Evaluation of the impact of IntoUniversity's Academic Support on Key Stage 2 attainment

Report from FFT Education Datalab to IntoUniversity

Dave Thomson

Natasha Plaister

April 2021

1. Executive summary

1.1 Methodology

- This report evaluates the impact of IntoUniversity's Academic Support programme on Key Stage 2 scaled scores in maths and reading between 2015/16 and 2018/19.
- Our analysis used data from the National Pupil Database (NPD) to compare the performance of IntoUniversity pupils to the performance of those in a matched comparison group.
- We also looked at how the impact of the programme varied with respect to dosage; that is, by how much a pupil had engaged with the programme, and by the length in time over which a pupil engaged with the project.

1.2 Main findings

- We found no evidence of significant positive effects on either scaled scores in maths or reading for IntoUniversity pupils overall.
- However, we did find some evidence of increased effect on pupils with higher engagement with the project. We found a significant positive effect for high dosage pupils (those who took part in 80 or more sessions) in 2018/19, the equivalent of four months of additional progress.
- When estimates were pooled across all years, we also found a significant positive effect for high dosage pupils, which was slightly lower than that found for 2018/19 alone, the equivalent of three months of additional progress.
- We did not find evidence to show increased effect for pupils who have been engaged with the programme over a longer period of time.
- We found a significant positive effect for pupils who have been involved in the programme for a medium period of time (four to eight terms) in 2018/19, the equivalent of five months of additional progress.

1.3 Limitations

- The approach used for this evaluation relies on creating a matched comparison group of pupils who are statistically similar to pupils who received support from IntoUniversity, using data from the NPD. Creating a comparison group in this way means that we are unable to control for factors not recorded in the NPD, such as pupils' motivation, parental occupation or school funding level.
- Some comparison pupils may have received similar support to that offered by IntoUniversity from other sources. If this was the case, it may have led to underestimation of effects.
- Due to low sample sizes, we were unable to provide estimates of how the effect varies with respect to dosage of length of time involved with the project for pupils who completed KS2 in 2015/16.
- Bias may have arisen from the way that pupils were selected for inclusion in this evaluation. Only those IntoUniversity pupils who had given consent for their data to be shared were eligible for inclusion.

2. Introduction

IntoUniversity's Academic Support is a holistic programme for primary and secondary students from disadvantaged backgrounds. Students are supported to develop social, emotional and study skills, as well as receiving help with homework, coursework, literacy and numeracy. This report focuses on the impact of this support on Key Stage 2 attainment, specifically scaled scores in reading and maths in national statutory tests.

We used data from the National Pupil Database to compare the performance of IntoUniversity pupils who completed Key Stage 2 between 2016 and 2019[1] to the performance of those in a matched comparison group. We looked at how the impact of the programme varied with respect to *dosage*; that is, by how much a pupil had engaged with the programme, in this case measured by the number of sessions attended. We also looked at how impact varied with respect to the length in time over which a pupil engaged with the project, measured by the number of terms over which a pupil received support.

2.1 Methodology

This evaluation uses what is known as a *quasi-experimental design*. It involves comparing the outcomes of pupils who received academic support from IntoUniversity to pupils in a matched comparison group of statistically similar pupils. This approach mimics what is done in a formal experiment such as a randomised control trial.

We selected pupils who were similar with respect to:

Pupil characteristics:

- whether they were eligible for the Pupil Premium
- IDACI score
- ethnic group
- whether they had English as an additional language
- gender
- month of birth
- special education needs
- prior attainment at foundation stage
- prior attainment at Key Stage 1

School characteristics:

- proportion of pupils who were eligible for the Pupil Premium
- attainment at Key Stage 2 for three years before the outcome year
- region

In some cases, data on some of the characteristics listed above was not available. For example, pupils who moved to England after age 7 would be missing both Foundation Stage and Key Stage 1 data. Pupils with missing data were matched to other pupils with missing data. Matching was done using a mixture of

[1] Throughout this report, years will refer to the year in which the academic year finished: 2016 will refer to the academic year 2015/16, for example.

nearest neighbour pair matching based on propensity score and coarsened exact matching. The matching process is described in more detail in section 3.2.

We then used regression models to compare the outcomes of IntoUniversity (IU) pupils to those of pupils in the matched comparison group, with a dummy variable to indicate whether a pupil had taken part in the programme or not. Confidence intervals were obtained by using bootstrapping with 1,000 iterations.

2.2 Data

The source data for this evaluation is the National Pupil Database (NPD), an administrative resource maintained by the Department for Education covering state-funded schools in England. This includes records of national curriculum tests taken in reading and maths at the end of Key Stage 2 (end of primary school) and details of pupil enrolments and pupil characteristics sourced from the termly School Census. These two key datasets are linked together and also linked to assessments taken by pupils earlier in their school career (Foundation Stage, Key Stage 1) provided they were in school in England at the time.

Between 400 and 600 pupils participated in IntoUniversity programmes in each cohort; numbers increased each year. Those used in the evaluation are those who gave consent for their details to be matched to NPD. Pupils giving consent in older cohorts all continued with IntoUniversity programmes after they were tested at Key Stage 2. There is a risk that they may not be representative of all participants and therefore we would urge caution when interpreting the results we present.

IntoUniversity supplied an original dataset that included 393 IntoUniversity pupils who completed KS2 between 2016 and 2019, and who had consented for their details to be matched to NPD. Records for all except one of these pupils were found in the NPD. However, one pupil who was identified in the original dataset as having completed KS2 in 2017 actually completed in 2018, according to the NPD. The final dataset included 392 pupils.

Table 1: IntoUniversity pupils included in the evaluation

Year	Pupils in cohort	Pupils in original dataset	Pupils in final dataset
2016	430	31	31
2017	450	70	69
2018	500	95	95
2019	600	197	197
Total	1980	393	392

Pupils were divided in categories for dosage and length of time involved in the programme. Those who had taken part in the programme for three terms or less were defined as having being involved for a *short* length of time, for four to eight terms, a *medium* length of time and for nine terms or more, a *long* length of time. Similarly, those who took part in less than 37 sessions were defined as *low*, between 37 and 80 sessions as *medium*, and 80 sessions or more as *high*.

2.3 Comparison to national average

In this section, we describe how the participating pupils compare to pupils nationally in terms of KS2 attainment.

The assessments currently used to measure attainment at Key Stage 2 were introduced in 2016. This report focuses on attainment in terms of scaled scores in reading and maths. Scaled scores are used instead of raw marks to allow for comparability across years. In tests, scaled scores range from 80-120. However, the Department for Education also extends the lower end of the scale to include pupils working below the standard of the test and assigns values based on teacher assessments. Pupils with a scaled score of at least 100 are deemed to have met the expected standard in the relevant area. Those with a scaled score of at least 110 are deemed to be working at a higher level.[2]

