



Department
for Education

An analysis of the performance of sponsored academies

Analytical research report
January 2019

Adam Hatton, Rebekah Hampson and
Rob Drake – Department for Education



Government
Statistical Service

Contents

Executive Summary	4
Secondary sponsored academies	4
Primary sponsored academies	4
Introduction	6
Peer review	7
Ofsted inspection frameworks	7
The context of key stage 2 (KS2) and key stage 4 (KS4) accountability reforms	7
Part 1: Performance of sponsored academies within the context of the national average performance	9
Sponsored academies and their operating contexts	9
Making comparisons with local authority schools	13
Accounting for how long sponsored academies have been open	15
Making comparisons over a number of years	16
Comparison with national average results	17
Summary of part 1	19
Part 2: Comparing performance of sponsored academies with similar schools using propensity score matching (PSM)	20
Making fairer comparisons using PSM	20
How to interpret the charts in this part of the report	20
Detailed results: Secondary sponsored academies (KS4)	22
Secondary sponsored academies with first academy results in 2011	22
Secondary sponsored academies with first academy results in 2012	23
Secondary sponsored academies with first academy results in 2013	24
Secondary sponsored academies with first academy results in 2014	25
Secondary sponsored academies with first academy results in 2015	26
Secondary sponsored academies with first academy results in 2016	27
Secondary sponsored academies with first academy results in 2017	28
Secondary sponsored academies with first academy results in 2018	29
Detailed results: Primary sponsored academies (KS2)	30
Primary sponsored academies with first academy results in 2013	30

Primary sponsored academies with first academy results in 2014	31
Primary sponsored academies with first academy results in 2015	32
Primary sponsored academies with first academy results in 2016	33
Primary sponsored academies with first academy results in 2017	34
Primary sponsored academies with first academy results in 2018	35
Conclusions	36
Appendices	37
Data and methodology	37
Propensity Score Matching (PSM)	37
Detailed results: choosing a model	39
Choosing how many iterations	40
Percentage of sponsored academies matched	41
Statistics tests	42
Glossary	46

Executive Summary

This report summarises the latest findings on the performance of sponsored academies up to and including 2018 results. It explains our current understanding of results by examining their historical context, their performance in comparison with national average performance and in comparison to matched similar schools. The complex evolving context in which schools have become sponsored academies since 2010 (involving changes to policy, key stage tests, accountability measures and the changes in the composition of the population of sponsored academies and remaining LA state funded schools) means the results for each annual cohort of sponsored academies need to be considered separately. A single over-time comparison between sponsored academies and similar LA schools would not be meaningful.

Secondary sponsored academies

- Secondary sponsored academies that have been open longer have improved, on average, so that they have results closer to the national average for all state-funded mainstream secondary schools. Secondary sponsored academies that opened prior to 2010, on average, perform most similarly to the national average.
- Our analysis shows that schools that became sponsored academies typically performed significantly less well than otherwise similar schools prior to academisation. Some of the groups of schools that became sponsored academies were on downward trajectories of performance prior to academisation.
- After academisation most groups of sponsored academies typically demonstrate improvement to a point where their performance is often indistinguishable from their similar schools and sometimes significantly better than them.
- For some groups of sponsored academies changes in performance between years were not always improvements. For some groups performance year-on-year is volatile and oscillates around that of similar schools. In other groups there is evidence to suggest year-on-year improvement has arrested or been reversed.
- There is substantial variation in performance and there are some schools with results well below national average performance, even in the groups that have been academies for more than seven years. In contrast, some sponsored academies perform well above the national average.

Primary sponsored academies

- As with secondary schools, it is the schools open for longer that tend to have performance which improved to be more similar to the national average.

- As with secondary sponsored academies, primary schools that became sponsored academies were typically performing significantly less well than similar schools prior to academisation. Some were on downward trajectories, others showing relatively little sign of improvement.
- Most of the groups of primary sponsored academies show signs of narrowing the gap with their similar schools after academisation; one group is now performing significantly better than its similar schools.
- Some groups are continuing to perform significantly less well than their similar schools.

Introduction

Sponsored academies generally replaced under-performing local authority schools, many with a history of low attainment. This was not the case for all sponsored academies; some former independent schools used academisation to join the state-sector. Other sponsored academies were entirely new schools built in response to a need for school places (known as “academy presumption” schools).

Converter academies have also been created which typically replaced high performing local authority schools. This means that the remaining local authority school population was increasingly unlike that of sponsored academies or converter academies. So, a direct comparison of the performance of each of these types of school is misleading.

Some researchers do not clearly describe the distinction between sponsored academies and converter academies. Making this distinction clear is important as these schools have often had very different histories and have different objectives they wish to achieve by becoming an academy. This brings additional complexity to understanding the impact of the programme. This report focuses specifically on those schools that have become sponsored academies.

We group sponsored academies into “cohorts” based on the results of the first full academic year that they were an academy. We might expect that the full impact of becoming an academy is more evident over a number of years and the method used allows us to examine that.

Part 1 of this report examines the historical context of sponsored academies and compares the performance of their different “cohorts” to the national average performance for state-funded schools.

To make fair comparisons, part 2 of this report uses an analytical technique called propensity score matching (PSM) which is explained in the appendix to this report. The Department has conducted similar analysis in the past, including a publication in June 2012¹. Since then, the academies programme has expanded dramatically, and we now have more data over a longer period from which to draw conclusions. This part of the report examines the performance of sponsored academies in comparison with groups of other similar schools over time, up to and including the provisional 2018 key stage 2 and 4 data².

¹ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/184062/DFE-RR223.pdf

² Provisional data means it has been quality assured by the department, but the underlying data has yet to be checked by schools. The statistics in this release are based on the results data that awarding organisations supply to the department. This includes the vast majority of pupils' results; however, it will not take account of any amendment requests made by schools during the September checking exercise.

Peer review

An earlier version of this report was reviewed by Ian Dadswell of Oxford Analytics. The authors of this report are grateful for his advice. As a result, we have brought into the final report more on the background context of the academies programme and an examination of academy performance against national average performance. We've also added more to convey to the reader the confidence in our estimates of the performance difference between sponsored academies and the matched similar schools. That said, methodological decisions, the presentation and interpretations of the findings in the report remain those of the authors.

Ofsted inspection frameworks

This report includes some analysis using Ofsted inspection outcomes. The authors recognise that the Ofsted inspection framework has changed over the period of the analysis. In particular, in September 2012 the “satisfactory” category was replaced with the term “requires improvement”.

The context of key stage 2 (KS2) and key stage 4 (KS4) accountability reforms

Due to changes in accountability measures, we have focussed on the most consistent and similar measures across all the years studied. For KS4, this was the percentage of pupils achieving GCSE English and mathematics grades A*-C (or 9-4 in 2017 onwards). For KS2, this was the percentage of pupils reaching the expected standard in English (or reading and writing) and mathematics. However, the number of changes that have occurred mean there is no one ideal approach.

History of changes in accountability measures

A short history of changes in primary accountability measures:

There were changes to the headline measures at KS2 within the period covered in this report. Between 2008 and 2012, pupils were measured at KS2 against an expectation of achieving level 4 or above in English and maths. From 2013, this changed to those achieving level 4 or above in reading, writing and maths individually. From 2016, this changed again to reflect those reaching a new, more challenging, expected standard in reading, writing and maths.

Effect of primary test boycott in 2010:

Interpreting the 2010 KS2 results requires some caution as a number of the schools boycotted national school tests in 2010. We have made no specific adjustments to our results because of the boycott.

