An investigation of the effect of early entry on overall GCSE performance, using a propensity score matching method

Tim Gill Research Division

Introduction

A report by Gill (2013) found that certain groups of students performed worse than expected in some GCSE subjects when they were taken early, even taking into account any improved performance from re-sitting. In particular, high attaining students (those achieving level 5 at Key Stage 2 [KS2] tests in the subject) were less likely to achieve a grade A in GCSE English or GCSE Mathematics if they took the exam early (even if they re-sat at the expected time). However, it may be that one reason for taking an exam early is to 'get it out of the way' to enable increased focus on other subjects in Year 11. An Ofsted survey (Ofsted, 2013) asked schools their reasons for entering students early and 44% responded that they did so "to allow students to focus on other subjects". Furthermore, schools were asked what they felt the benefits of early entry were given their experience, and 51% responded "the freed time allowed students to do better in other subjects".

If early entry leads to better than expected performance in the other exams then the overall impact of early entry may not be detrimental and could even be advantageous. This article investigates this issue by looking at whether students entering early for GCSEs perform better or worse across all their GCSEs (or equivalents) than those who do not enter for any GCSEs early.

Data and methods

The data for this analysis came from the National Pupil Database (NPD) for 2011. This is a database of student level attainment and personal characteristics compiled by the Department for Education from data supplied by centres and awarding bodies. The Key Stage 4 (KS4) extract, which records all attainment by students who are at the end of KS4, was used. The database includes exams taken by these students in previous years, meaning it was possible to identify early entry.

To compare the overall GCSE performance of early entry students with non-early entry students, three different outcome measures were used:

1. Mean GCSE score. This was calculated by transforming each GCSE grade into a number (A*=8, A=7 etc.) and then generating a mean value for each student. The grade used was the best grade attained in each subject (i.e. after re-sits).
2. Indicator of whether or not the student passed the statutory target of five or more GCSEs (or equivalents) at A* to C including English and Mathematics. This is an important accountability measure for schools, and is used in school performance tables. The outcome measure to compare between the different groups was therefore the percentage of students passing this threshold.
3. Total KS4 points score. For all KS4 qualifications a score is allocated to each grade (for example, an A* grade at GCSE is worth 58 points, an A grade 52 points and a B grade 46 points)[1]. The total points score is the sum of the points received on all qualifications taken by the candidate. This was included as an alternative to the GCSE mean score because it gives more value to a candidate with, for instance, nine A* grades than one with eight A* grades. This might be an important difference in particular circumstances (e.g. allowing more options at A level).

Early entry was defined as having taken the exam for the first time prior to starting Year 11. This means that students taking an exam for the first time in January of Year 11 were not considered to be early entry. Students taking the qualification early and then re-sitting in Year 11 were counted as early entry, despite the fact that some of them will not have had more time in Year 11 to focus on other subjects (as hypothesised). However, counting these students as not early entry would potentially have been more problematic because of the way in which they would have been 'allocated' to this group. Had we done this, it is likely that anyone who didn't achieve at least a C would probably be entered again, thus moving from an early entry group to a non-early entry group. In effect this could mean that the outcome measure (GCSE performance) determines which group students are in, which would invalidate the analysis. Students taking fewer than five GCSEs were excluded from the analysis.

Propensity score matching

A propensity score matching method was used in this research (see Caliendo & Kopeinig, 2008; Morgan & Harding, 2006). This method is useful when we have a 'treatment', and want to compare the outcomes for a 'treated' group with those of a 'non-treated group' but we are not able to randomly assign people to the groups. For this research, treatment refers to early entry in at least one GCSE, and the outcome refers to each of the three performance measures detailed above. In theory, to know for certain the effect of a treatment, we would need to compare the outcomes for the same participants with and without treatment at the same time. In practice this is not possible, so other methods are necessary. The treated and non-treated groups could just be compared in terms of their mean outcomes, but this would not be comparing like with like because of differences between the two groups in terms of background characteristics (covariates). The propensity score method attempts to overcome this by manipulating the data such that the treated and non-treated groups are made similar enough for comparisons between the groups to be valid.

There are a number of different ways of doing this, the most common of which is to 'match' each individual in the treated group with one (or more) individuals in the non-treated group in terms of covariates. However, this is a computationally demanding method when dealing with large data sets and so a different method was employed here, involving

1. For a full list of qualifications and points scores visit http://register.ofqual.gov.uk/Qualification, enter the qualification and click on "View performance measures".

the creation of subgroups in the treated and non-treated groups such that the members of each subgroup were very similar in the two different groups in terms of covariates. Weights were then used to compensate for the imbalance of treated and non-treated individuals in each subgroup. This method is now described in some detail.

First, it was necessary to identify individuals in each group who were similar in terms of covariates. To do this, individuals were classified by their 'propensity' for being in the treated group. A logistic regression model was run, with being in the treated group (i.e. early entry) as the dependent variable and all the covariates of interest as independent variables. The coefficients from this model allowed us to estimate the probability an individual with any particular set of background characteristics would be in the treated group. This probability is referred to as the propensity score. Groups of students with similar propensity scores are very likely to be similar in terms of their background characteristics.