Figure 1: Proportion of pupils achieving the expected level

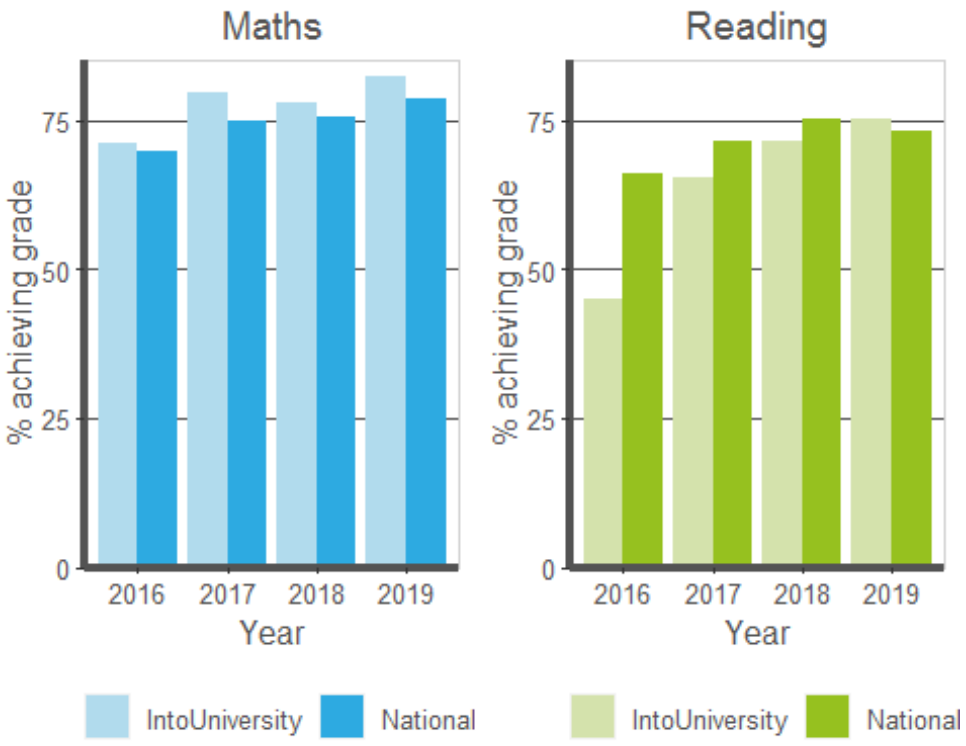


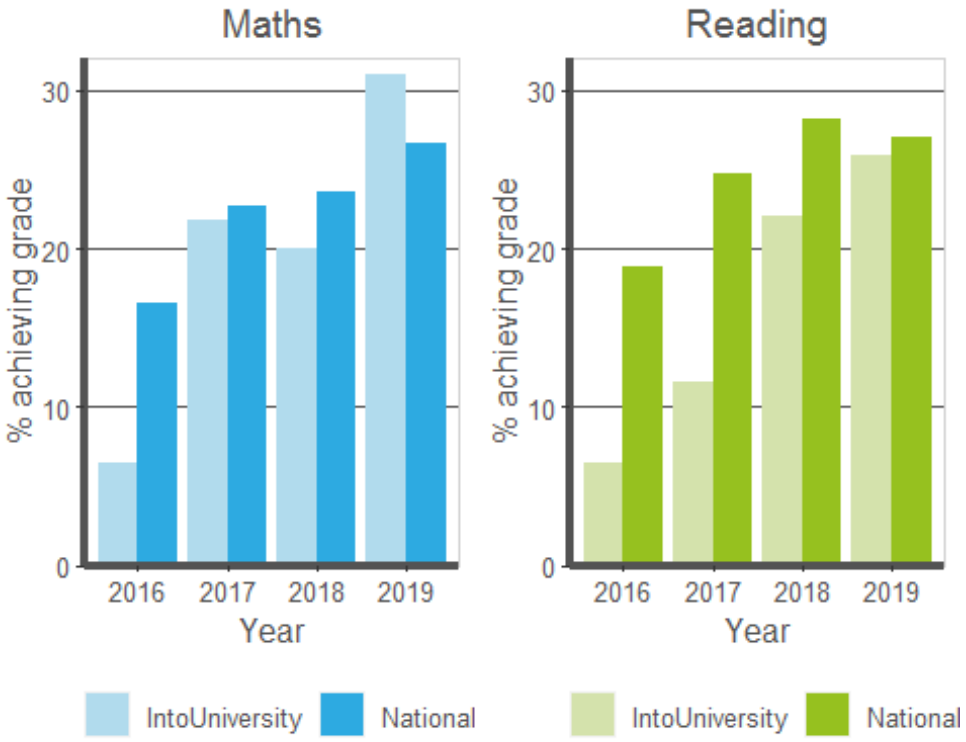
Figure 1 shows how the proportion of IntoUniversity pupils achieving the expected standard in maths and reading at KS2 compared to the national average. In each year from 2016-19, IntoUniversity pupils were more likely to meet the standard in maths than other pupils; in 2019, for example, 82.2% of IntoUniversity pupils met the expected standard compared to 78.6% overall. However, IntoUniversity pupils did tend to perform below the national average in reading. The only exception was the proportion meeting the expected standard in 2019: 75.1% of IntoUniversity pupils met the standard compared to 73.2% nationally.

---

[2] For more about scaled scores and how they are calculated, see: www.gov.uk/guidance/understanding-scaled-scores-at-key-stage-2

Figure 2 compares the proportion of IntoUniversity pupils achieving a higher level in maths and reading to the national average. For maths, the picture is less clear than for the proportion reaching the expected level; in 2016, 2017 and 2018, IntoUniversity pupils were less likely to reach a higher level than the national average. However, in 2019 they were more likely: 31% were working at a higher level, compared to 26.7% nationally. In reading, however, IU pupils were less likely to reach a higher level than the national average in every year from 2016-19.

Figure 2: Proportion of pupils achieving a higher level



In figure 3, we present boxplots showing the distribution of KS2 scaled scores in maths and reading for IU pupils and nationally. We also show the distribution for IU pupils by dosage - that is, by how much a pupil had engaged with the programme, in this case measured by the number of sessions attended – and with respect to the length in time over which a pupil engaged with the project, measured by the number of terms over which a pupil received support.

In general, the attainment of the 2016 to 2018 IntoUniversity cohorts was below the national average in reading. Attainment in maths was broadly similar although exhibited less variation than the national average. By contrast, the attainment of the 2019 cohort was slightly above average in both reading and maths.

The differences in attainment seen in this section should be seen in the context of how the demographics and prior attainment of IntoUniversity pupils compare to those of other pupils. This is discussed in the next section.