A short history of changes in secondary accountability measures:

At KS4, there were accountability reforms, which affect the interpretation of this analysis, particularly those occurring in 2014. These include the Wolf Review, which led to restrictions on the inclusion of “equivalent” vocational qualifications within the performance tables. At the same time, there were further changes to discourage early and repeated entry of pupils for examinations. Such changes included using pupils’ first exam results (which may not have been their best result) in performance measures. From then onwards, performance measures excluded a number of non-GCSE equivalent qualifications and those that remained were restricted in their size to the equivalent of no more than two GCSEs. These changes had a disproportionate negative effect on measured performance of sponsored academies, and so they performed less well after the changes. From 2016, new headline measures, including progress 8, were introduced. In 2017, pupils sat reformed GCSEs in English language, English literature and mathematics for the first time, graded on a 9-1 scale.

These changes have led us to focus solely on measuring achievement based on the percentage of pupils achieving a “good” GCSE in both English and mathematics. We have defined “good” as A*-C in years up to 2016 and as grades 9-4 in 2017 and 2018. Achievement at grades 9-4 is broadly similar to A*-C in earlier years.

Part 1: Performance of sponsored academies within the context of the national average performance

Sponsored academies and their operating contexts

As at November 2018 there were over 2,200 sponsored academies in England. The oldest of these opened in September 2002 and the newest on 1 November 2018.

A school may become a sponsored academy at any point during the year. The Department for Education's performance tables allocate the first set of performance data to the academic year that covers the first year in which they were open on 12 September³. This means that new trusts or sponsors may not always have had the same amount of time to influence a sponsored academy's first set of outcomes in the performance tables. The first set of results officially allocated to an academy are typically those that come at the end of the academy's first full academic year of operation.

Most sponsored academies directly replaced an existing school. Some sponsored academies were formed to replace more than one existing school. Some sponsored academies were entirely new schools meeting need for school places in an area.

To be manageable, the analysis specifically examines those schools that directly replaced one existing school. Table 1 below highlights that, of almost 1800 sponsored academies with a single predecessor school, around two in three were community schools prior to becoming a sponsored academy. The next largest group were foundation schools, of which more than 200 have become sponsored academies. The table also shows that a small number of city technology colleges and independent schools have become sponsored academies.

³ This has not always been the case. It is since 2011, but this analysis assumes that it has always been true to create groups on the same basis for consistency.

Table 1: Cohorts of sponsored academies and the school type of their predecessor schools

First year results would be included in performance tables (cohort)	Academy converter	Academy sponsored	City technology college	Community school	Community special school	Foundation school	Foundation special school	Other independent school	Pupil referral unit	Voluntary aided school	Voluntary controlled school	Total
2002/03										1		1
2003/04				4								4
2004/05				2								2
2005/06			2	2						2		6
2006/07			1	12		1				1		15
2007/08			4	13		4		2		2		25
2008/09			2	21		7		3			1	34
2009/10				32		5		1		2		40
2010/11				32		6		1		5		44
2011/12				23		10				3		36
2012/13				112	4	26				6	3	151
2013/14		2		245	3	32		1		20	24	327
2014/15				215	7	34	1		4	32	34	327
2015/16	1			142	6	28	2		9	39	20	247
2016/17		1		91	8	24	1		2	14	12	153
2017/18	1			149	9	26			5	32	21	243
2018/19				70	5	17			2	20	18	132
Total	2	3	9	1165	42	220	4	8	22	179	133	1787

Table 2 shows that in most of these cohorts (i.e. groups based on the year schools would first be expected to appear in school performance tables) of sponsored academies, the majority of their predecessor schools were rated less than good by Ofsted (based on inspections between September 2005 and August 2015). In three of the cohorts at least half of the sponsored academies were rated as inadequate prior to academisation. These ratings do not necessarily tell the full story as a predecessor school's last inspection outcome does not necessarily reflect the grading they might have received had they been inspected *immediately* prior to academisation. In some cases, these inspections took place a few years prior to academisation. Therefore, inspection outcomes may be under- or over- estimations of school performance immediately prior to academisation. There have been changes to the Ofsted inspection framework over this period, but we do not believe those changes invalidate the general conclusions drawn from examining them.

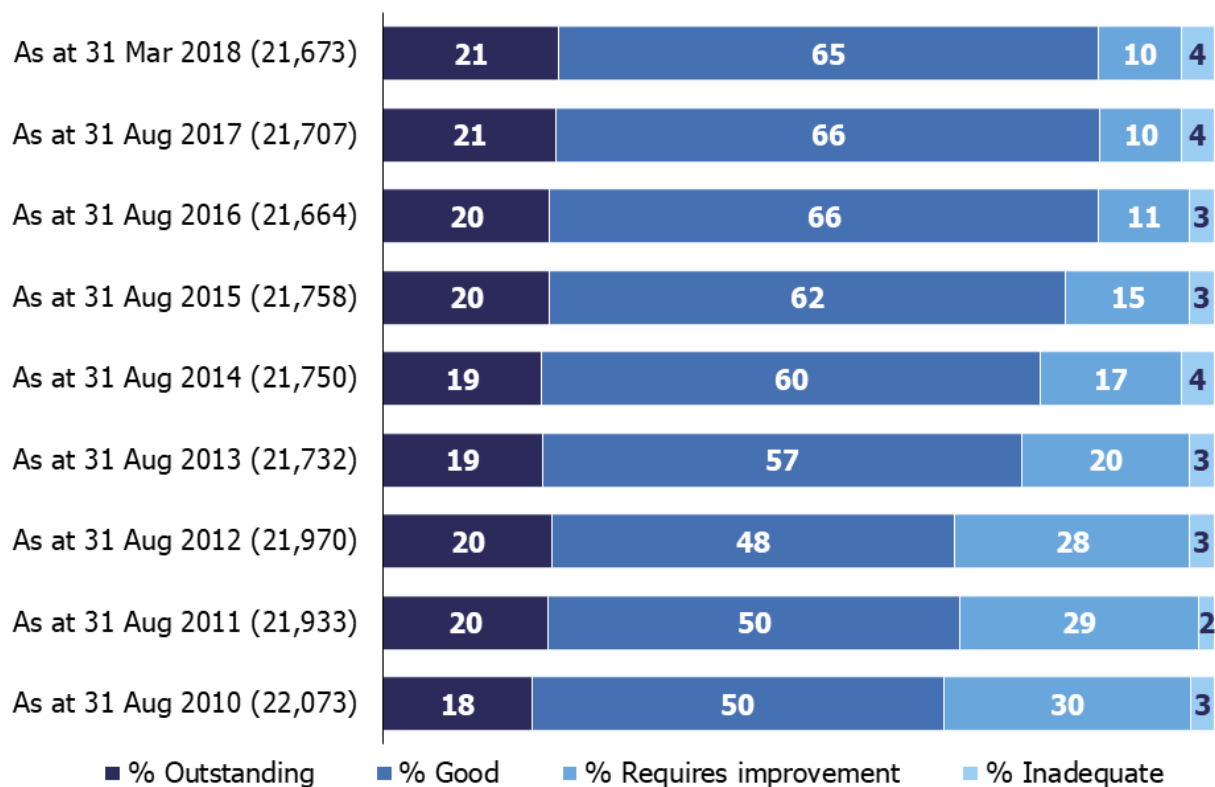
Table 2: Ofsted inspection outcomes for schools which preceded sponsored academies

First year results would be included in performance tables	% less than good	% inadequate	Number of schools
2006/07	100%	25%	4
2007/08	53%	27%	15
2008/09	76%	7%	29
2009/10	82%	10%	39
2010/11	77%	12%	43
2011/12	83%	11%	36
2012/13	89%	25%	150
2013/14	85%	39%	326
2014/15	92%	61%	327
2015/16	87%	54%	247
2016/17	86%	50%	153
2017/18	61%	16%	242
2018/19	44%	9%	131

A total of 646 of the 1,787 schools above that became sponsored academies were rated inadequate by Ofsted before becoming an academy.

Figure 1 shows that in 2010 around a third of schools overall were rated less than good; by March 2018 this has declined to less than one in six schools. This demonstrates that the schools that subsequently became sponsored academies were much more likely to be rated less than good (and more likely to be rated inadequate) than state-funded schools are now nationally.