Once the propensity score measure was calculated, individuals were classified into ten subgroups[2], based on their propensity score. Thus, subgroup 1 consisted of those with the lowest propensity score (lowest probability of being treated) and subgroup 10 those with the highest propensity score (highest probability of being treated). The equivalent subgroups in the treated and non-treated groups should now have been similar in terms of their background characteristics, enabling comparisons to be made. However, within each subgroup the balance of the number of treated and non-treated individuals was not even (particularly in groups 1 and 10), and thus it was also necessary to apply weights to the non-treated individuals to account for this imbalance.

Following the application of the weights, the distribution of covariates in the treated and non-treated group should have been approximately the same. This was checked to make sure that the weighting had worked correctly. Then, the outcome variable (weighted in the non-treated group) was compared in the treated and non-treated groups. The results for the weighted non-treated group could be thought of as the outcome for the treated group *had they not been treated*. Using the technical language, we were estimating the *average treatment effect for the treated* (ATT).

This method was applied in a number of different situations. The first of these had just one treated group (entering early for at least one GCSE). The same method was then applied to a situation with two treatments, either taking one GCSE early or two or more GCSEs early. For this analysis the principle was the same but the method was modified in two important ways. Firstly, a different method for generating propensity scores was used. This was necessary because using logistic regression with two treatment groups generated propensity scores that meant the groups were not well-matched on covariates. Instead, a Generalised Boosted Model (GBM) was used to generate the propensity scores. GBMs use an automated, data-adaptive algorithm to estimate a smooth function, by adding together a large number of simple functions (see McCaffrey, Ridgeway & Morral, 2004, for an example application to propensity score estimation). They are flexible because they allow the function being modelled to be non-linear, and generate propensity scores that are well-matched to the empirical probability of treatment.

The second modification was in the meaning of the propensity score, which now referred to a student's propensity for being in the *non-treated* group. Students were then classified by their propensity score into 15 subgroups and, in contrast to the single treatment situation, data from the *treated* groups was weighted to match the non-treated group in terms of the number of students in each subgroup. This means that the results for the weighted, treated groups can be thought of as the outcome for the non-treated group *had they been treated*. In the technical language, we are estimating the *average treatment effect for the non-treated* (ATNT).

Finally, both the single treatment and two treatment models described were applied to subgroups of students to see if there were different treatment effects within different groups. Students were classified by three variables; gender, school type and prior attainment. The methods described were then applied to each subgroup in turn.

Subgroup analyses

To investigate the effect of prior attainment, students were classified by their mean KS2 level across the three tests (English, Mathematics and Science) into three approximately equally sized groups. Normal KS2 levels range from 2 to 5 and students given either a level 'B' ('Working below the level assessed by the test') or 'N' ('No test level awarded') were allocated a level 1 so that these results could be included in the calculation of their mean KS2 level.

For the school type analysis, the schools that students attended were classified into four types – comprehensives (including academies and free schools), grammar schools, independent schools and secondary modern schools.

Covariates

For the logistic regression models only the covariates that had a statistically significant effect on the probability of being in the treated group were included. These were selected from the following list:

- Total number of GCSEs taken
- Number of other qualifications taken
- Number of BTECs taken
- Number of OCR Nationals taken
- Prior attainment (as measured by KS2 levels in English, Mathematics and Science)
- Deprivation measure (IDACI)
- Ethnicity
- Gender
- Age
- School type

These variables were chosen because they were available in the NPD, and were potentially influential in determining both a student's likelihood of entering early and the outcome measures.

When calculating the propensity scores, only variables with a statistically significant parameter estimate were included in the final logistic regression model. Furthermore, when analysing by subgroup, the relevant subgroup variable was excluded from the model. So, for instance, in the analysis of comprehensive students the school type variable was excluded because all students were in the same category. For other subgroups some variables were missing for all students, so these were excluded. For example, for the analysis of independent school students', ethnicity and Income Deprivation Affecting Children Index (IDACI) score were removed because these are not recorded in the NPD for these students.

2. According to Rosenbaum and Rubin (1983), subclassification into five subclasses is enough to remove 90% of the bias for many distributions, so ten subclasses should be more than sufficient.

Results

Data exploration

Table 1 presents a breakdown of the number of GCSEs entered for early by students in the 2011 cohort taking at least five GCSEs in total. Thus, almost 38% of students entered early for at least one GCSE, with most of those just entering one early (25%). Only a very small minority entered for three or more GCSEs early.

Table 1: Number of GCSEs entered early

Number of early entry	*Students (n)*	*Students (%)*
0	328,246	62.7
1	130,738	25.0
2	43,286	8.3
3	13,704	2.6
4	3,919	0.8
5+	3,635	0.7

Figures 1 and 2 present the distribution of mean GCSE for early entry and non-early entry students (and for different numbers of GCSEs entered early).

