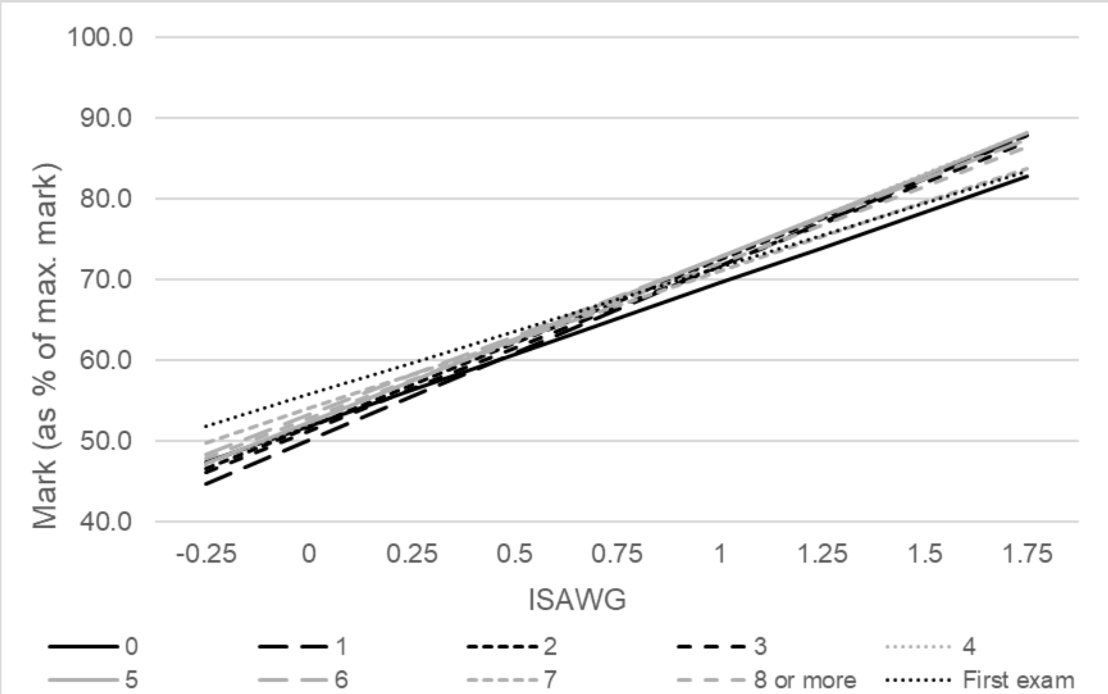

Figure 5: Predicted mark percentage for different levels of ISAWG and "days since previous exam" variable (model 3)

In models 4 and 5, there were small but significant negative effects for most categories of the "number of previous exams" variable. The results of model 5 indicate that students performed worse for each number of previous exams (apart from 2) than they did when the number was 0 (i.e., their first exam). Figure 6 shows the predicted mark percentages for the different categories of this variable, for a typical student. This shows that the differences in predicted performance were all very small, with a maximum difference of 2.5 marks (for 6 previous exams).

In model 6, the significant interactions between the number of previous exams and the contextual variables are included. There was only one significant interaction, with the ISAWG. This was positive for all numbers of previous exams and (mostly) increased with the number of previous exams taken. This means that the negative effect of more previous exams was lower for students of higher ability. Figure 7 illustrates this effect by comparing the mark percentage (for a typical student) for different numbers of previous exams and across a range of ISAWG scores. This shows that for high ability students there was a substantial positive effect of more previous exams, while for low ability students the effect was negative. The largest effect was for 9 or more previous exams. Students with an ISAWG score of 1.75 were predicted a mark percentage 7.1 percentage points higher for 9 or more previous exams compared with 0 previous exams. For students with an ISAWG score of -0.25 the predicted mark percentage for 9 or more previous exams was 11.2 percentage points lower than for 0 previous exams.

13 Note that centre type did not have a statistically significant effect, so was removed from the models.

Table 8: Regression results (models including "number of previous exams" variable)