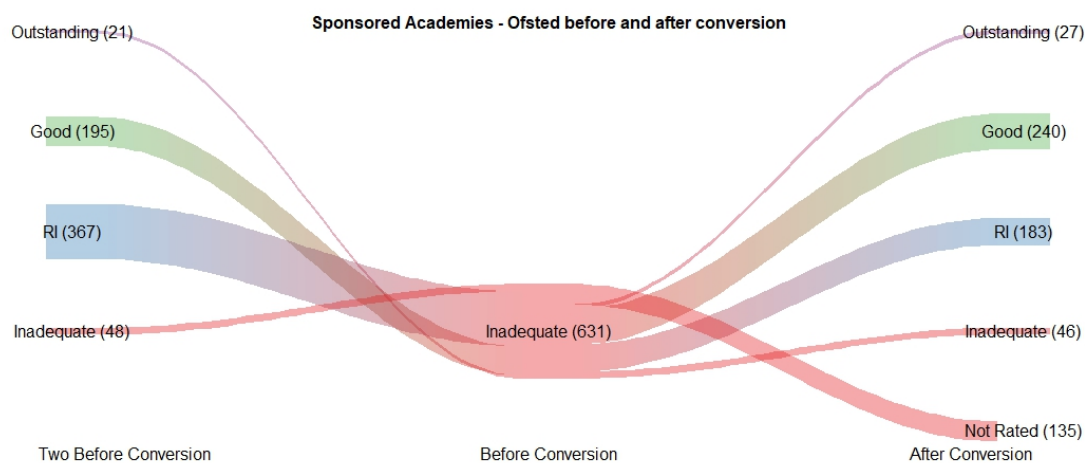
Figure 1: Most recent overall effectiveness grades of schools over time⁴



For the schools that were rated inadequate prior to becoming a sponsored academy, we can also use Ofsted data to look at the penultimate inspection before becoming an academy. These transitions are shown in figure 2. Two inspections prior to becoming a sponsored academy two-thirds were rated less than good. This gives some support to claims that many of the schools that were replaced by sponsored academies were performing poorly for several years.

⁴ <https://www.gov.uk/government/publications/state-funded-schools-inspections-and-outcomes-as-at-31-march-2018/state-funded-schools-inspections-and-outcomes-as-at-31-march-2018-main-findings> (Accessed: 10 August 2018)

Figure 2: Ofsted outcomes for inadequate schools that became sponsored academies prior to and following inadequate inspection outcome



Making comparisons with local authority schools

It is not necessarily appropriate to make direct comparisons between the performance of sponsored academies and the performance of schools that remained in the local authority sector. These two groups of schools differ across a number of characteristics.

Table 3 below exemplifies some similarities and differences by examining what happened (up to June 2018) to a set of local authority schools that were rated inadequate by Ofsted as of 31 August 2013⁵. This analysis encompasses 492 schools; the first observation is that those which remained LA schools number less than half of those that became sponsored academies. This means that the Academies Programme has assumed responsibility for many more of the inadequate schools than remained as local authority maintained schools.

The table shows that the group that became sponsored academies had some similarities, on average, to those that remained LA schools. In particular the percentage of pupils in major ethnic groups other than White and their GCSE examination outcomes. However, there were also some key differences. Those that became sponsored were more likely to have a greater percentage of pupils known to be eligible for, or claiming free school meals, so on this measure were more disadvantaged. Schools that became sponsored academies were more likely to have fewer pupils meeting the expected standard of

⁵ Historical data - <https://www.gov.uk/government/statistics/changes-to-ofsteds-statistical-reporting-of-inspection-outcomes-for-state-funded-schools-an-analysis-of-the-changes> (Accessed 20 July 2018)

attainment (at that time level 4) in reading, writing teacher assessment and maths at the end of primary school.

Table 3: Numbers and characteristics of inadequate local authority schools as at 31 Aug 2013

	Became converter	Became sponsored	Remained LA school	Closed at any time between Aug 13 and Jun 18
Number of schools	32	304	142	14
Percentage of pupils known to be eligible for, or claiming, free school meals	20%	24%	21%	26%
Percentage of pupils in major ethnic groups other than 'White'	21%	23%	23%	15%
Percentage of pupils achieving level 4 or above in reading, writing TA and maths	68%	60%	69%	66%
Percentage of pupils achieving good GCSEs including English and maths	52.7%	47.5%	48.2%	40.1%
Percentage of pupils achieving five good GCSEs including English and maths	51.6%	46.3%	47.4%	39.0%

This analysis supports decisions taken in other research, such as National Audit Office (2010)⁶, Department for Education (2012)⁷, Worth (2016)⁸ to use more complex analytical methods to construct fairer comparison groups. It also highlights that, on average, across this selection of factors we observe, to turn around the schools that became sponsored academies could be said to offer a greater degree of challenge. This is particularly the case with the primary schools, that on average have results 15 percentage points behind the national average at the end of key stage 2⁹. It also

⁶ <https://www.nao.org.uk/wp-content/uploads/2010/09/1011288.pdf> (Accessed 20 July 2018)

⁷ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/527043/Attainment_by_pupils_in_academies_2012_supplementary_analysis_to_the_academies_annual_report_2011_to_2012.pdf (Accessed 20 July 2018)

⁸ <https://www.nfer.ac.uk/analysis-of-academy-school-performance-in-2015/> (Accessed 20 July 2018)

⁹ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/527043/

demonstrates that, of this set of schools, the sponsored academy system took on more of the inadequate schools than remained local authority maintained.

Accounting for how long sponsored academies have been open

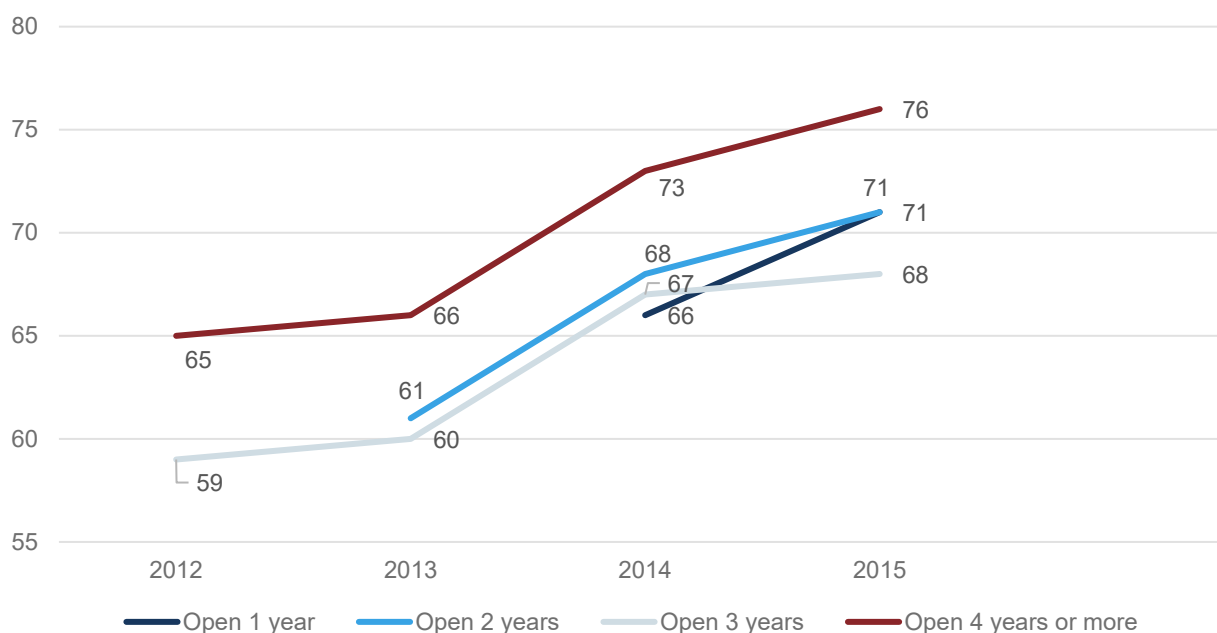
Another fundamental factor in the analysis of sponsored academies is the length of time that they have been open. This can be shown in a variety of ways. Firstly, in terms of summary results published in DfE statistical releases. Figure 3 shows performance trends between 2012 and 2015 for sponsored academies that had opened by 2015. This highlights the very different levels of performance of these academies by length of time open. The first point in each series (except that for those open four or more years) represents the final set of results allocated to the LA predecessor schools. The group open for four years or more consistently performed at least five percentage points ahead of the next best performing group. The improvement rate between 2014 and 2015 for the group open for three years was lower than for the other groups.