Figure 3a: Distribution of KS2 scaled scores, IntoUniversity pupils and national

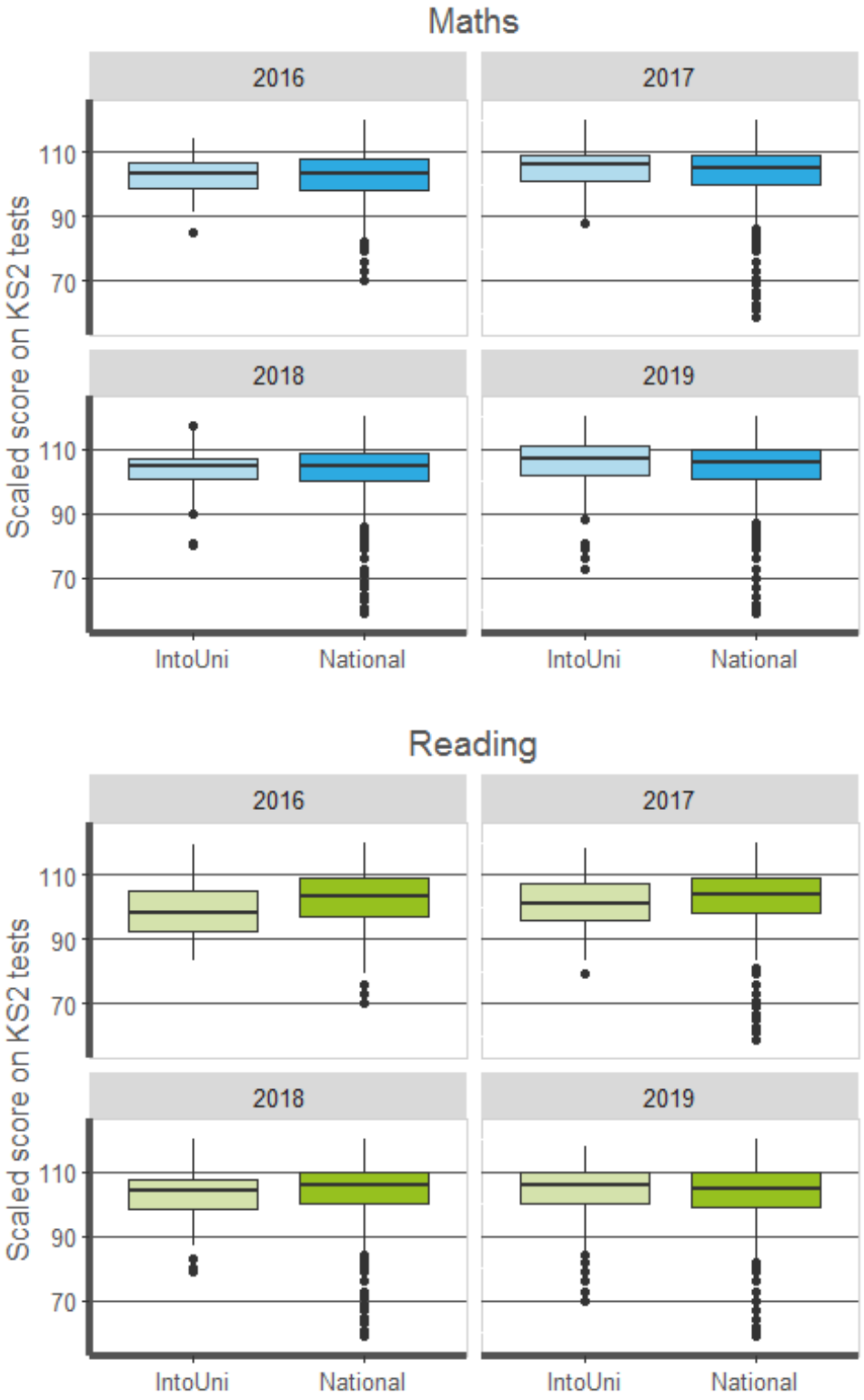


Figure 3b: Distribution of KS2 scaled scores, IntoUniversity pupils by dosage

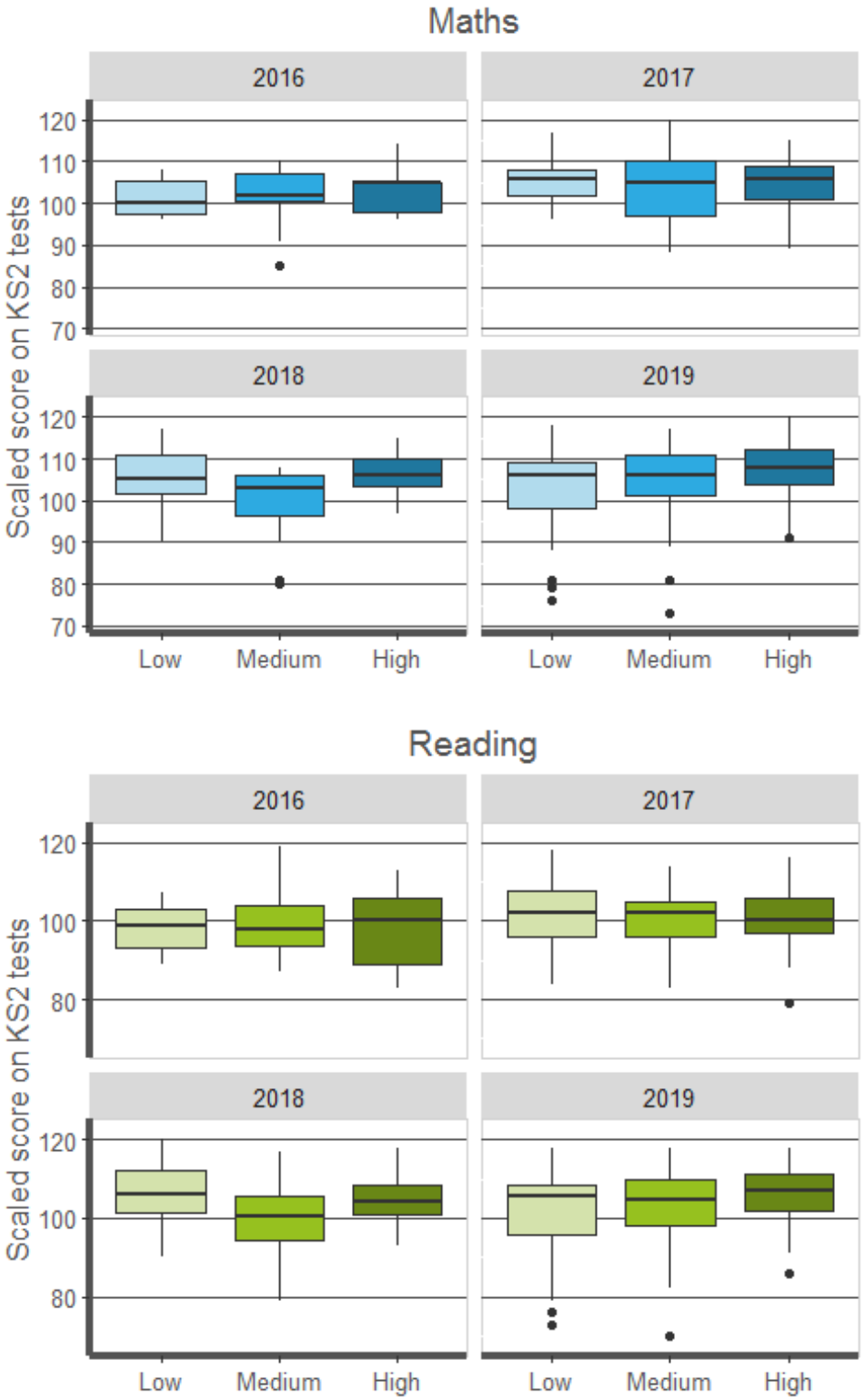
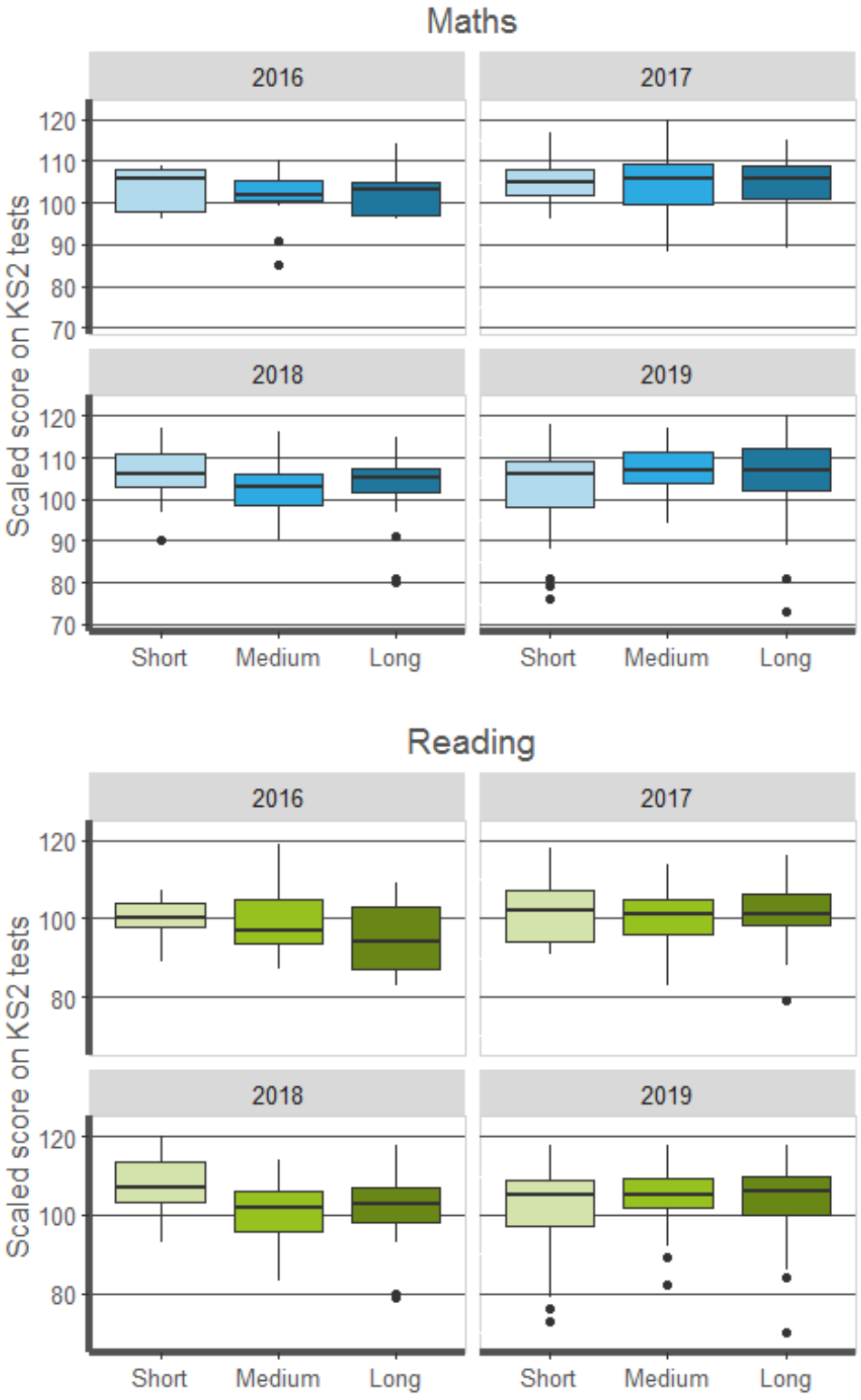


Figure 3c: Distribution of KS2 scaled scores, IntoUniversity pupils by length of time involved with the project



3. Mitigation of confounding effects

3.1 Differences before matching

This section shows how IntoUniversity pupils compared to all other pupils with respect to pupil and school characteristics.

Unsurprisingly, given the nature of the project, IntoUniversity pupils were much more likely to be disadvantaged than other pupils. In 2019, for example, the majority of IU pupils were eligible for the Pupil Premium (59.4%), compared to 29.8% of all other pupils. Their average IDACI score was much higher, at 0.4 compared to 0.2 for all other pupils, and they tended to go to schools with a higher proportion of disadvantaged pupils.

IU pupils were more likely to have English as an additional language than other pupils; in 2019 64.5% of IU pupils were EAL compared to 20.9% of other pupils. There were also some striking differences in ethnicity: the largest ethnic group among IU pupils was black African, while among all other pupils it was white British. In 2019, 38.6% of IU pupils were black African, compared to just 4.0% of other pupils, and 15.2% of IU pupils were white British, compared to 65.8% of other pupils.

The differences described above were similar for all of the years covered in this evaluation. Some differences were less consistent. For example, IU pupils were generally more likely to be female than other pupils (53.3% in 2019, compared to 49% of other pupils), but IU pupils who completed KS2 in 2017 were slightly less likely to be female (43.5% compared to 48.9% of other pupils). IU pupils generally had lower prior attainment at both foundation stage and Key Stage 1, but the difference was smaller in 2019 than in previous years, perhaps suggesting that the pupils included in the analysis of earlier cohorts are substantively different from IU participants in general.

Finally, due to the nature of the programme, IU pupils were concentrated in certain regions. IU pupils who completed KS2 in 2016 were located in just four of the nine regions; by 2019, there were IU pupils in eight regions. However, even in 2019, 31.5% of IU pupils were from London, compared to 15.5% of all other pupils; 17.8% of were from Yorkshire and the Humber, compared to 10.1% of all other pupils.

3.2 Extent of success in creating a matched comparison group

The initial matching process was carried out using the nearest neighbour method, pairing treated and comparison pupils based on propensity scores. A propensity score can be thought of as a measure of how typical each pupil is of the pupils in the treated group. As shown in section 3.1, treated pupils were more likely to be disadvantaged than potential comparison pupils, a higher proportion were black African, and they tended to go to schools with a track record of relatively low attainment. So a pupil who isn't disadvantaged or black African, and who went to a school with a track record of high attainment would be likely to have a low propensity score, and vice versa.