Overall, students entering at least one GCSE early had a higher mean GCSE (5.33) than those who did not enter any GCSEs early (5.16). Figure 3 shows that students who entered early for one, two or three GCSEs had the highest mean GCSE scores (5.34, 5.33 and 5.33 respectively).

However, this analysis takes no account of differences in the background characteristics of students in each category. If these characteristics have an impact on the variable of interest (GCSE mean grade) then it is important to account for any differences in them between the groups.

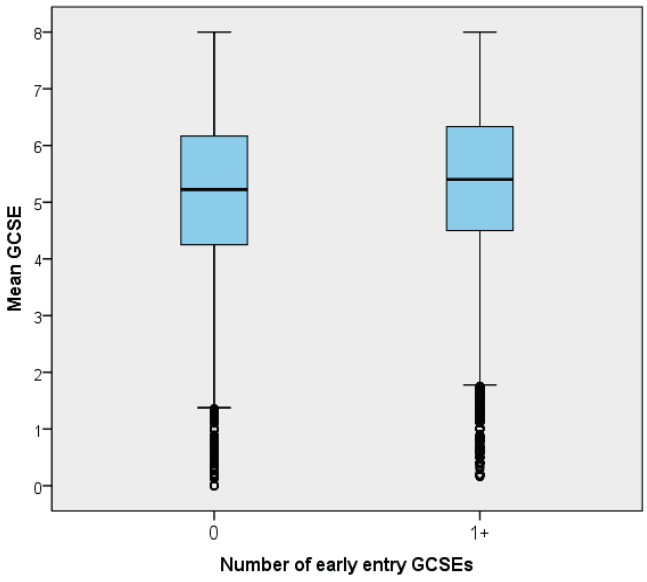
Analysis 1 – Mean GCSE, single treatment group

The first set of analyses compared the mean GCSE scores for those taking at least one GCSE early with those not taking any early. It should be noted that there were fewer students in this analysis than in Figure 1 because it was not possible to estimate a propensity score for students with missing data for any of the covariates used in the logistic regression model. For example, there were a substantial number of students (40,759) with missing KS2 levels across all three tests (mainly from independent schools). For these analyses there were 453,421 students of which 38.2% entered early for at least one GCSE.

Checking quality of matching

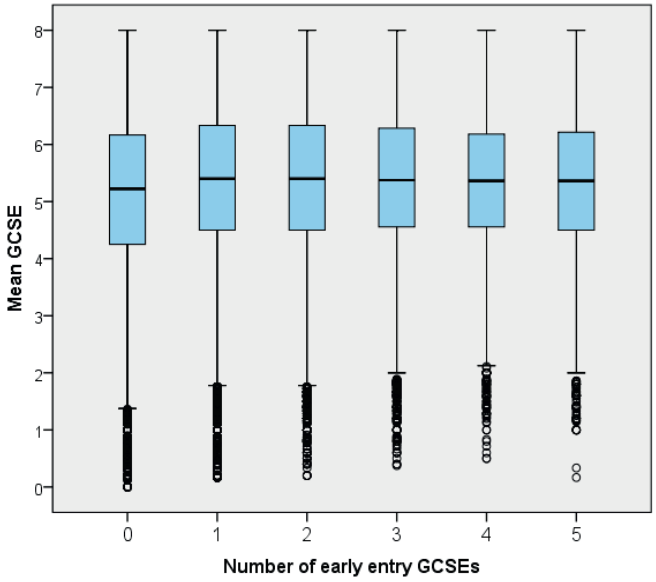
Before presenting the results for each of the analyses in terms of the outcome variable (GCSE mean), it is important to check the quality of the process undertaken to match the treated and non-treated group in terms of their covariates. This was done by comparing the mean values for all covariates between the non-treated and treated groups, before and after weighting. It is normally recommended that statistical tests are undertaken to check whether there are significant differences and, should any be found, the model may need to be re-specified. However, because of the very large sample sizes involved, statistical tests (e.g. t-test) are highly likely to come out as significant even if the differences are very small. Thus, the approach used here was to note any particularly large differences and take them into account when interpreting the results.

For the sake of brevity the full results of this checking are not presented here. However, we note that for this analysis the matching worked very well for all the covariates, with the values for the weighted non-treated group very close to the values for the treated group.