Fixed effect		Model 4 (n=95 129)	Model 5 (n=95 094)	Model 6 (n=95 094)
Intercept		63.67 (0.68)*	68.82 (2.18)*	68.68 (2.13)*
Number of previous exams	0			
	1	-1.66 (0.31)*	-1.09 (0.19)*	-0.95 (0.19)*
	2	-0.19 (0.32)	0.25 (0.19)	0.40 (0.19)*
	3	-1.19 (0.34)*	-1.28 (0.21)*	-0.99 (0.21)*
	4	-1.88 (0.36)*	-2.04 (0.22)*	-1.83 (0.22)*
	5	-1.35 (0.37)*	-1.45 (0.23)*	-1.20 (0.23)*
	6	-2.31 (0.39)*	-2.52 (0.24)*	-2.27 (0.24)*
	7	-0.00 (0.42)	-1.34 (0.26)*	-1.30 (0.26)*
	8	0.70 (0.45)	-1.37 (0.28)*	-1.67 (0.28)*
	9 or more	5.55 (0.48)*	-0.03 (0.30)	-1.54 (0.31)*
Gender	Female			
	Male		0.51 (0.08)*	0.50 (0.07)*
ISAWG			20.14 (0.43)*	14.85 (0.46)*
Subject group	Creative			
	English		7.26 (3.57)	7.34 (3.54)*
	Humanities / Social Sciences		-0.48 (2.30)	-0.05 (2.28)
	Information Technology		-11.19 (4.11)	-11.01 (4.07)*
	Maths		-14.51 (2.41)	-13.92 (2.39)*
	Other		0.27 (2.57)	0.49 (2.55)
	Professional		2.52 (2.91)	2.87 (2.89)
	Sciences		-11.83 (2.44)	-11.32 (2.42)*
Series	2016			
	2017		-1.79 (0.50)*	-1.84 (0.50)*
	2018		3.26 (0.51)*	3.24 (0.50)*
	2019		15.40 (0.51)*	15.38 (0.50)*
	2023		2.53 (0.51)*	2.42 (0.50)*
ISAWG*Series	2016			
	2017		2.48 (0.47)*	3.20 (0.46)*
	2018		2.06 (0.45)*	2.92 (0.45)*
	2019		4.84 (0.44)*	6.06 (0.44)*
	2023		6.77 (0.44)*	7.71 (0.44)*
Number of previous exams*ISAWG	0			
	1			3.54 (0.22)*
	2			3.20 (0.22)*
	3			4.60 (0.22)*
	4			3.79 (0.22)*
	5			4.40 (0.22)*
	6			4.89 (0.22)*
	7			4.44 (0.22)*
	8			9.22 (0.24)*
	9 or more			9.15 (0.31)*

Note: Standard errors in parentheses. * $p < 0.05$.

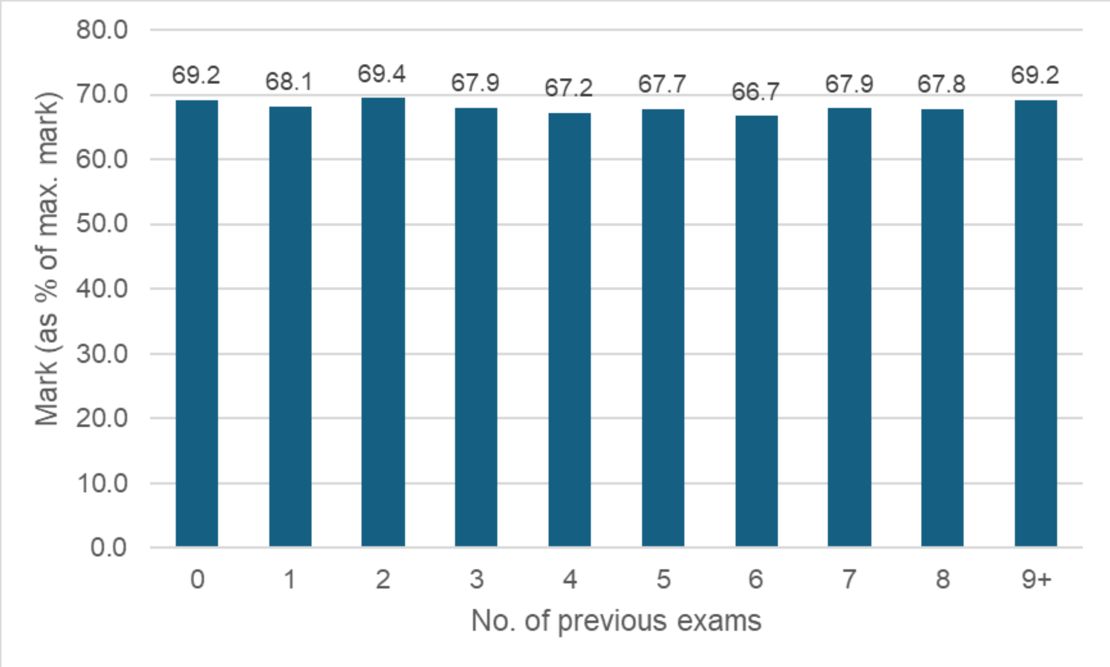

Figure 6: Predicted mark percentage for different levels of "number of previous exams" variable (model 5)