Propensity scores were calculated by fitting a logit model, controlling for the matching variables described in section 2.1. Treated pupils are then paired with the potential comparison pupil with the closest propensity score.

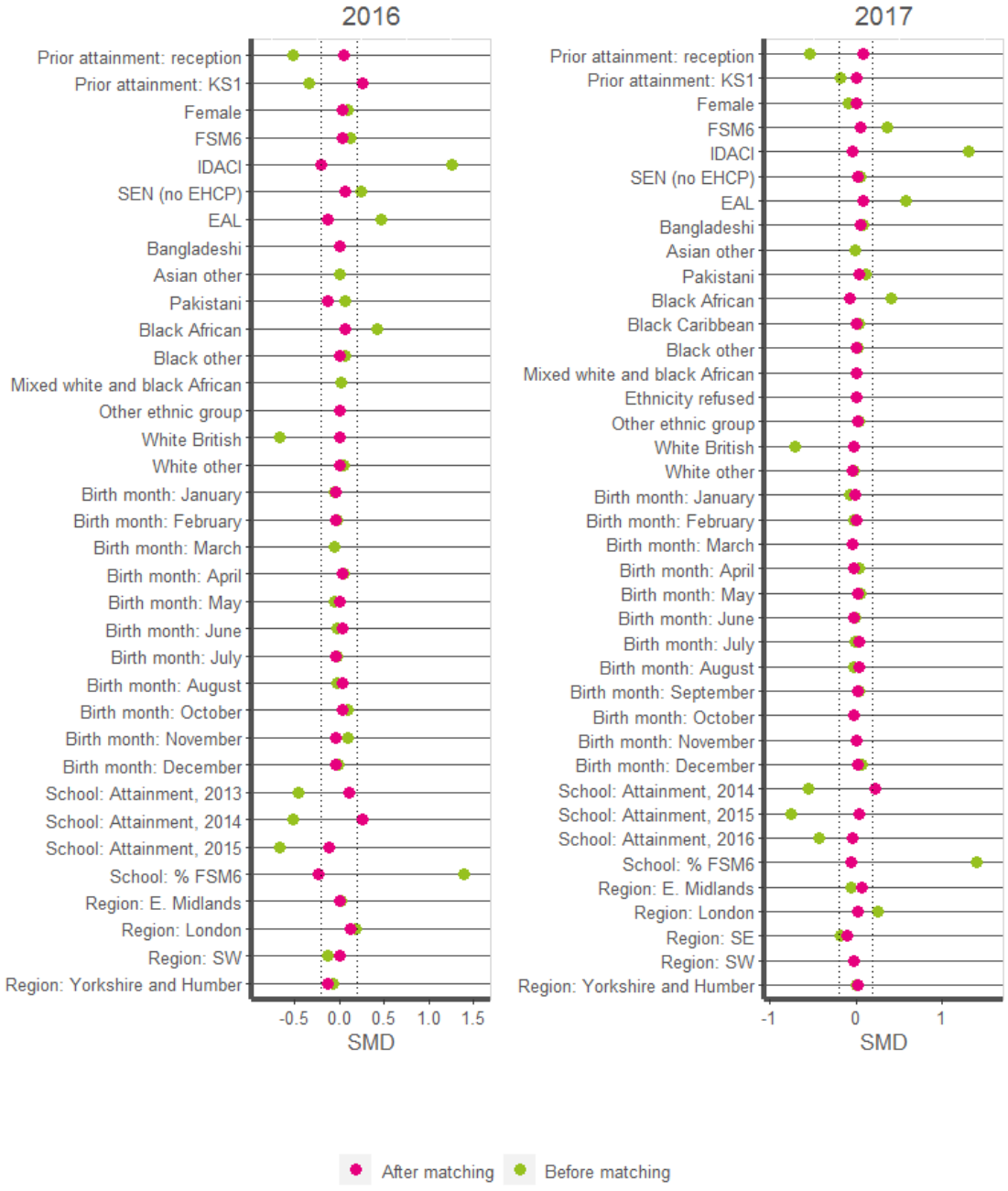
For a small number of pupils, data on some of the matching variables was not available. This was the case for pupils who did not attend a mainstream state-maintained school, including those who attended independent schools or pupil referral units; the NPD does not hold the same data for these institutions as it does for mainstream state-maintained schools. Some pupils did not complete either the foundation stage or Key Stage 1 at an English school; for these pupils, data on prior attainment at these points was not available.

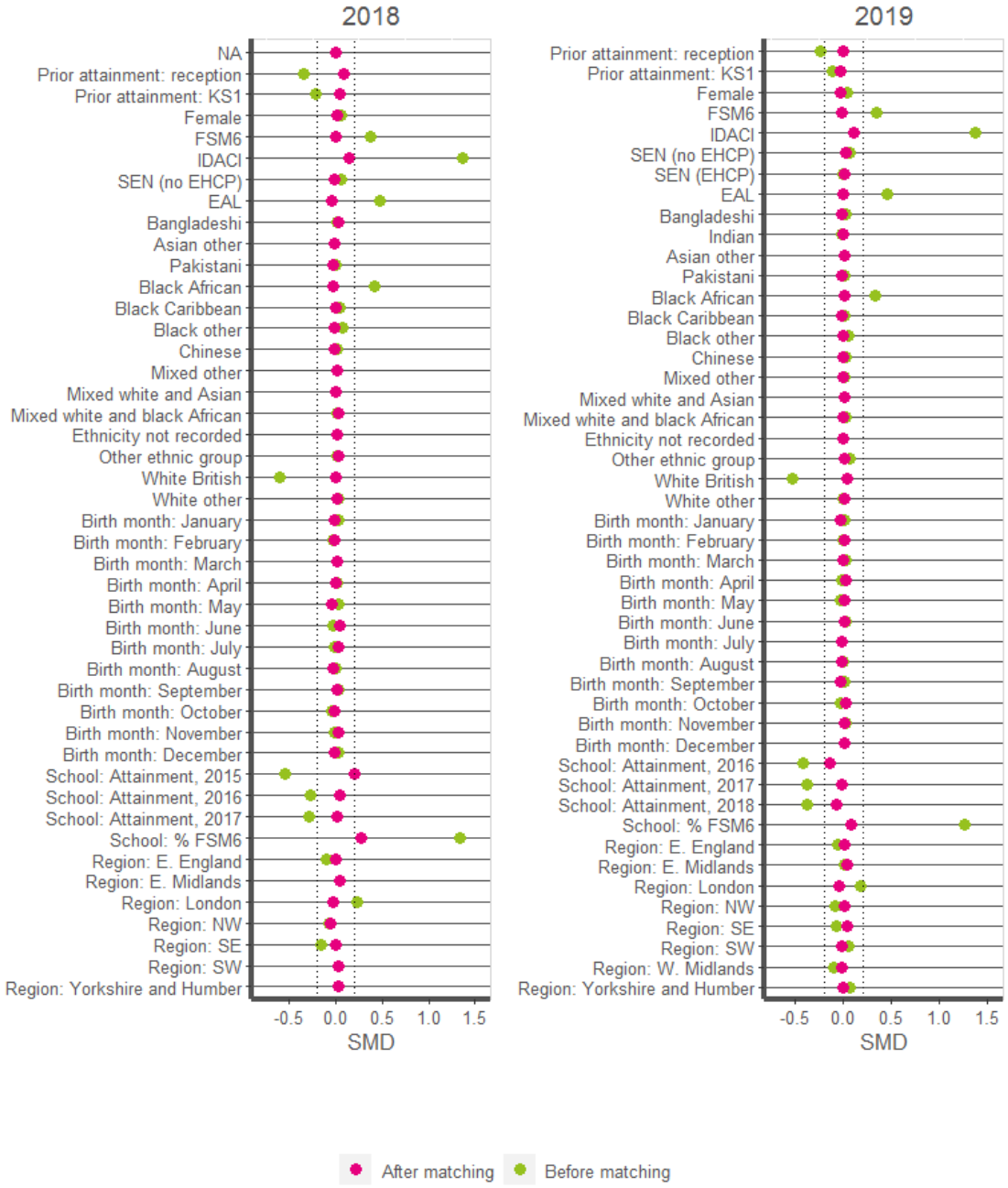
Pupils with missing data were matched to comparison pupils who were also missing data in the same fields, and who were similar with respect to the data that was available. An alternative method, known as coarsened exact matching (CEM) was used for matching these pupils; this method is better suited for matching when the treated group is small than methods based on propensity scores. In this case, each treated pupil was paired with a comparison pupil in the same stratum.

The graphs shown in figure 4, known as love plots, show how similar IU pupils and comparison pupils were to one another before and after matching. The plots use a measure known as *standardised mean difference* (SMD). The mean difference is simply the difference between the average value for the IU pupils and the average value for the comparison pupils. Standardising this measure means that we can compare balance across different variables. Generally, a standardised mean difference of 0.2 or less is considered to indicate good balance. This threshold is indicated on the plots as a dotted line.