	0	*1+*
Mean	5.16	5.33
Std Dev.	1.41	1.38
N	328,246	195,282

Figure 1: Distribution of mean GCSE scores for early entry and non-early entry students



	0	*1*	*2*	*3*	*4*	*5*
Mean	5.16	5.34	5.34	5.33	5.29	5.26
Std Dev.	1.41	1.38	1.39	1.35	1.32	1.36
N	328,246	130,738	43,286	13,704	3,919	3,636

Figure 2: Distribution of mean GCSE scores by number of early entry GCSEs

Table 2: Mean GCSE performance of early entry and non-early entry groups

Analysis	*Subgroup*	*No of students*	*% early entry*	*GCSE mean (T)*	*GCSE mean (NT–UW)*	*GCSE mean (NT–W)*	*Prob*
Main		453,421	38.2	5.28	5.03	5.27	0.4598
Gender	Female	223,024	39.0	5.50	5.26	5.48	0.5074
	Male	215,973	37.8	5.19	4.95	5.18	0.7363
Prior attainment	High	107,946	45.8	6.44	6.40	6.44	0.9600
	Medium	160,455	39.1	**5.40**	5.37	**5.43**	**0.0314**
	Low	188,234	33.0	4.21	4.12	4.22	0.6532
School type	Comp	414,712	38.0	5.24	4.97	5.23	0.7576
	Independent	23,426	22.2	6.57	6.47	6.52	0.4186
	Selective	20,538	36.7	6.75	6.62	6.74	0.8995
	Secondary Modern	16,912	46.9	4.92	4.69	4.94	0.7907

Comparison of means

The results in terms of the mean GCSE variable are shown in Table 2. This shows the number of students in each subgroup, the percentage taking at least one GCSE early and the mean of the mean GCSE variable for the treated (T) and non-treated groups (both un-weighted [NT-UW] and weighted [NT-W]). The difference in means between the treated and non-treated (weighted) groups was assessed with a test of statistical significance[3]. Statistically significant differences are highlighted in bold.

In terms of GCSE mean the students in the treated group performed better than those in the non-treated group (un-weighted), both overall and in each of the subgroups. However, after weighting these differences almost disappeared and in three of the subgroups the students in the non-treated group performed better. There was only one statistically significant difference, for students in the medium prior attainment group, where students entering early had a significantly lower mean GCSE (5.40) than those not entering early (5.43). However, this difference was very small (only 1/33rd of a grade), or the equivalent to 0.3 of a grade in one GCSE for someone who took ten GCSEs in total.

Analysis 2 – Mean GCSE, two treatment groups

The second analysis investigated whether there was an effect of different numbers of early entries on GCSE performance. Students were classified into groups based on how many GCSEs they entered for early (none, one, or two or more). Table 3 presents the number and percentage of students in each group.

Table 3: Number of students in early entry groups

No of early exams	*Number of students*	*% of students*
0	280,121	61.8
1	115,257	25.4
2+	58,098	12.8
Total	453,476	

3. The 'Surveyreg' procedure in SAS was used to test for differences in the means. This accounts for the effect of clustering of students within schools.

Thus, the majority of students did not take any exams early (61.8%) and about 13% took two or more. As before, analyses were undertaken comparing the performance for the whole cohort of students and then separately for students in each subgroup.

Checking quality of matching

In contrast to the first analysis, the data from the treated groups (one early entry and two or more early entry) were weighted to match the data from the non-treated. The quality of this procedure was checked by comparing the mean values for all covariates between the non-treated and treated groups, before and after weighting.

Overall, the matching was very good, and there were no issues at all between the first treated group (one early entry) and the non-treated group. However, for two of the subgroups the matching between the second treated group (two or more early entry) and the non-treated group was not ideal on all variables. Specifically, for the high attaining subgroup, there was a mismatch on the school type variable after weighting, with 76.10% of the second treated group attending a comprehensive school, compared with 71.68% of the non-treated group. For the selective school subgroup the matching was poor on the gender variable after weighting, with 58.69% of the second treated group being female, compared with 50.83% of the non-treated group. Furthermore, for this subgroup, 75.50% of the second treated group were white, compared with 81.15% of the non-treated group. Therefore, we need to acknowledge these differences when interpreting the results for these subgroups.

Comparison of means

Table 4 presents a comparison of mean GCSE scores for the non-treated group (NT), the treated, un-weighted group (T-UW) and the treated, weighted group (T-W). Separate rows in the table compare the students in the treated groups (one GCSE early or two or more GCSEs early) with those not taking any GCSEs early. Thus for the analysis of all students ('Main') the non-treated group had a mean GCSE of 5.10. The mean for the group taking one GCSE early was 5.29 (un-weighted) and 5.07 (weighted). The figures for the group taking two or more GCSEs early were 5.30 (un-weighted) and 4.98 (weighted).

Again, statistical tests were undertaken to assess whether differences in the mean between the non-treated group and the treated, weighted groups were significant.