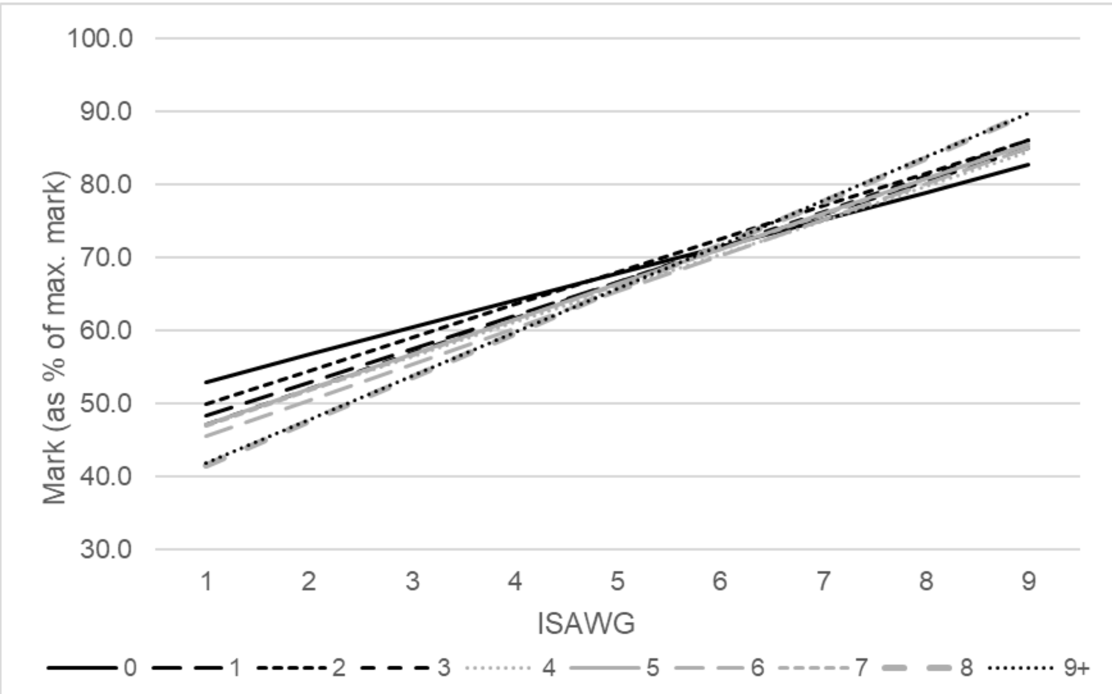

Figure 7: Predicted mark percentage for different levels of ISAWG and "number of previous exams" variable (model 6)

Discussion

This research investigated the current exam load on A Level students in England and the impact of exam scheduling on performance. The focus was on students taking OCR A Levels only.

In terms of the exam load, the median number of exams taken was 9 and the average total time was around 17.5 hours. The average number of days between exams was 3.1, but a substantial proportion of students had an average of around 2 days.

In terms of the impact of exam scheduling on performance, the results of the analyses showed no significant effect on performance of the number of days since the first exam (after accounting for gender, concurrent attainment, exam series, and subject). This means that there was no evidence of a fatigue effect, whereby students get tired or bored towards the end of the exam period. This finding contrasts with previous research (Goulas & Megalokonomou, 2018) which found a small negative effect for this variable, but for exams in STEM subjects only.

There was a significant and positive effect on performance of the number of days since the previous exam. However, this was essentially limited to any number of days being better than 0. There was no evidence that cumulatively more days was associated with higher performance. This suggests that students were less likely to do well on the second (or third) exam taken on the same day, compared with longer gaps between exams. Overall, this effect was very small (between 2 and 3.5 mark percentage points).

However, the effect was different for some groups of students. Firstly, it was slightly larger for male students than for females. Secondly, for students of higher ability there was a clear effect of improved performance for a gap of 1 day or more, but this was not the case for lower ability students. This finding is similar to that from previous research (Goulas & Megalokonomou, 2018) which found a positive effect of an increase in the number of days between exams for high ability students (but only on exams in non-STEM subjects). Their suggested reason for this was that high ability students may be better than low ability students at cramming for exams and therefore additional days between exams may be beneficial for these students.

For the final measure of exam scheduling (number of previous exams) there was a significant negative effect for most numbers of prior exams compared with it being the first exam. However, there was no evidence of larger effects for larger numbers of prior exams. Furthermore, the significant effects were all very small.

However, there was clear evidence of a differential effect of the number of previous exams on performance for students of different ability. For high ability students the more previous exams taken the better the performance. For low ability students the reverse was true. This effect was larger than the main effect, amounting to a 7-percentage point difference for students with an ISAWG score of 1.75 with 9 or more previous exams compared with their first exam. This effect may be due to different levels of motivation for low and high attainers. Low attainers may tend to be less motivated and therefore may be more likely to suffer exam fatigue in later exams. In contrast, high attainers are usually more motivated and therefore more able to benefit from practice effects of taking more exams. Practice can potentially help with different aspects of exam taking technique such as time management, stress management and preparation strategies.