The love plots indicate good balance between the IU pupils and the matched comparison group. The vast majority of variables are matched to well within the 0.2 threshold. The matches are slightly less strong for pupils who completed KS2 in 2016, with two variables slightly above the 0.2 threshold.

Figure 4: Loveplots showing balance before and after matching




2018
NA
Prior attainment: reception
Prior attainment: KS1
Female
FSM6
IDACI
SEN (no EHCP)
EAL
Bangladeshi
Asian other
Pakistani
Black African
Black Caribbean
Black other
Chinese
Mixed other
Mixed white and Asian
Mixed white and black African
Ethnicity not recorded
Other ethnic group
White British
White other
Birth month: January
Birth month: February
Birth month: March
Birth month: April
Birth month: May
Birth month: June
Birth month: July
Birth month: August
Birth month: September
Birth month: October
Birth month: November
Birth month: December
School: Attainment, 2015
School: Attainment, 2016
School: Attainment, 2017
School: % FSM6
Region: E. England
Region: E. Midlands
Region: London
Region: NW
Region: SE
Region: SW
Region: Yorkshire and Humber
-0.5
0.0
0.5
1.0
1.5
SMD
2019
Prior attainment: reception
Prior attainment: KS1
Female
FSM6
IDACI
SEN (no EHCP)
SEN (EHCP)
EAL
Bangladeshi
Indian
Asian other
Pakistani
Black African
Black Caribbean
Black other
Chinese
Mixed other
Mixed white and Asian
Mixed white and black African
Ethnicity not recorded
Other ethnic group
White British
White other
Birth month: January
Birth month: February
Birth month: March
Birth month: April
Birth month: May
Birth month: June
Birth month: July
Birth month: August
Birth month: September
Birth month: October
Birth month: November
Birth month: December
School: Attainment, 2016
School: Attainment, 2017
School: Attainment, 2018
School: % FSM6
Region: E. England
Region: E. Midlands
Region: London
Region: NW
Region: SE
Region: SW
Region: W. Midlands
Region: Yorkshire and Humber
-0.5
0.0
0.5
1.0
1.5
SMD
After matching
Before matching

4. Results

4.1 Format

Results are given in three different forms: estimated impact, effect size, and months of progress.

Estimated impact is given in the same units as the outcome measure. In this report, there are two outcome measures: KS2 scaled score in maths, and KS2 scaled score in reading. In both cases, an estimated impact of one would mean that we'd expect an IntoUniversity pupil to achieve one point higher than a non-IU pupil.

However, when using estimated impact it is difficult to compare across different outcome measures. It's not necessarily the case that an estimated impact of 0.75, for example, on KS2 scaled scores in maths is the equivalent of an estimated impact of 0.75 on KS2 scaled scores in reading; having an impact on reading scores may be more challenging than having an impact on maths scores, for example. It is also difficult to compare the effect of IntoUniversity to the effect of another project that focuses on a different outcome measure using estimated impact.

The effect size is used to get around this problem. It is a standardised version of the estimated impact. That is, it is the estimated impact divided by the standard deviation in the outcome measure among all pupils entered for a particular subject. Because it is a standardised measure, it can be compared across different outcomes.

However, effect sizes can be difficult to interpret; it is not immediately obvious whether an effect size of, for example, 0.5 is large or small. Months of progress are a measure used in education research to try and help with this. In this report, effect sizes were translated into equivalent months of progress using guidance developed by the Education Endowment Foundation[3], as shown in table 1. In our example, an effect size of 0.5 would be the equivalent of six months of additional progress; expressed using the months of progress measure, it is clear that this is a large effect.

Table 2: Effect sizes and equivalent months of progress

Effect size from	To	Months of progress
-0.04	0.04	0
0.05	0.09	1
0.10	0.18	2
0.19	0.26	3
0.27	0.35	4
0.36	0.44	5
0.45	0.52	6
0.53	0.61	7

[3] As described at https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/evaluator-resources/writing-a-research-report, accessed January 2020

0.62	0.69	8
0.70	0.78	9
0.79	0.87	10
0.88	0.95	11

4.2 Pooling

Results are reported separately for each outcome year, and we have also included a pooled estimate for each outcome. The pooled estimates are simply a weighted average of the estimates for each individual year.[4] They provide an estimate of the overall effect, based on data from all of the outcome years. This increases the sample size, meaning that pooled estimates will tend to be more precise – that is, have narrower confidence intervals – than estimates for individual years.

This is particularly useful where some sample sizes are small. In this evaluation, the sample size for 2019 is far larger than for other years. This may mean that estimates for this year are more reliable, and more precise, than for other years. Some of the results from earlier years have very wide confidence intervals, and in some cases are inconsistent with those from 2019. In these cases, the pooled estimates may be more helpful than the results from individual years.

[4] Estimates were weighted with the inverse variance of the bootstrapped estimates. More information on the methodology we used can be found in Borenstein *et al* (2009).

4.3 Overall

4.3.1 Maths

Estimates of the impact of IntoUniversity academic support on Key Stage 2 scaled scores in maths are shown in table 3 below, and summarised in figure 5. All numbers in table 3 have been rounded to two decimal places.

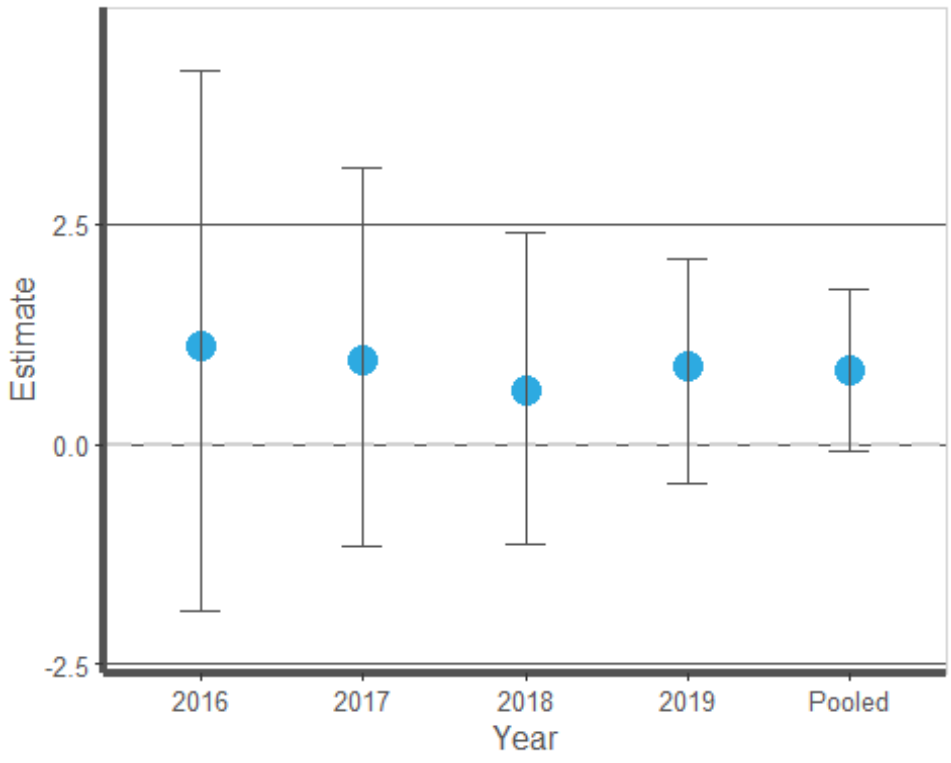
These results do not provide conclusive evidence that IntoUniversity support has a positive impact on attainment in Key Stage 2 maths. Although the estimates for each year, and the pooled estimate, are positive, the equivalent of between one and two months of progress, none of the results are statistically significant; that is, all of the confidence intervals contain zero. This means that we cannot be confident that IU support has any effect.

As shown in figure 5, the confidence intervals for the estimates are narrower for later years. This is because of the increased sample size; more pupils were included in the evaluation in later years.