Table 4: Mean GCSE performance of (multiple) early entry and non-early entry groups

Analysis	*Subgroup*	*No of early entry*	*No of students*	*GCSE mean (NT)*	*GCSE mean (T–UW)*	*GCSE mean (T–W)*	*Prob*
Main		1	126,562	5.10	5.29	5.07	0.2968
		2+	63,288	**5.10**	5.30	**4.98**	**0.0087**
Gender	Female	1	63,899	5.27	5.44	5.24	0.3948
		2+	33,647	**5.27**	5.46	**5.17**	**0.0332**
	Male	1	62,633	4.93	5.13	4.90	0.3422
		2+	29,641	**4.93**	5.12	**4.80**	**0.0079**
Prior attainment	High	1	33,820	6.49	6.50	6.52	0.2251
		2+	18,521	**6.49**	6.43	**6.41**	**0.0266**
	Medium	1	42,985	**5.42**	5.44	**5.39**	**0.0166**
		2+	21,649	**5.42**	5.37	**5.30**	**<0.0001**
	Low	1	43,613	4.15	4.23	4.13	0.3615
		2+	19,917	**4.15**	4.18	**4.00**	**0.0002**
School type	Comp	1	111,805	4.95	5.20	4.93	0.4433
		2+	55,321	4.95	5.23	4.87	0.0624
	Independent	1	3,971	6.45	6.58	6.49	0.573
		2+	1,320	6.45	6.47	6.37	0.534
	Selective.	1	4,807	6.67	6.81	6.68	0.931
		2+	3,178	6.67	6.72	6.63	0.845
	Secondary Modern	1	5,406	4.58	4.90	4.57	0.909
		2+	3,115	4.58	4.76	4.38	0.151

For the analysis using all data ('Main') there was a statistically significant difference in means for the second treated group only (p=0.0087). The performance of the non-treated group (5.10) was better than the second treated group (4.98) after weighting had been applied. Similar results were also found in each of the subgroups, with the differences being statistically significant except in the school type subgroups. In each case the performance of the non-treated group was better than the second treated group after weighting. The differences varied from 0.07 of a grade (comprehensives) to 0.20 of a grade (secondary moderns). This suggests that early entry (of two or more subjects) had a negative impact on overall performance at GCSE, for students overall and for several of the subgroups that were analysed.

Comparing students in the first treated group with those in the non-treated group, the differences were very small. The only statistically significant difference was for students in the medium attaining group, where students in the non-treated group performed better (5.42) than those in the treated group (5.39).

Analysis 3 – Accountability measure, single treatment group

This analysis is with a single treatment group, but with the outcome measure being the percentage of students passing the school accountability target of five or more GCSE grades A* to C including English and Mathematics. As before, this analysis was done for all students and then each of the subgroups. The same propensity scores were used as in the GCSE mean analyses, so there was no need to check the quality of matching.

Comparison of percentages

Table 5 compares the percentage of students passing the threshold measure in the treated group with the percentage[4] passing in the non-treated groups (weighted and un-weighted). A test of statistical significance was undertaken of the difference in percentage between the treated group and the weighted non-treated group.

There is a clear pattern in these results with a significantly higher percentage of students in the treated group passing the threshold measure than those in the non-treated group (after weighting), overall and in most of the subgroups. Amongst all students, 73.77% of the treated group passed, compared with 70.97% of the non-treated group.

The exceptions to this pattern were in the low attaining and the independent school groups, where a significantly higher percentage of the non-treated group passed (62.55% and 91.10% respectively) than the treated group (57.97% and 88.73% respectively).

These results suggest that there may have been some advantage in schools entering students early for some GCSEs, in terms of getting more students to pass the threshold measure (except for low attaining students and those in independent schools).

Analysis 4 – Accountability measure, two treatment groups

This analysis investigated whether there was an effect of different numbers of early entries on the percentage of students passing the accountability measure.

Comparison of percentages

Table 6 compares the percentage of students passing the target measure in the non-treated group with the percentage passing in the two treated groups.

4. The 'surveylogistic' procedure in SAS was used to test for differences in the proportions, taking into account the clustering of students within schools.

Table 5: Threshold measure success rate of early entry and non-early entry groups

Analysis	*Subgroup*	*No of students*	*% early entry*	*% passing (T)*	*% passing (NT–UW)*	*% passing (NT–W)*	*Prob*
Main		453,421	38.2	**73.77**	63.19	**70.97**	**<0.0001**
Gender	Female	223,024	39.0	**76.61**	66.75	**73.84**	**<0.0001**
	Male	215,973	37.8	**72.20**	61.95	**69.99**	**<0.0001**
Prior attainment	High	107,946	45.8	98.49	97.69	98.42	0.5747
	Medium	160,455	39.1	**85.23**	80.39	**83.62**	**<0.0001**
	Low	188,234	33.0	**57.97**	66.62	**62.55**	**<0.0001**
School type	Comp	414,712	38.0	**73.06**	61.89	**70.31**	**<0.0001**
	Independent	23,426	22.2	**88.73**	90.72	**91.10**	**0.0325**
	Selective	20,538	36.7	99.23	98.55	99.35	0.6485
	Secondary Modern	16,912	46.9	66.81	54.59	64.13	0.2613

Table 6: Threshold measure success rate of (multiple) early entry and non-early entry groups