Secondly, other analysis, such as Machin and Veroit (2011) and Worth (2016), provides evidence that the positive link between becoming a sponsored academy and school improvement is stronger in schools that have been a sponsored academy for longer.

This reinforces the point that each cohort of sponsored academies is different and including sponsored academies open for different lengths of time within a single grouping of schools introduces an increased risk of unsatisfactorily explained variation in school performance. This is the rationale for the choices made within this report to consider sponsored academy performance within groups structured by their length of time open as an academy.

Figure 3:

Percentage of pupils achieving level 4 or above in reading, writing and mathematics in sponsored academies by length of time open



Making comparisons over a number of years

Changes to the system of testing, examination and accountability in England over the last eight years make comparisons over time difficult. When assessing the potential impact of a change in policy, it is important to try and understand change with respect to a baseline, or earlier position. In recent years Department for Education statistical releases have, where possible, given time series of sponsored academy performance starting with the last year of performance as an LA school to act as this baseline (such as in figure 3 above). DfE's previous methodology for measuring the impact of multi-academy trusts¹⁰ took steps to produce a robust baseline to assess improvement.

If analysis looks at a single year of performance in isolation, particularly one a few years after academisation, the analysis may find no statistically significant performance difference between the sponsored academies and similar schools and conclude that there has been no benefit arising from academisation, such as in Worth (2016). Taking a longer-term view may lead to different conclusions.

However, comparisons over several years may be undermined by changes in the characteristics of pupils at either sponsored academies or the similar schools groups. For

¹⁰ <https://www.gov.uk/government/statistics/multi-academy-trust-performance-measures-2014-to-2015> (Accessed 20 July 2018)

example, if the prior attainment of pupils attending a sponsored secondary academy improves, this may be a stronger driver of improvement in key stage 4 results than other factors. Andrews et al (2017)¹¹ investigated this and found changes in pupil intake are less marked in post-2010 sponsored academies than in pre-2010 sponsored academies.

Comparisons over time can also affect the similar schools groupings themselves. The similar school analysis in Andrews et al (2017) allows control groups to include schools that have become academies. The risk to this approach is that estimates of the difference in performance between the groups may be affected by an academisation effect taking place within schools in the control group. Our analysis in part 2 explicitly excludes schools that later became academies from our similar school groups.

Comparison with national average results

The previous section showed that schools that became sponsored academies were purposively selected, as they were often performing less well than other schools. This means that these schools typically come from the lower end of the distribution of school performance.

Within figures 4 and 5 below the areas in the coloured bars represent the middle two quartiles of pupil outcomes separated by the line through that area which represents the median. The range from the top of the coloured bar to the top of the “whisker” represents the upper quartile (excluding outliers). The range from the bottom of the coloured bar to the bottom of the whisker represents the lower quartile (excluding outliers).

Figure 4 below shows the distribution of pupil outcomes in 2016, 2017 and 2018 in secondary sponsored academies for pupils achieving grade C/4 or above in GCSE English and mathematics relative to the national average performance (for state-funded mainstream schools). This chart shows that, on average, those sponsored academies open for longest (the pre-2010 group) are performing better than more recent sponsored academies and have performance closer to the national average. What it also shows is that the performance of sponsored academies within these groups is variable, notably in those sponsored academies open for longest.

¹¹ <https://epi.org.uk/publications-and-research/impact-academies-educational-outcomes/> (Accessed 23 July 2018)

Figure 4:

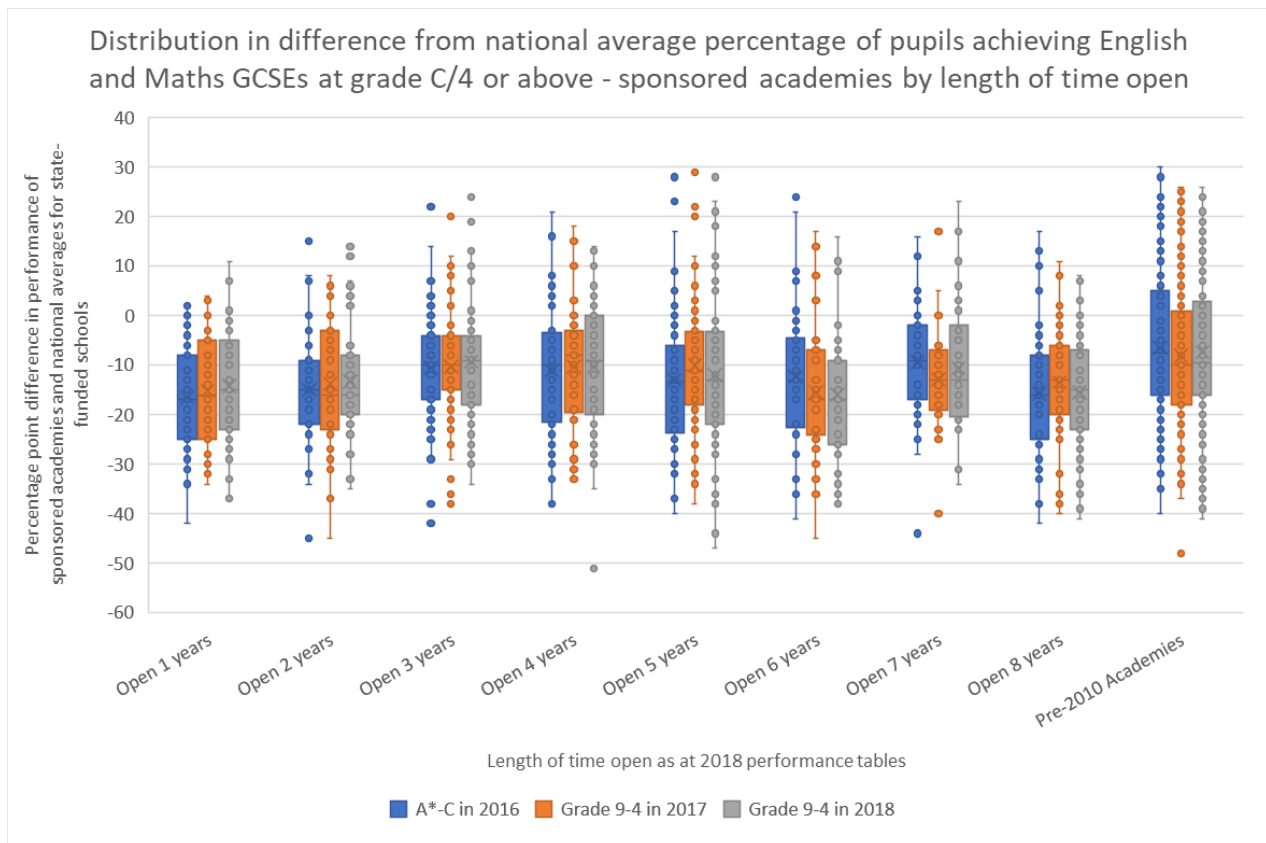
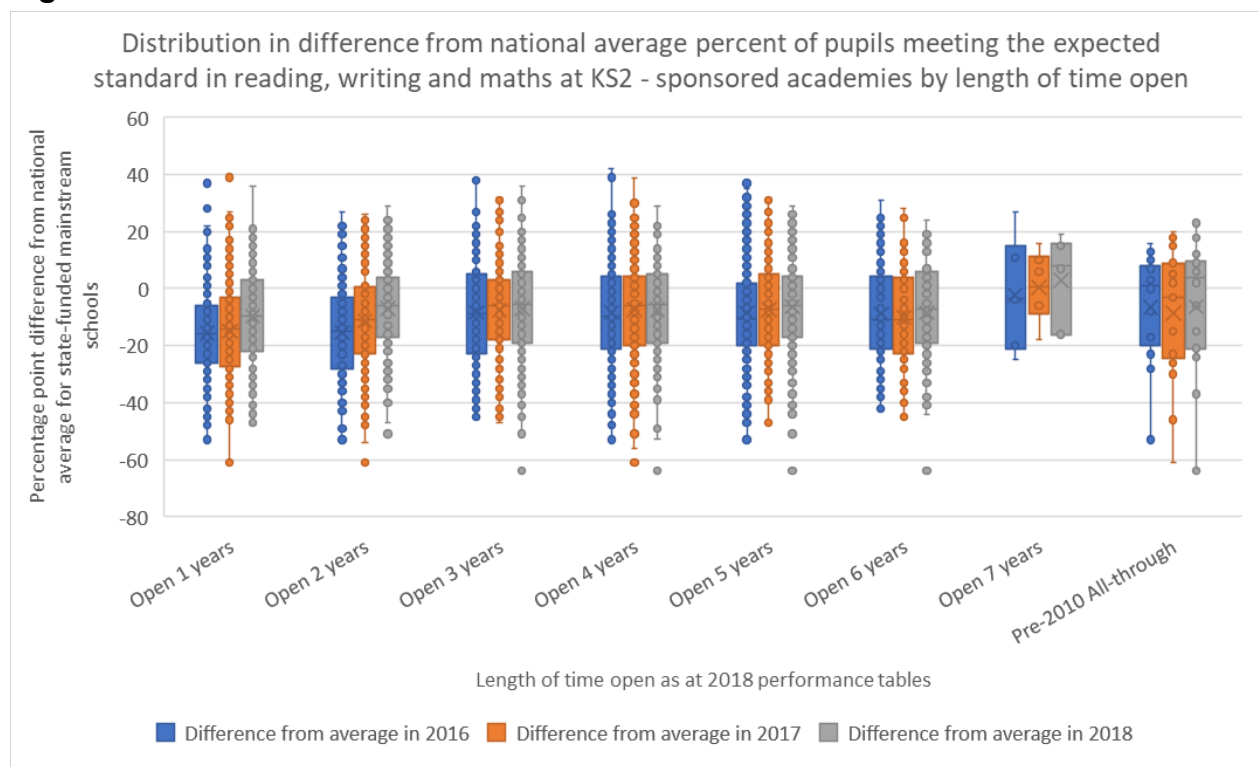