This finding was the opposite of previous research (Goulas & Megalokonomou, 2018), which found that, while the overall effect of the number of previous exams was positive, for higher ability students it was significant and negative. They argue that high ability students have stronger meta-cognitive skills and may therefore experience lower returns to practice. However, it is worth noting that the context of the previous study was different, involving students in 11th grade taking exams in an average of 13 different subjects in Greece.

In this research, we found no significant interaction effects between the measures of exam scheduling and subject group. This is in contrast to previous research (Goulas & Megalokonomou, 2018), which found differential effects depending on whether the exam was in a STEM or non-STEM subject. However, this could be because in this analysis, we used a more detailed breakdown of subjects, and the context was different in the previous study.

The overall findings from this research suggest that there is not a strong case for making substantial changes to the way that exams are currently scheduled. Nonetheless, a minor change which could be considered would be to try to minimise the number of occasions when candidates have two (or even three) exams on the same day, as candidates are slightly disadvantaged by this. This, however, may not be possible given the constraints involved in trying to schedule a large number of exams into a relatively short period of time.

A limitation of this research was the fact that we only had data from OCR. Inevitably, there would have been some students in our data taking three or more A Levels with OCR who also took one or more A Levels with a different board. For these students we did not know when they took their non-OCR exams, and therefore some of their exam scheduling data will have been inaccurate. We do not know how many students this applies to, but we can assume it is a small minority. For example, only 4.7 per cent of students took more than three A Levels in 2023.[14]

Another caveat is that by restricting to OCR exams only we do not know how representative this is of the population of all A Level students. The most common combinations of subjects in our data might be different from those for A Level students as a whole. However, according to Ofqual (https://analytics.ofqual.gov.uk/apps/Alevel/SubjectCombinations/), the most common combination in 2019 and 2023 was the same as we found (Biology, Chemistry, Maths),[15] which is a positive sign that our dataset shares characteristics with the wider cohort.

A final limitation relates to the specification of the multilevel model. We chose to nest results within centres and centres within components. An alternative

14 See https://www.gov.uk/government/publications/infographic-a-level-results-2023/infographics-for-a-level-results-2023

15 We could only compare subject combinations between our dataset and the wider A Level cohort for 2019 and 2023 due to the data available.

specification would have been to nest results within candidates, on the basis that results were more likely to be similar within candidates than between candidates. Not doing this clustering within candidates may have led to reductions in standard errors and hence increased the chance of variables being found to be significant. Further research could explore the possibility of using more complex multilevel models, which also account for the clustering within candidates.

Further research could also investigate in more detail the differential effects of exam scheduling on low and high achievers, to explore, in particular, why high achievers benefit from a practice effect while low achievers do not.

It would also be interesting to investigate the impact of the COVID-19 changes to scheduling, with exams from the same qualification being more spaced apart. In particular, did this have any effect on performance in the later exams because, for instance, the longer gaps meant candidates had more time to revise, or alternatively, candidates forgot what they revised for the previous exams?

Finally, further research could look at absence rates and whether they are higher towards the end of the exam period. This is another way in which exam scheduling could impact on results, but which was not captured by the analysis presented here.

References

Benton, T. (2017, November 9–11). *Pooling the totality of our data resources to maintain standards in the face of changing cohorts* [Conference presentation]. 18th annual AEA-Europe conference, Prague, Czech Republic.

Department for Education (2023, April 28). Exams in 2023 – everything you need to know. *The Education Hub*.

Goulas, S., & Megalokonomou, R. (2018). *Marathon, hurdling or sprint? The effects of exam scheduling on academic performance*. *IZA DP No. 11624*. Institute of Labor Economics.

Pope, D. G., & Fillmore, I. (2015). The impact of time between cognitive tasks on performance: Evidence from advanced placement exams. *Economics of Education Review, 48,* 30–40.

Appendix: List of subjects, by subject group

Subject group	Subjects
Creative	Design and Technology, Drama and Theatre, Film Studies, Music, Media Studies
English	English Language, English Language and Literature, English Literature
Humanities / Social Sciences	Ancient History, Business, Classical Civilisation, Critical Thinking, Economics, General Studies, Geography, Government and Politics, History, Psychology, Religious Studies, Sociology
Information Technology	Computing, Computer Science, Information and Communications Technology
Maths	Further Mathematics A, Further Mathematics B (MEI),[16] Mathematics A, Mathematics B (MEI)
Professional	Accounting, Health and Social Care, Law, Travel and Tourism
Sciences	Biology A, Biology B (Advancing Biology), Chemistry A, Chemistry B (Salters), Geology, Physics A, Physics B (Advancing Physics)
Other	Classical Greek, Dutch, French, German, Gujarati, Latin, Persian, Physical Education, Portuguese, Spanish, Turkish

16 Mathematics in Education and Industry.