Analysis	*Subgroup*	*No of early entry*	*No of students*	*% passing (NT)*	*% passing (T–UW)*	*% passing (T–W)*	*Prob*
Main		1	126,562	64.22	71.60	64.60	0.5910
		2+	63,288	64.22	76.62	65.84	0.1570
Gender	Female	1	63,899	66.73	73.54	67.21	0.5335
		2+	33,647	66.73	78.82	68.90	0.0825
	Male	1	62,633	61.67	69.63	61.97	0.7220
		2+	29,641	61.67	74.12	62.57	0.5030
Prior attainment	High	1	33,820	97.68	98.24	97.60	0.7230
		2+	18,521	97.68	98.75	97.29	0.1510
	Medium	1	42,985	81.14	84.18	81.21	0.8800
		2+	21,649	81.14	87.46	81.99	0.2990
	Low	1	43,613	**34.23**	40.51	**36.10**	**0.0052**
		2+	19,917	**34.23**	45.93	**37.28**	**0.0096**
School type	Comp	1	111,805	61.25	70.31	62.21	0.1936
		2+	55,321	**61.25**	75.82	**64.50**	**0.0085**
	Independent	1	3,971	90.13	88.74	88.95	0.3279
		2+	1,320	**90.13**	86.89	**85.37**	**0.0348**
	Selective	1	4,807	98.82	99.11	98.53	0.5224
		2+	3,178	98.82	99.34	97.36	0.0943
	Secondary Modern	1	5,406	51.54	63.41	52.31	0.7870
		2+	3,115	51.54	65.14	50.79	0.8430

For the main analysis and most of the subgroup analyses there was very little difference in the percentages of students passing the threshold after weighting, although the treated groups tended to do slightly better. There was only one subgroup with a statistically significant difference between the non-treated and first treated group. This was for the low attaining students, with 36.10% of the first treated group achieving the threshold compared with 34.23% of the non-treated group (p=0.0052). For this subgroup, students in the second treated group were also significantly more likely to achieve the threshold than the non-treated group (37.28%, p=0.0096). There were two other subgroups with significant differences between the second treated group and the non-treated group. For comprehensive school students, 64.60% of the second treated group achieved the threshold, compared with 61.25% of the non-treated group. In contrast, a lower percentage of independent school students in the second treated group achieved the threshold (85.37%) than those in the non-treated group (90.13%).

These results suggest that there seemed to be little advantage for those taking just one GCSE early (except for low attaining students), and the advantage for those taking two or more GCSEs early was limited to comprehensive school students and low attainers. Independent school students were disadvantaged if they took two or more GCSEs early.

These results are somewhat at odds with the results for the single treatment group (Table 5), which had significant differences in most of the subgroups and larger differences in percentage of students passing. This finding is discussed further in the conclusion.

Analysis 5 – Total points score, single treatment group

For the final two analyses the outcome measure was the total points score, across all GCSEs and equivalents. As before this analysis was undertaken for all students and then each of the subgroups. Again, the same propensity scores were used as in the GCSE mean analyses, so there was no need to check the quality of matching.

Comparison of means

Table 7 compares the mean total points score in the treated and non-treated groups.

For the main analysis and each subgroup analysis the treated group had a higher mean total points score than the non-treated (weighted) group. This difference was statistically significant in the main analysis and in the female, male and comprehensive schools subgroups. However, the differences were not large, being about 6 points, equivalent to one GCSE grade in one GCSE.

This suggests that students entering early for some GCSEs, whilst not doing significantly better on their GCSEs (see Table 2), tend to perform better on the GCSE equivalent qualifications, leading to a higher total points score.

Analysis 6 – Total points score, two treatment groups

Table 8 presents the results of the analysis of mean total points score with two treatment groups.

The differences between the first treated group and the non-treated

Table 7: Mean total points score of early entry and non-early entry groups

Analysis	*Subgroup*	*No of students*	*% early entry*	*Mean points total (T)*	*Mean points total (NT–UW)*	*Mean points total (NT–W)*	*Prob*
Main		453,421	38.2	**540.9**	474.0	**534.4**	**0.0183**
Gender	Female	223,024	39.0	**557.5**	491.2	**551.2**	**0.0318**
	Male	215,973	37.8	**528.3**	463.1	**522.3**	**0.0424**
Prior attainment	High	107,946	45.8	629.3	568.3	624.6	0.1185
	Medium	160,455	39.1	549.7	496.8	547.3	0.3809
	Low	188,234	33.0	458.4	411.4	453.8	0.1137
School type	Comp	414,712	38.0	**538.2**	470.6	**532.3**	**0.0393**
	Independent	23,426	22.2	483.0	455.5	479.1	0.5723
	Selective	20,538	36.7	628.5	557.1	623.6	0.6991
	Secondary Modern	16,912	46.9	518.4	467.8	515.6	0.7758

Table 8: Mean total points score of (multiple) early entry and non-early entry groups