Table 3: Estimated impact on maths scaled scores

Year	Sample size	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2016	31	-1.89	1.11	4.24	0.14	2
2017	69	-1.16	0.96	3.15	0.11	2
2018	95	-1.13	0.60	2.41	0.07	1
2019	197	-0.45	0.88	2.11	0.10	2
Pooled	392	-0.07	0.85	1.76	0.10	2

Figure 5: Estimated impact on maths scaled scores



4.3.2 Reading

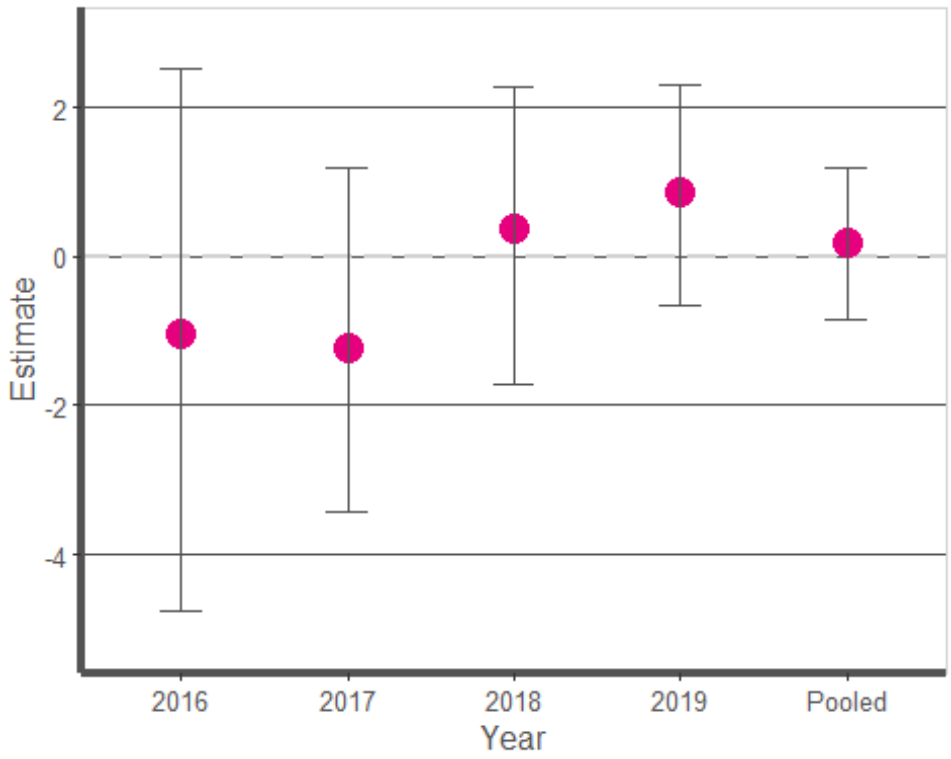
Estimates of the impact of IntoUniversity academic support on Key Stage 2 scaled scores in reading are shown in table 4 below, and summarised in figure 6. All numbers in table 4 have been rounded to two decimal places.

These results do not provide conclusive evidence that IntoUniversity support has a positive impact on attainment in Key Stage 2 reading. For both 2016 and 2017, the estimate is actually negative. This means that we would expect a pupil receiving academic support for IntoUniversity to achieve a lower scaled score in reading than a matched comparison pupil. However, the confidence intervals for both estimates are wide and neither are statistically significant. The estimates for 2018 and 2019, which had larger sample sizes, are both positive, and the estimate for 2019 is very similar to the estimate for maths. Likewise, the pooled estimate is positive, although it isn't large enough to be the equivalent of any additional months of progress. However, the confidence intervals for all estimates contain zero; they are not statistically significant and we cannot be confident that there is any effect.

Table 4: Estimated impact on reading scaled scores

Year	Sample size	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2016	31	-4.77	-1.03	2.52	-0.11	<0
2017	69	-3.42	-1.23	1.21	-0.13	<0
2018	95	-1.70	0.37	2.28	0.04	0
2019	197	-0.66	0.87	2.30	0.09	1
Pooled	392	-0.83	0.18	1.20	0.02	0

Figure 6: Estimated impact on reading scaled scores



4.4 Dosage

4.4.1 Maths

Estimates of the impact of IntoUniversity academic support on Key Stage 2 scaled scores in maths by dosage are shown in table 5 below, and summarised in figure 7. All numbers in table 5 have been rounded to two decimal places.

Due to low numbers, no estimates are provided for pupils who completed Key Stage 2 in 2016.

These results provide some evidence that the impact of IntoUniversity support on attainment in Key Stage 2 maths increases as dosage increases. In 2019, this pattern is clear, and the estimated impact for the high dosage group is both positive and significant. We would estimate that a high dosage IntoUniversity pupil who completed KS2 in 2019 would achieve a scaled score of 2.48 more than a matched comparison pupil. This is the equivalent of four months of additional progress.

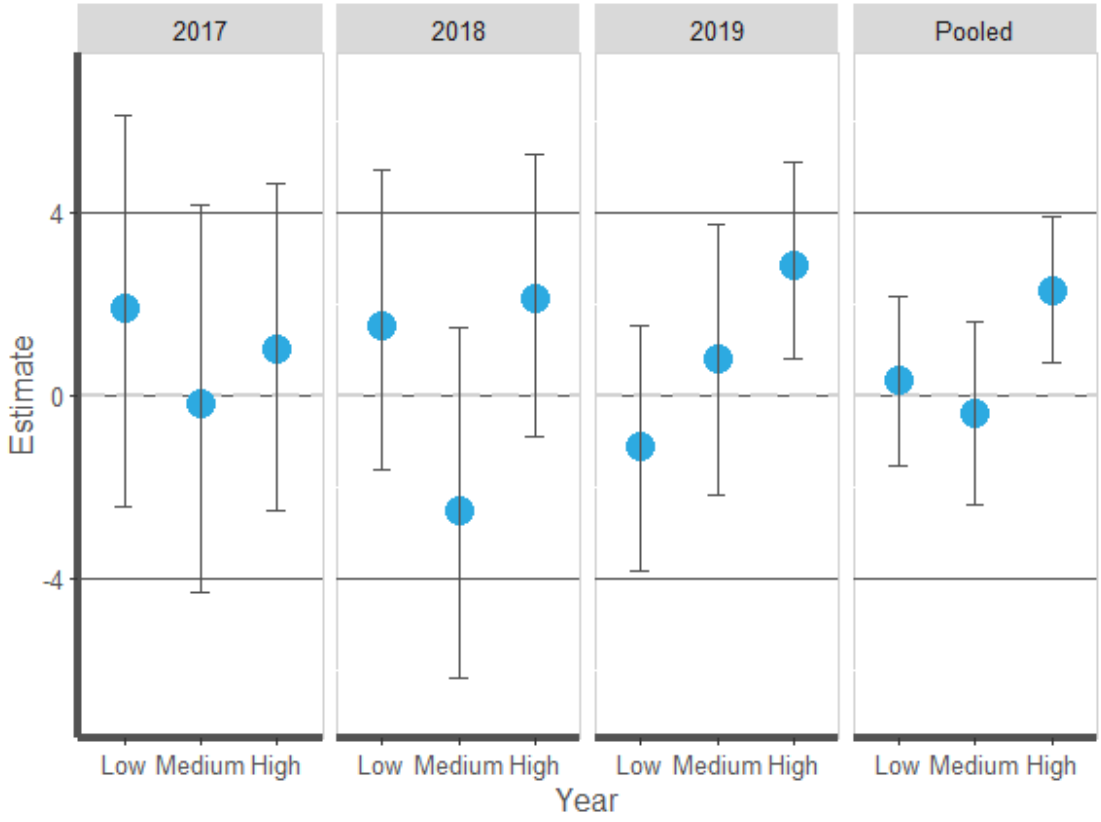
However, the results for 2017 and 2018 do not follow the same pattern, and none of the estimates are significant. This may be a reflection of the lower sample size.

The pooled estimates show a positive and significant effect on Key Stage 2 maths for the high dosage group. Based on the pooled results, we would estimate that a high dosage IntoUniversity pupil would achieve a scaled score of 2.29 more than a matched comparison pupil. This is the equivalent of 3 months of additional progress.

Table 5: Estimated impact on maths scaled scores by dosage

Year	Sample size	Dosage	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2017	25	High	-2.54	1.00	4.64	0.11	2
2017	25	Medium	-4.33	-0.17	4.17	-0.02	0
2017	19	Low	-2.45	1.90	6.13	0.21	3
2018	30	High	-0.89	2.11	5.25	0.24	3
2018	30	Medium	-6.18	-2.52	1.50	-0.29	<0
2018	35	Low	-1.62	1.54	4.94	0.18	2
2019	73	High	0.78	2.84	5.08	0.32	4
2019	56	Medium	-2.19	0.79	3.72	0.09	1
2019	68	Low	-3.83	-1.13	1.51	-0.13	<0
Pooled	128	High	0.69	2.29	3.89	0.26	3
Pooled	111	Medium	-2.41	-0.41	1.59	-0.05	<0
Pooled	122	Low	-1.53	0.32	2.17	0.04	0

Figure 7: Estimated impact on maths scaled scores by dosage



4.4.2 Reading

Estimates of the impact of IntoUniversity academic support on Key Stage 2 scaled scores in reading by dosage are shown in table 6 below, and summarised in figure 8. All numbers in table 6 have been rounded to two decimal places.

Due to low numbers, no estimates are provided for pupils who completed Key Stage 2 in 2016.

As for maths, these results provide some evidence that the impact of IntoUniversity support on attainment in Key Stage 2 reading increases as dosage increases. The pattern is clear in 2019; the estimated impact for medium dosage pupils is higher than that for short term pupils, and for long term pupils it is higher again. However, this pattern is not seen in the results for 2017 or 2018. None of the results for any dosage group are significant in any year.

The pooled estimates do show a higher estimated impact for high dosage pupils. However, as the confidence interval for this estimate contains zero, it is not statistically significant. This means that we cannot be confident that the programme had any effect on this outcome.