Figure 5 shows the distribution of pupil outcomes in primary sponsored academies for pupils achieving the expected standard at the end of key stage 2 relative to the national average performance (for state-funded mainstream schools). The data shows that at KS2, like KS4, the older sponsored academies (the pre-2010 group) are performing closer to national average performance. Here the “pre-2010” group were “all through” sponsored academies open prior to 2010 that offer primary and secondary education within the same school. This “pre-2010” group also includes some schools that have key stage 2 results considerably lower than the national average.

Figure 5:



Summary of part 1

Within this part of the report we have examined the historical context of sponsored academies. We noted that most but not all sponsored academies replaced a pre-existing school. We examined an example group of schools that were rated inadequate at one point in time. That example provides some evidence that the schools that became sponsored academies had greater challenges to overcome in improving than those that remained local authority maintained.

We have examined a comparison of sponsored academy results against the national average results. Within that there is some evidence to suggest that, on these measures, the older cohorts of sponsored academies have performance level closer to the national average than more recent sponsored academies. These results also demonstrate the variation in outcomes of sponsored academies and show that there are some that have been open for several years and are performing well below national averages.

Part 2 of the report shows the results of making a fairer comparison between the performance of the different “cohorts” of sponsored academies and similar schools.

Part 2: Comparing performance of sponsored academies with similar schools using propensity score matching (PSM)

Making fairer comparisons using PSM

When creating a group of similar schools with which to compare sponsored academies, we are aiming to be able to contrast two sets of schools in which conditions are as similar as possible. Even though data measuring a variety of aspects of the schools' experience exists, we do not hold sufficient data to fully control for all the influences on pupils within schools. The similar schools groups are constructed primarily on similarity in terms of basic pupil characteristics. However, the matching model we use in this part of the report also includes some of the judgements arrived at through Ofsted inspections. These are included to encapsulate variation arising from the quality of the teaching and leadership within schools. We have made a deliberate decision not to include proxy measures of school quality (or effectiveness) derived from test or examination results of pupils that attended these schools prior to academisation since these are essentially the outcomes of interest. However, we have run alternative analysis including these and the broad conclusions this report arrives at remain consistent.

Unlike Andrews et al (2017), the analysis in this part of the report does not solely focus on pupils who have been enrolled in these schools for the full period of their education prior to the end of key stage 2 or 4. However, the similar schools analysis in this part is accompanied by data in the appendices which summarise pupil characteristics in the sponsored academies and similar schools over time. The appendices also contain further detail on the underpinning methodology of propensity score matching. Additionally, we have listed the names of the sponsored academies and their respective trusts in a separate annex to this report.

How to interpret the charts in this part of the report

The charts in part 2 of this report compare the performance of a group of sponsored academies with a group of similar non-academy schools. The figures plotted on the charts are calculated by subtracting the results for similar schools from the results for the sponsored academies. So positive differences represent sponsored academies having results better than similar schools, negative differences are worse results than similar schools. The area shaded in blue represents our estimation of the degree of uncertainty around the results, it shows what results might have looked like had we compared this group to a set of different comparison schools, but that were generated using the same data and modelling approach.

Our estimates are dependent, to an extent, on the schools that remain under local authority control. Given that this is an ever-changing set, from one year to another it is