Analysis	*Subgroup*	*No of early entry*	*No of students*	*Mean points total (NT)*	*Mean points total (T–UW)*	*Mean points total (T–W)*	*Prob*
Main		1	126,562	470.77	520.12	469.44	0.5587
		2+	63,288	**470.77**	575.42	**463.96**	**0.0459**
Gender	Female	1	63,899	485.15	533.78	484.13	0.6730
		2+	33,647	485.15	589.73	479.75	0.1590
	Male	1	62,633	456.25	506.18	454.72	0.5536
		2+	29,641	**456.25**	559.18	**447.12**	**0.0184**
Prior attainment	High	1	33,820	559.08	605.57	560.31	0.6090
		2+	18,521	559.08	660.27	556.22	0.4680
	Medium	1	42,985	493.24	531.52	491.52	0.4180
		2+	21,649	**493.24**	582.47	**485.70**	**0.0178**
	Low	1	43,613	409.72	445.52	408.90	0.7373
		2+	19,917	**409.72**	489.71	**399.58**	**0.0075**
School type	Comp	1	111,805	468.51	519.37	467.46	0.6680
		2+	55,321	468.51	575.43	464.44	0.2660
	Independent	1	3,971	453.37	479.31	453.76	0.9520
		2+	1,320	453.37	487.83	442.76	0.3230
	Selective.	1	4,807	565.06	607.53	561.68	0.6810
		2+	3,178	565.06	667.01	571.03	0.6550
	Secondary Modern	1	5,406	459.59	502.23	452.73	0.4743
		2+	3,115	459.59	526.22	435.87	0.0528

group were generally very small, and none were statistically significant. However, students in the non-treated group did perform significantly better than those in the second treated group, for the main analysis and in three of the subgroup analyses (males, medium attainers and low attainers). In the main analysis, the difference was around 7 points, equivalent to just over one grade in a GCSE. The differences were slightly larger amongst males (9 points), medium attainers (7.5 points) and low attainers (10 points). There was also a considerably larger difference in the secondary modern subgroup (24 points), although this was not quite large enough to be statistically significant. Thus, students entering early for two or more GCSEs seemed to be disadvantaged in terms of their overall KS4 points score.

Again, these results are somewhat at odds with the results with a single treatment group (Table 7), which had students in the treated group performing better, on average, than those in the non-treated group. This finding is discussed further in the next section.

Discussion

The purpose of this research was to investigate whether students who take one or more GCSEs before Year 11 perform any differently than those not doing so, across all GCSEs (and equivalents). Many schools enter some students early for at least one GCSE and this may be for a number of reasons, such as trying to get students over the crucial threshold grade C, or getting some qualifications out of the way to allow students to focus on other subjects in Year 11. Students may also be entered early in order to give them a chance to re-sit if they do not do as well as expected.

By looking at three different measures of success at GCSE it was possible to investigate the effect of early entry on individual students and also on schools' performance through the percentage of students passing the important threshold measure of five grades A* to C including English and Mathematics.

For individual students there does not seem to be any advantage in early entry in terms of overall GCSE performance. Comparing those taking at least one GCSE early with those not doing so showed there to be almost no difference in mean GCSE. Indeed, Table 4 showed that for students taking two or more GCSEs early there was a significant disadvantage compared with non-early entry students. The analysis of the whole cohort found a difference of 0.12 of a grade, equivalent to one grade lower in one GCSE for a student taking eight GCSEs. This significant effect was also present for all the subgroups investigated apart from the school type subgroups.

However, it seems there may be some advantage in early entry if we consider other measures of performance. Early entry students had a statistically significantly higher mean total points score than those not taking any GCSEs early (Analysis 5). This amounted to 6.5 points in the cohort as a whole, equivalent to about one GCSE grade (which is, perhaps, not a large difference over ten or more GCSEs, but significant nonetheless). This difference was present in all subgroups (although only significant in the male, female and comprehensive school subgroups). When the analysis was limited to students with two or more early entry GCSEs a different pattern emerged, with these students tending to perform worse on this measure than those not taking any GCSEs early. This was the case for all students taken together and amongst males, low attainers and medium attainers (Analysis 6).

When looking at the percentage of students passing the threshold measure (Analysis 3), students in the early entry group performed significantly better (73.77% passed) compared with the non-early entry students (70.97% passed). This was also the case for most of the subgroups. In the low attaining and independent school subgroups, however, the early entry students performed significantly worse than non-early entry students. When comparing students with different numbers of early entry GCSEs the positive effect of early entry on the threshold measure was only present for low attainers and for comprehensive school students taking two or more early (Analysis 4).

It should be noted that there is an important difference in the interpretation of any differences between treated and non-treated groups depending on whether we are talking about the single treatment or multiple treatment case, because of the different weighting methods in each case. In the single treatment case we were estimating the average treatment effect for the treated (ATT). This was done by comparing the actual results for the treated group with the estimated results for the treated group *had they not been treated* (the weighted, non-treated group). In contrast, the two treatment case involves estimating the average treatment effect for the non-treated (ATNT). This is done by comparing the actual results for the non-treated group with the estimated results for the non-treated group *had they been treated* (the weighted, treated groups). So any observed differences are only really relevant for the group in question.