Table 6: Estimated impact on reading scaled scores by dosage

Year	Sample size	Dosage	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2017	25	High	-6.00	-2.00	2.00	-0.21	<0
2017	25	Medium	-6.03	-1.82	2.24	-0.19	<0
2017	19	Low	-4.99	-0.08	4.62	-0.01	0
2018	30	High	-2.36	1.07	4.66	0.11	2
2018	30	Medium	-7.59	-3.15	0.72	-0.34	<0
2018	35	Low	-0.87	2.30	6.01	0.25	3
2019	73	High	0.00	2.48	5.04	0.25	3
2019	56	Medium	-2.28	0.69	3.58	0.07	1
2019	68	Low	-3.73	-0.70	2.03	-0.07	<0
Pooled	128	High	-0.62	1.20	3.02	0.12	2
Pooled	111	Medium	-2.95	-0.84	1.28	-0.09	<0
Pooled	122	Low	-1.64	0.35	2.33	0.04	0

Figure 8: Estimated impact on reading scaled scores by dosage

4.5 Length of time involved with project

4.5.1 Maths

Estimates of the impact of IntoUniversity academic support on Key Stage 2 scaled scores in maths by length of time involved with the programme are shown in table 7 below, and summarised in figure 9. All numbers in table 7 have been rounded to two decimal places.

Due to low numbers, no estimates are provided for pupils who completed Key Stage 2 in 2016.

These results do not provide evidence that IntoUniversity support on attainment in Key Stage 2 maths increases as the length of time involved with the programme increases. In 2019, the impact on short term pupils is lower than that of other groups, but the impact on medium term pupils is higher than that on long term pupils. It is also significant, while that for the long term group is not. We would estimate that a medium term IntoUniversity pupil who completed KS2 in 2019 would achieve a scaled score of 3.33 more than a matched comparison pupil, the equivalent of five months of additional progress.

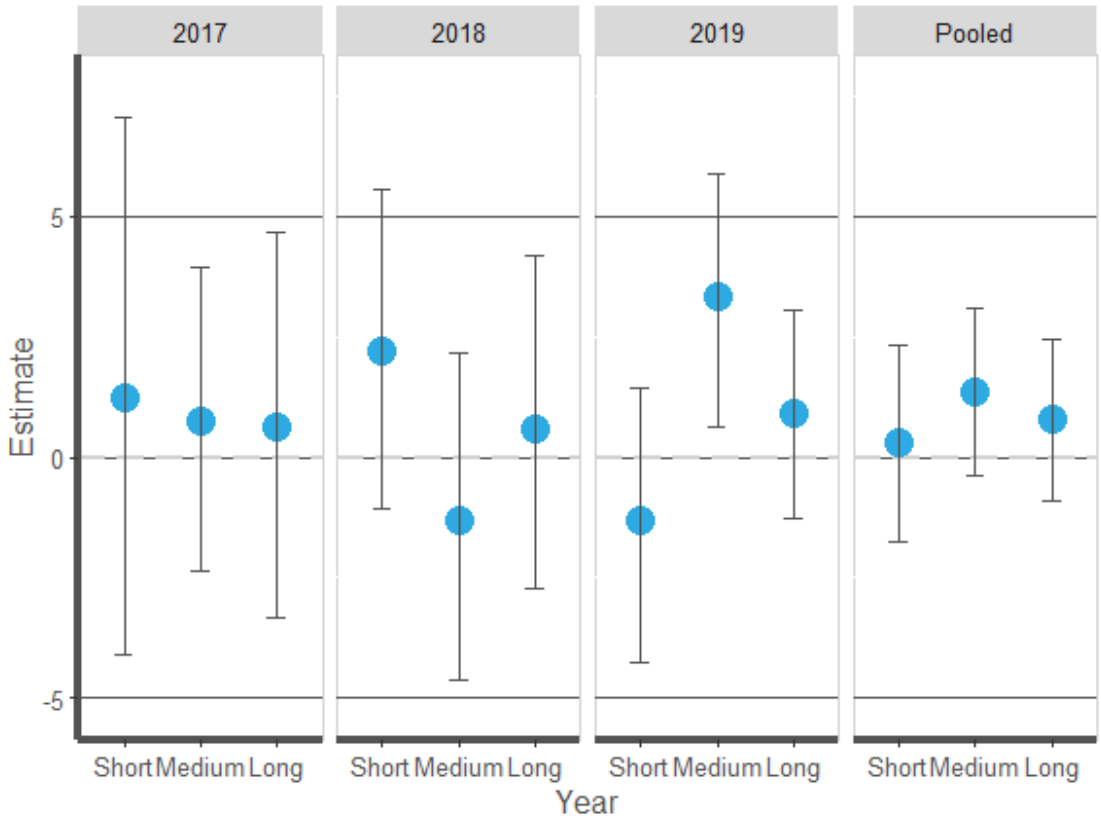
However, the results for 2017 and 2018 do not show any increased effect for medium term pupils, and none of the estimates are significant.

The pooled estimates do show a higher estimate for the medium term group of pupils, but none of the estimates are significant, meaning that we cannot be confident that the programme had any effect on this outcome in those years.

Table 7: Estimated impact on maths scaled scores by length of time involved

Year	Sample size	Time	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2017	23	Long	-3.33	0.61	4.68	0.07	1
2017	33	Medium	-2.36	0.75	3.97	0.08	1
2017	13	Short	-4.11	1.24	7.07	0.14	2
2018	32	Long	-2.73	0.59	4.18	0.07	1
2018	32	Medium	-4.61	-1.30	2.15	-0.15	<0
2018	31	Short	-1.06	2.22	5.58	0.25	3
2019	81	Long	-1.26	0.90	3.04	0.10	2
2019	55	Medium	0.65	3.33	5.91	0.38	5
2019	61	Short	-4.27	-1.33	1.43	-0.15	<0
Pooled	136	Long	-0.91	0.78	2.46	0.09	1
Pooled	120	Medium	-0.38	1.36	3.09	0.15	2
Pooled	105	Short	-1.74	0.29	2.33	0.03	0

Figure 9: Estimated impact on maths scaled scores by length of time involved



4.5.2 Reading

Estimates of the impact of IntoUniversity academic support on Key Stage 2 scaled scores in reading by length of time involved with the programme are shown in table 8 below, and summarised in figure 10. All numbers in table 8 have been rounded to two decimal places.

Due to low numbers, no estimates are provided for pupils who completed Key Stage 2 in 2016.

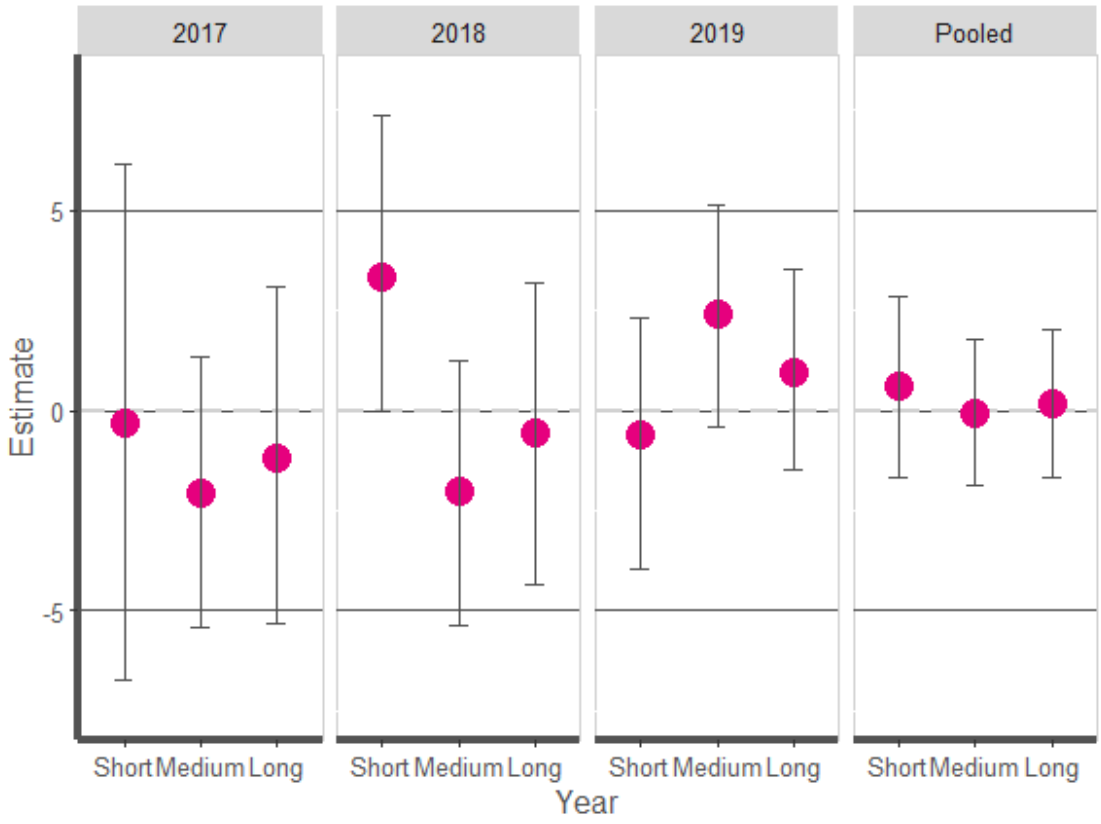
As for maths, these results do not provide evidence that IntoUniversity support on attainment in Key Stage 2 reading increases as the length of time involved with the programme increases. Again as for maths, the impact on short term pupils in 2019 is lower than that of other groups, but the impact on medium term pupils is higher than that on long term pupils. Unlike maths, none of the results here are significant.