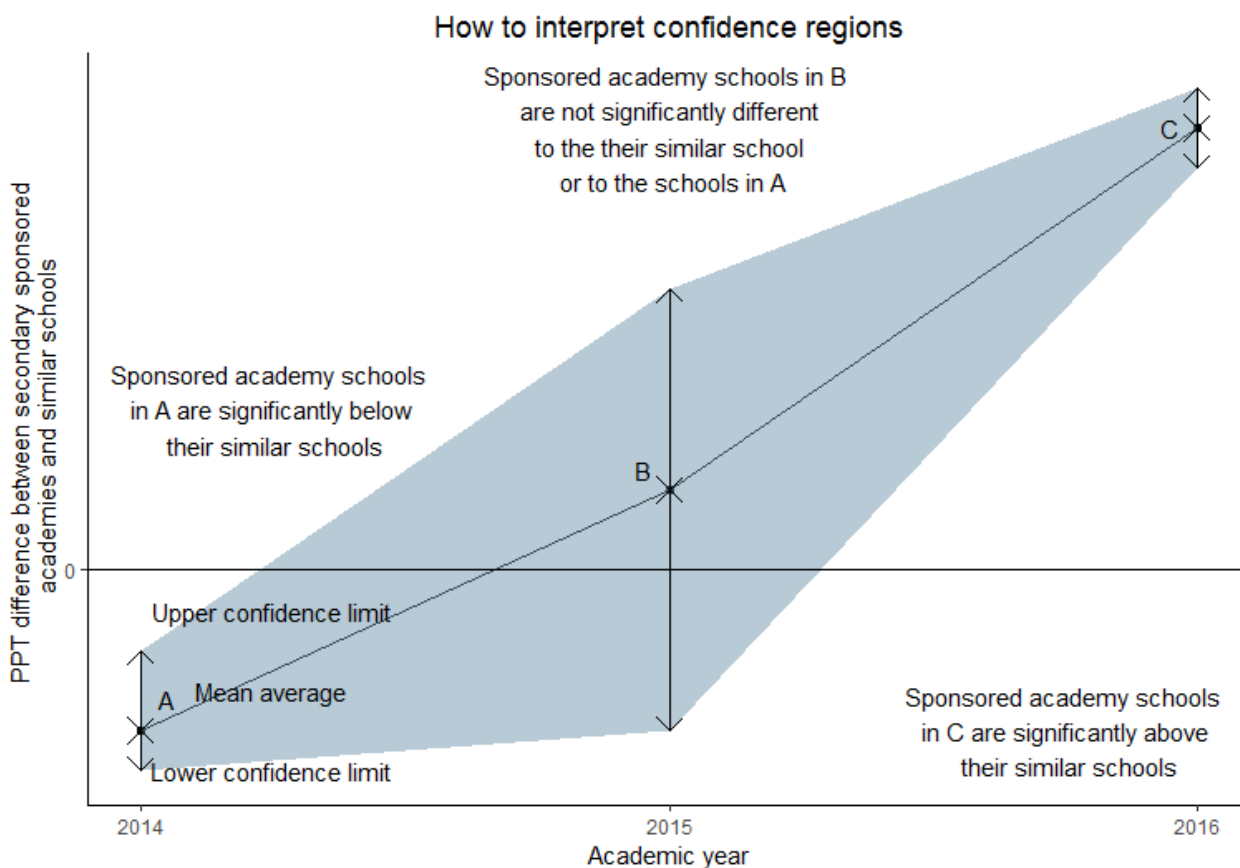
not possible to maintain a constant similar school (control) group. If we were to repeat this analysis in a year's time, we would expect slightly different results, even for the same cohorts of sponsored academies, as the comparison groups must change as time passes. This change in data and the impact on the underlying modelling approach might lead to estimates that do not fall within the confidence regions illustrated in this report.

When we measure our outcome numerous times and calculate a 95% confidence interval each time, 95% of our confidence intervals contain the true mean. When comparing two independent samples, if the confidence intervals do not overlap then the two different means are significantly different; if they do overlap, the results are unlikely to be significantly different. This is shown diagrammatically in Figure 6.

$$\bar{x} = \text{Result for sponsored academies} - \text{results for similar school}$$

$$95\% \text{ CI} = \bar{x} \pm 1.96 \sigma$$

Figure 6:



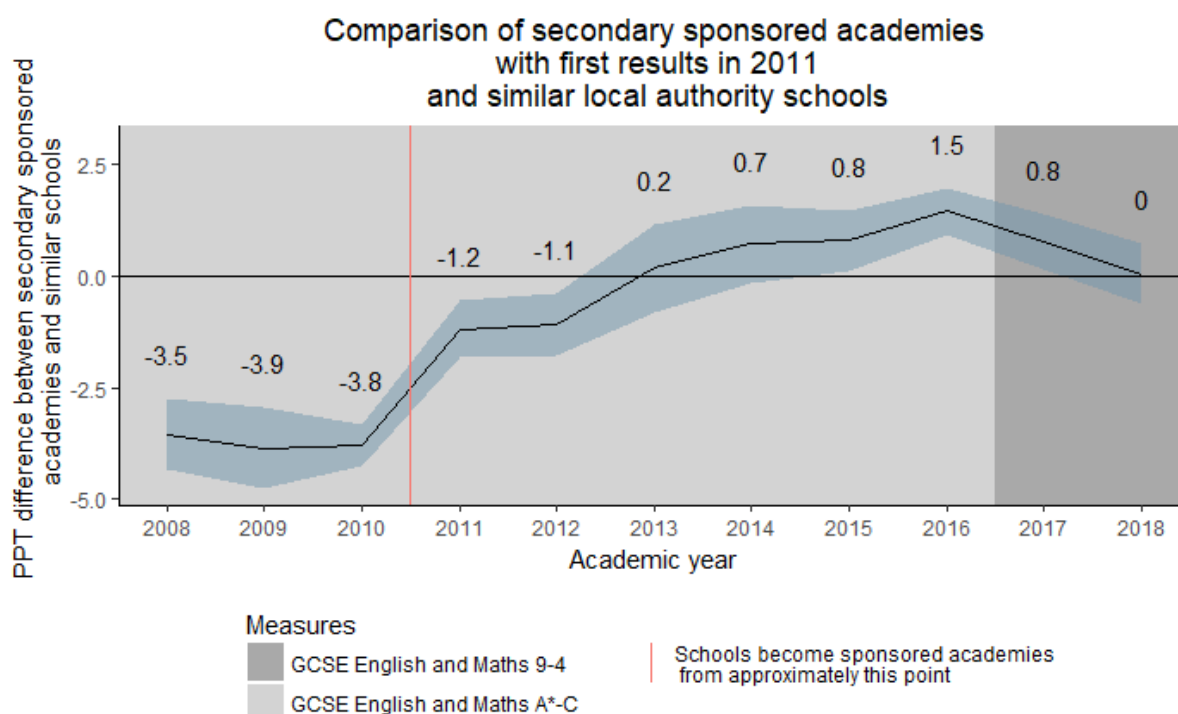
Detailed results: Secondary sponsored academies (KS4)

Secondary sponsored academies with first academy results in 2011

Figure 7 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2011 and similar schools. The results are based on 43 schools in each group, which had an average of 914 pupils in each sponsored academy in 2018 (total of 39,300 pupils)¹². It shows that:

- In the years (of our analysis) prior to becoming sponsored academies, this set of sponsored academies was falling further behind the set of similar schools. In 2008, 2009 and 2010 the average performance of these sponsored academies is significantly lower than in the similar schools.
- From 2011 onwards, the year of their first set of academy results, the average performance of these sponsored academies begins to improve and is either similar to or significantly better than the similar schools from 2013 onwards.

Figure 7:



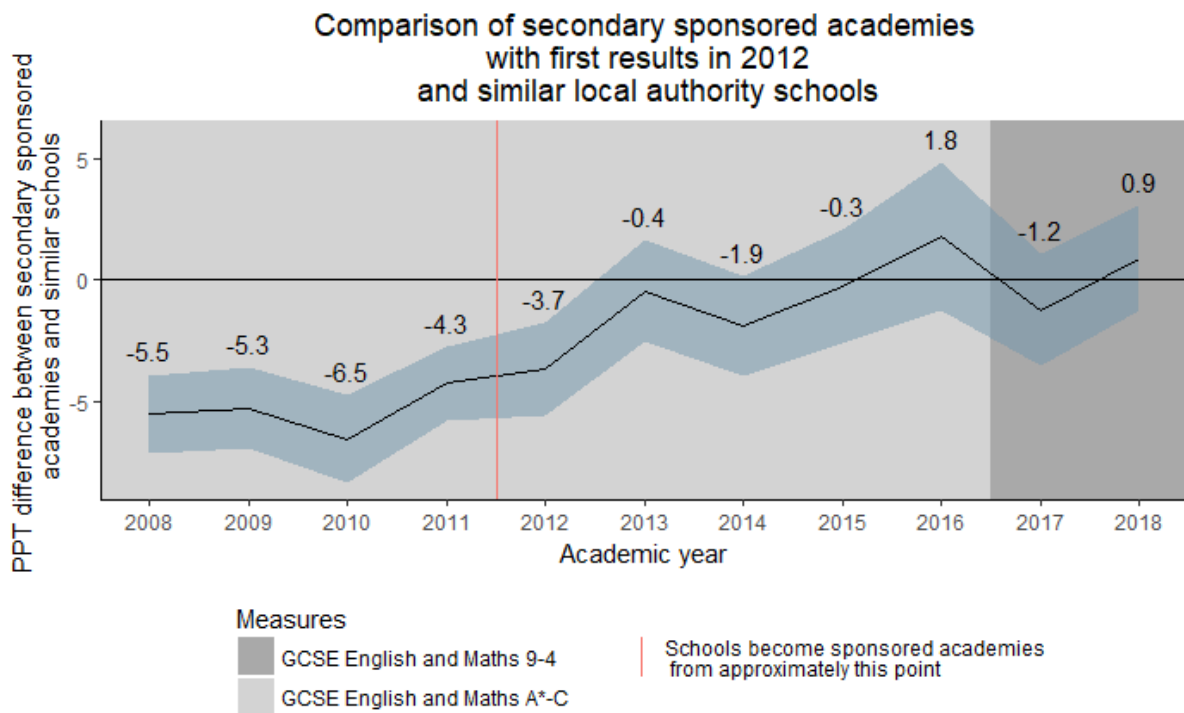
¹² This figure represents the average total number of pupils in each of the academies in the sponsored academy group, and is not simply the number of pupils at the end of the key stage, which is often reported alongside school performance data.