This distinction may be the reason why there was a positive effect of early entry when looking at the single treatment group (in terms of accountability measure and total points score) and a less positive (for accountability measure) or even negative (for total points score) effect when looking at the two treatment groups. In other words, the effect of early entry seems to have been more positive (on average) for the treated, than it would have been for the non-treated. This suggests that, to a degree, teachers are correctly choosing the early entry students as those most likely to benefit from it.

It is interesting that early entry seems to be successful in getting a larger percentage of students to pass the threshold measure, but it is not better for individual students (at least in terms of GCSE mean). This apparent contradiction is presumably because early entry is more successful in getting students around the C boundary to improve their grade than getting A and B grade students to perform to their potential. These results also corroborate the findings from previous studies (Gill, 2013) that high attaining students are least likely to benefit from early entry (in individual subjects).

It is also interesting that independent school students who take exams early have a higher mean GCSE than those that don't (although this difference is not significant), but are significantly less likely to pass the threshold measure, suggesting that independent schools' focus is on individual students rather than the threshold measure (which they are not judged on).

One interesting hypothesis that may be worthy of further research is whether students who were disadvantaged by early entry in individual subjects (e.g. high performing students in English and Mathematics) are able to make this up in Year 11. From the results presented here we cannot know whether this is the case because we do not identify students who performed below expectations on their early entry exams. All we can say is that, on average, students who take at least one GCSE early are not disadvantaged in terms of overall GCSE, and actually perform better in terms of overall points score. In contrast, it is estimated that those who did

not enter early would have performed worse if they had taken two or more GCSEs early. Further research could also estimate the average treatment effect for the treated in the case of two treatment groups, to see if taking two or more GCSEs early is beneficial to these students or not.

Finally, it will be interesting to see the impact of GCSE reforms on the amount of early entry. Students will still be able to sit GCSEs in Year 10, but changes to accountability measures mean that only the result from the first sitting of a GCSE will count in performance tables. This is likely to lead to a fall in early entry because schools may want to wait until students are ready to achieve their best possible grade, rather than getting them to sit GCSEs early and then re-sit if they underperform.

References

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.

Gill, T. (2013). Early entry GCSE candidates: Do they perform to their potential? *Research Matters: A Cambridge Assessment Publication, 16*, 23–40.

McCaffrey, D.F., Ridgeway, G., & Morral, A.R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9*(4), 403–425.

Morgan, S.L., & Harding, D.J. (2006). Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods & Research, 35*(1), 3–60.

Ofsted (2013). Schools' use of early entry to GCSE examinations. Its usage and impact. Manchester: Ofsted.

Rosenbaum, P.R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Big data and social media analytics

Vikas Dhawan and **Nadir Zanini** Research Division

Introduction

'Big data' is fast becoming an area of great importance for businesses in many areas, including education. In simple terms it refers to the combination of data from various sources and understanding patterns in the data which can be used for various purposes such as improving market intelligence and educational research. Businesses, large and small, are implementing (or planning to implement) big data strategies. Apart from market intelligence, it is being applied in diverse areas such as healthcare and other scientific research, complex manufacturing industries such as aviation and heavy machinery, improving public utilities and traffic management, oil and gas exploration, telecoms, retail, banking and insurance, defence and security.

In this article we give an introduction to big data and some of its applications in various fields, including education. We also describe the use of big data for the monitoring of social media (for instance LinkedIn, Facebook and Twitter) for market growth and brand management. Some training courses in big data offered by various universities are mentioned in the article.

Applications in the education industry mentioned in this article include the combination of various sources of information about pupils such as test records, behaviour patterns, and teacher observations over a period of time for providing more accurate and timely interventions. In addition to this, we discuss new forms of assessment such as e-assessment and adaptive testing which will provide new streams of data which could be tapped for studying the performance of test takers in more detail and for monitoring and evaluation of tests.

Big data

Technological advances in recent years have led to a significant amount of data which is now generated in everyday life, such as shopping, travelling, banking, manufacturing and trading, public utilities, state and governance, sports, entertainment, science, education and health. Commercial organisations, research bodies and governments have started to realise the importance of using this data for their growth. As a result, the study of big data has gained prominence among scholars in different areas of research (Einav & Levin, 2013; Mayer-Schönberger & Cukier, 2013) as well as generating interest from the non-academic world (BBC, 2013; Lohr, 2012).

The concept of big data encompasses the collection of data, the combination of the data collected from various sources, processing it and using the results so obtained. Specifically, big data is a term used for large databases requiring complex processing and visualisation which cannot be efficiently handled by traditional data processing software (Wikipedia, 2014a). According to the McKinsey Global Institute, "Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika *et al*., 2011). A well-known model (known as 3V's model) of big data attributed to Gartner Inc. defines it as "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" (Beyer & Laney, 2012). The term 'volume' here indicates the complexity of datasets and not necessarily their size. 'Variety' refers to the different type of structured or unstructured data such as text and