The results for 2017 and 2018 do not show any increased effect for medium term pupils, or any evidence of an increased effect for longer term involvement. This is also the case when we look at the pooled estimates. None of the estimates for 2017 and 2018, or the pooled estimates, are significant, meaning that we cannot be confident that the programme had any effect on this outcome.

Table 8: Estimated impact on reading scaled scores by length of time involved

Year	Sample size	Time	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2017	23	Long	-5.33	-1.20	3.08	-0.12	<0
2017	33	Medium	-5.43	-2.06	1.36	-0.21	<0
2017	13	Short	-6.73	-0.32	6.14	-0.03	0
2018	32	Long	-4.36	-0.56	3.17	-0.06	<0
2018	32	Medium	-5.38	-1.99	1.27	-0.21	<0
2018	31	Short	0.00	3.35	7.35	0.36	5
2019	81	Long	-1.49	0.96	3.55	0.10	2
2019	55	Medium	-0.43	2.42	5.11	0.25	3
2019	61	Short	-3.95	-0.62	2.30	-0.06	<0
Pooled	136	Long	-1.70	0.15	2.00	0.02	0
Pooled	120	Medium	-1.89	-0.05	1.79	-0.01	0
Pooled	105	Short	-1.66	0.60	2.85	0.06	1

Figure 10: Estimated impact on reading scaled scores by length of time involved



5. Conclusions

5.1 Limitations

Many of the limitations of this piece of research arise from the fact that the matched comparison group was created using data from the National Pupil Database (NPD). While the NPD is an excellent resource for education research, it is necessarily limited in scope; for example, it doesn't contain data on parental occupation, social class or school funding levels. Not accounting for these unobserved variables may introduce bias into our estimates.

We also have no way of knowing whether pupils in the matched comparison group received any support similar to that offered by IntoUniversity from other sources. If so, this may have led to underestimation of the impact, assuming that the alternative support had a positive impact on attainment. We would note, however, that this may be the best approach, representing an evaluation of IntoUniversity support against current conditions, which may include pupils' decision to engage with alternative support.

The sample sizes for the earlier years, particularly 2016, were relatively low. With lower sample sizes, we are less likely to be able to identify significant effects. For 2016, the sample size was too low to allow us to provide any estimates on how the effect of the project varied by dosage of length of time involved.

It is also possible that selection bias has played a part. It may be the case that pupils who choose to engage with IntoUniversity are those who are seeking extra support because they are falling behind in school; these pupils might be expected to achieve lower scores at KS2 than their prior attainment at KS1 and foundation stage might suggest. If this is the case, it could have led to underestimation of impact, because they might have been matched to comparison pupils who were not falling behind. On the other hand, it may be that pupils who choose to engage with IntoUniversity are those who are highly motivated. If this is the case, their level of motivation may be at least partly responsible for any improved results, rather than support received from IntoUniversity. This could lead to overestimation of results.

Similarly, bias may arise from the way that pupils were selected for inclusion in this evaluation. Only those IU pupils who had given consent for their data to be shared were eligible for inclusion. This type of consent was requested from pupils for the first time in 2019; unsurprisingly, more pupils from the 2019 cohort responded than pupils from earlier cohorts. Pupils from earlier cohorts who gave consent were primarily those who remained involved with the project long term. This longer term involvement may indicate that those pupils were more in need of support than the average IU pupil; if this is the case, it could have led to an underestimation of the impact of the project. On the other hand, it may be the case that pupils who gave consent were those who were most committed to and engaged in the project. This could have led to overestimation of the impact, assuming that the project had a greater impact on those pupils who were most engaged.

5.2 Discussion

This evaluation does not provide conclusive evidence that IntoUniversity academic support has a positive impact on scaled scores in Key Stage 2 maths and reading. The estimates of impact on maths scaled scores, while positive, were not significant. Estimates of impact on reading scaled scores were likewise not significant, and were lower than those for maths; in 2016 and

2017 they were actually negative, suggesting that IntoUniversity pupils achieved scaled scores slightly lower than those of matched comparison pupils.

However, the evaluation does provide some evidence that the impact of IU support on attainment increases as dosage increases, and that there is a significant positive impact on maths scaled scores for pupils who attend a high number of sessions (80 or more). This can be seen in the estimates for pupils who completed KS2 in 2019 - the year with by far the largest sample size. We would estimate that a high dosage IntoUniversity pupil who completed KS2 in 2019 would achieve a scaled score of 2.84 more than a matched comparison pupil. This is the equivalent of four months of additional progress. However, the estimates for 2017 and 2018 do not fit this pattern, and do not include any significant estimates. Pooling the estimates across all years gives us an estimated impact that is slightly lower than that for 2019 alone, but that is still both significant and positive. Based on the pooled estimate, we would estimate that a high dosage IntoUniversity pupil would achieve a scaled score of 2.29 more than a matched comparison pupil. This is the equivalent of 3 months of additional progress.

Our results do not provide consistent evidence of an increase in impact for pupils who have been involved in the programme for a longer period of time. For pupils who completed KS2 in 2019, the estimated impact for short term pupils is lowest, but that for medium term pupils is higher than that for long term pupils. In fact, the impact on medium term pupils for maths scaled scores is significant; we would estimate that a medium term IntoUniversity pupil would achieve a scaled score of 3.33 points higher than a matched comparison pupil, the equivalent of five additional months of progress. However, no similar increased impact on medium term pupils was seen in 2017 or 2018; in 2018, the estimated effect on both maths and reading scaled scores was actually lower for medium term pupils than for either short or long term pupils. The pooled estimates did not find a significant impact for either short, medium or long term pupils, and did not provide any evidence that the impact increases with length of time involved.

As the estimates for 2019 were based on a far larger sample size than for other years, it may be the case that the estimates for this year are more reliable than those for other years. This is particularly likely for the estimates of effect by dosage or time involved with the project, where the sample has been broken down into smaller subgroups, and may go some way to explaining the inconsistent year-on-year results for these parts of the evaluation. The pooled estimates provide a more reliable estimate that incorporates the findings from all outcome years.

6. Technical appendix: Sensitivity analysis

Ideally, in an evaluation like this we would use a *doubly robust approach*; that is, we would control for the matching variables as part of the regression model. This approach will yield unbiased estimates provided at least one of the models used for matching or regression is correct.

However, a number of treated pupils in this evaluation were missing data for one or more of the matching variables, as described in section 3.2. These pupils were matched to other pupils who were also missing data for the relevant variables. We were therefore unable to use the matching variables as controls during regression; to do so would have required us to exclude those pupils with missing data. We were unwilling to do so as sample sizes were already relatively small. It would also have meant excluding any pupils who did not complete both foundation stage and KS1 at an English school, potentially skewing the results.

In order to get an indication of whether failing to apply the doubly robust method affected our conclusions, we carried out a sensitivity analysis. Using only those pupils for whom we had complete data for all matching variables, we estimated impact with and without controls applied at the regression stage. The results of this analysis are shown below.

Although there are some differences in the estimates, they are generally very similar to one another; in all but one case, the estimated months of progress is the same, and the significance matches. The only case where months of progress differs is the estimated impact on maths in 2016; the sample size on which this estimate was based is small and the confidence interval wide, so more variability in the point estimate is to be expected and not a cause for much concern.

We would conclude that failing to use the doubly robust method is unlikely to have biased our estimates in this case.

6.1 Maths

Table 9: Estimated impact on maths scaled scores by controls type

Year	Sample size	Type	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2016	27	Controls	-3.62	0.12	4.38	0.01	0
2016	27	No controls	-2.74	0.50	4.04	0.06	1
2017	56	Controls	-1.62	0.65	2.73	0.07	1
2017	56	No controls	-1.84	0.69	3.20	0.08	1
2018	82	Controls	-1.48	0.26	1.91	0.03	0
2018	82	No controls	-1.73	0.31	2.34	0.04	0
2019	171	Controls	-0.60	0.54	1.66	0.06	1
2019	171	No controls	-0.85	0.59	1.93	0.07	1

6.2 Reading

Table 10: Estimated impact on maths scaled scores by controls type

Year	Sample size	Type	Lower CI	Estimate	Upper CI	Effect size	Months of progress
2016	27	Controls	-6.90	-1.81	3.25	-0.19	<0
2016	27	No controls	-5.70	-1.58	2.26	-0.17	<0
2017	56	Controls	-4.02	-1.58	0.89	-0.16	<0
2017	56	No controls	-3.77	-1.28	1.50	-0.13	<0
2018	82	Controls	-2.28	-0.16	1.88	-0.02	0
2018	82	No controls	-2.45	-0.13	1.94	-0.01	0
2019	171	Controls	-0.83	0.63	1.98	0.06	1
2019	171	No controls	-1.04	0.69	2.32	0.07	1