Secondary sponsored academies with first academy results in 2012

Figure 8 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2012 and similar schools. The results are based on 31 schools in each group, which had an average of 873 pupils in each sponsored academy in 2018 (total 27,100). It shows that:

- Prior to becoming a sponsored academy, these schools were performing significantly less well than the set of similar schools. However, they were on an improving trajectory.
- After becoming academies performance continued to improve. Since 2013 there has been no significant difference in performance between the sponsored academies and the similar schools.

Figure 8:

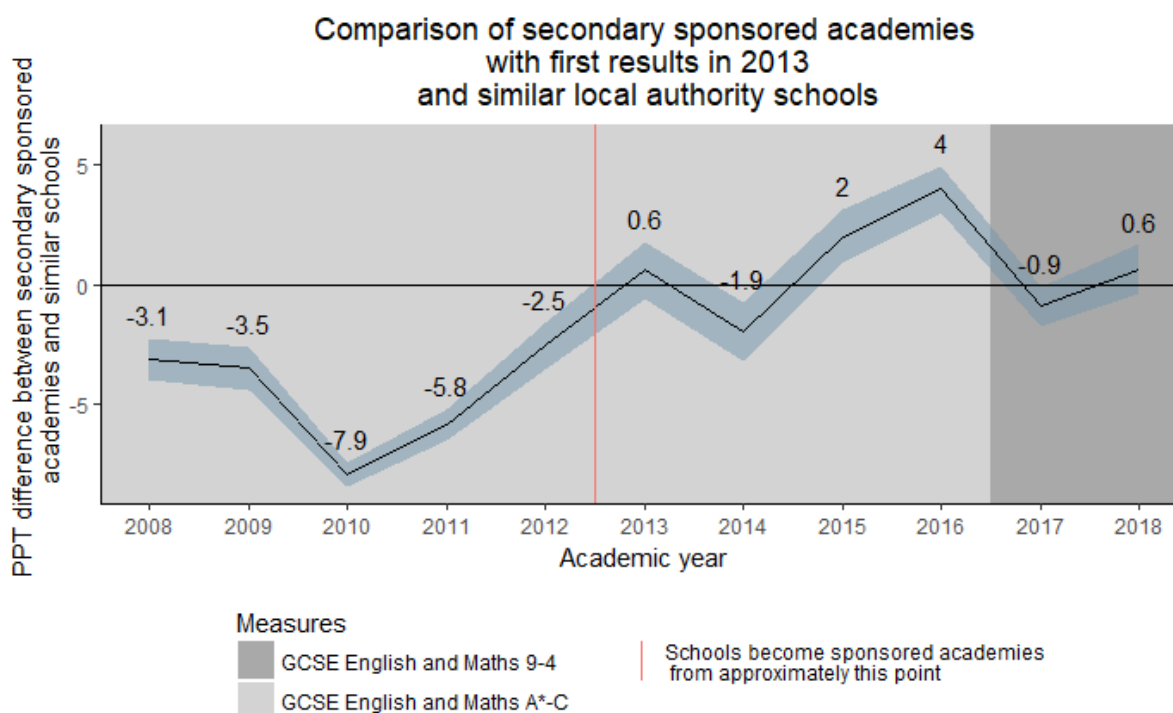


Secondary sponsored academies with first academy results in 2013

Figure 9 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2013 and similar schools. The results are based on 39 schools in each group, which had an average of 824 pupils in each sponsored academy in 2018 (32,100 total). It shows that:

- In the three years prior to first results as sponsored academies, the group of sponsored academies appear to narrow the gap with the set of similar schools.
- Results from 2013 onwards fluctuated, but in 2015 and 2016 we can be confident that sponsored academies outperformed the similar schools.
- These schools saw a decline in performance compared to their similar schools coinciding with the changes to KS4 performance measures in 2017, but their performance has remained comparable to the similar schools.

Figure 9:

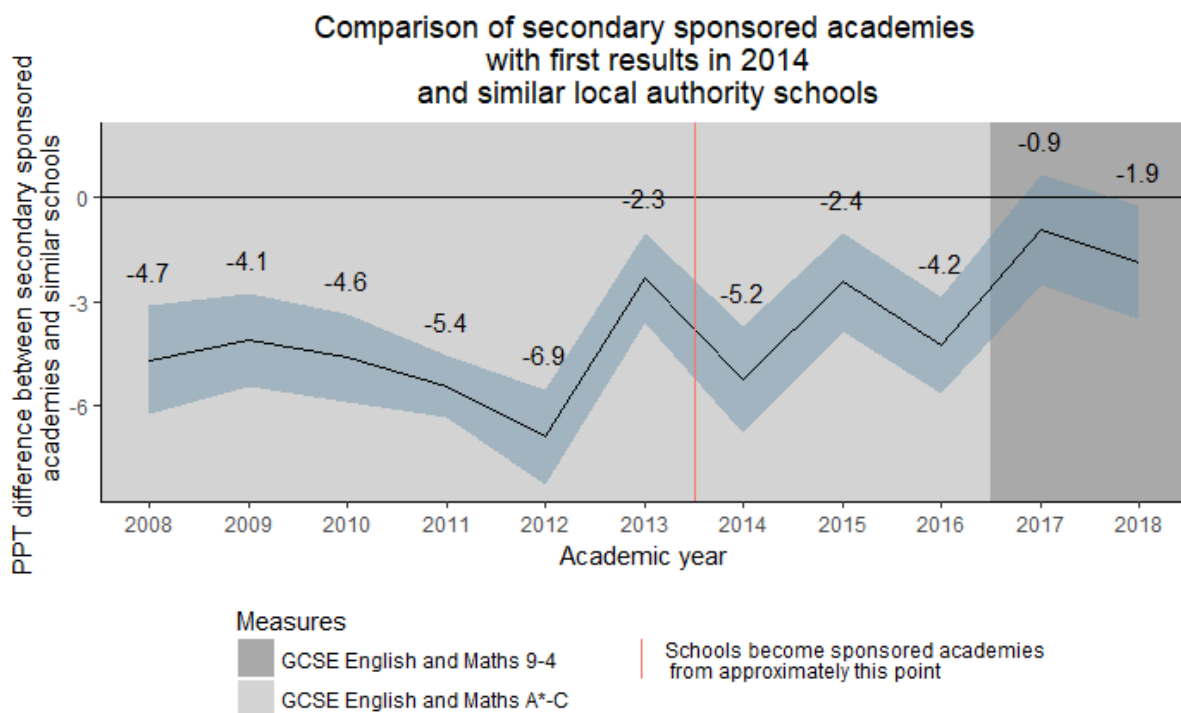


Secondary sponsored academies with first academy results in 2014

Figure 10 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2014 and similar schools. The results are based on 40 schools in each group, which had an average of 802 pupils in each sponsored academy in 2018 (32,100 total). It shows that:

- In all years of our analysis before becoming a sponsored academy, these schools performed significantly less well than the set of similar schools.
- In the years immediately after becoming sponsored academies these schools continued to perform less well than the similar schools, with fluctuating levels of relative performance. The performance decline in 2014 coincides with performance measures including a pupil's first entry in an examination subject rather than their best entry.
- In 2017, the performance of schools in this group that became sponsored academies was not significantly different to that of the similar schools, but was below their similar schools' average in 2018.
- In 2018, pupils in these schools will typically have had their full secondary education while their school was an academy.

Figure 10:

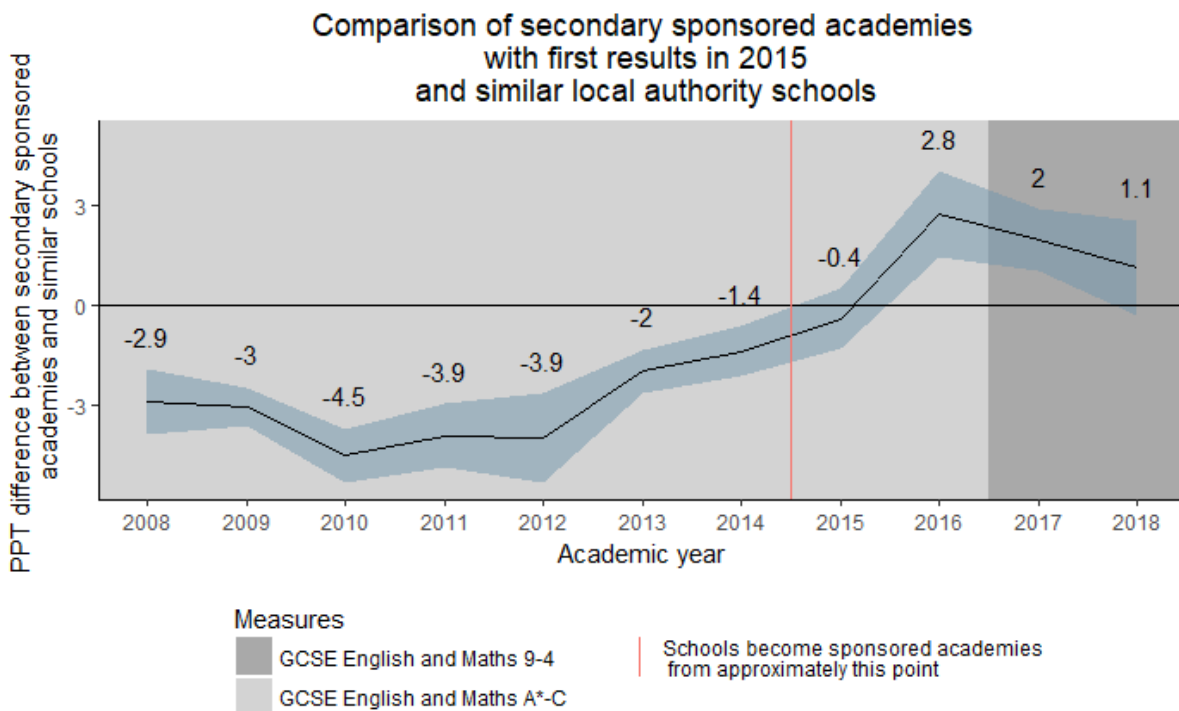


Secondary sponsored academies with first academy results in 2015

Figure 11 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2015 and other similar schools. The results are based on 33 schools in each group, which had an average of 737 pupils in each sponsored academy in 2018 (24,300 total). It shows that:

- In the years prior to becoming an academy, the schools that became sponsored academies performed significantly worse than their similar schools, but did move closer to the similar schools over the years prior to academisation.
- Immediately after becoming a sponsored academy in 2015, performance of sponsored academies had continued to improve relative to, but was not significantly different from, the set of similar schools.
- We can be confident that the sponsored academies performed better than the set of similar non-academies in 2016 and 2017
- Although there has been a decrease in relative performance in 2018, we can be confident that the sponsored academies are not performing significantly worse than the set of similar schools.

Figure 11:

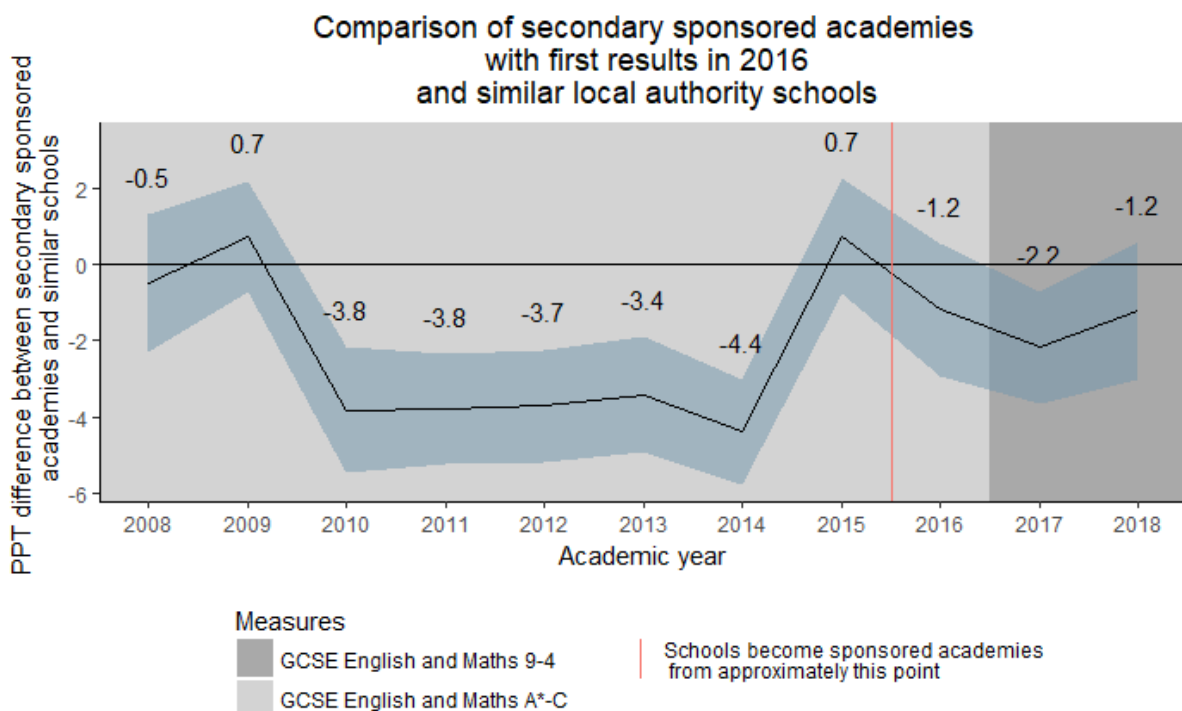


Secondary sponsored academies with first academy results in 2016

Figure 12 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2016 and similar schools. The results are based on 40 schools in each group, which had an average of 775 pupils in each sponsored academy in 2018 (31,000 total). It shows that:

- In the majority of the years (in our analysis) prior to academisation the schools that became sponsored academies performed significantly less well than similar schools, although appeared to show some improvement immediately prior to conversion that was not sustained.
- In the first two years of results after academisation performance of the sponsored academies declined in comparison with the set of similar schools.
- Between 2017 and 2018 the sponsored academies improved relative to the set of similar schools and were not significantly different from them on average by 2018.
- By 2018 pupils in these schools will typically have had just over half of their secondary education while their school was an academy.

Figure 12:

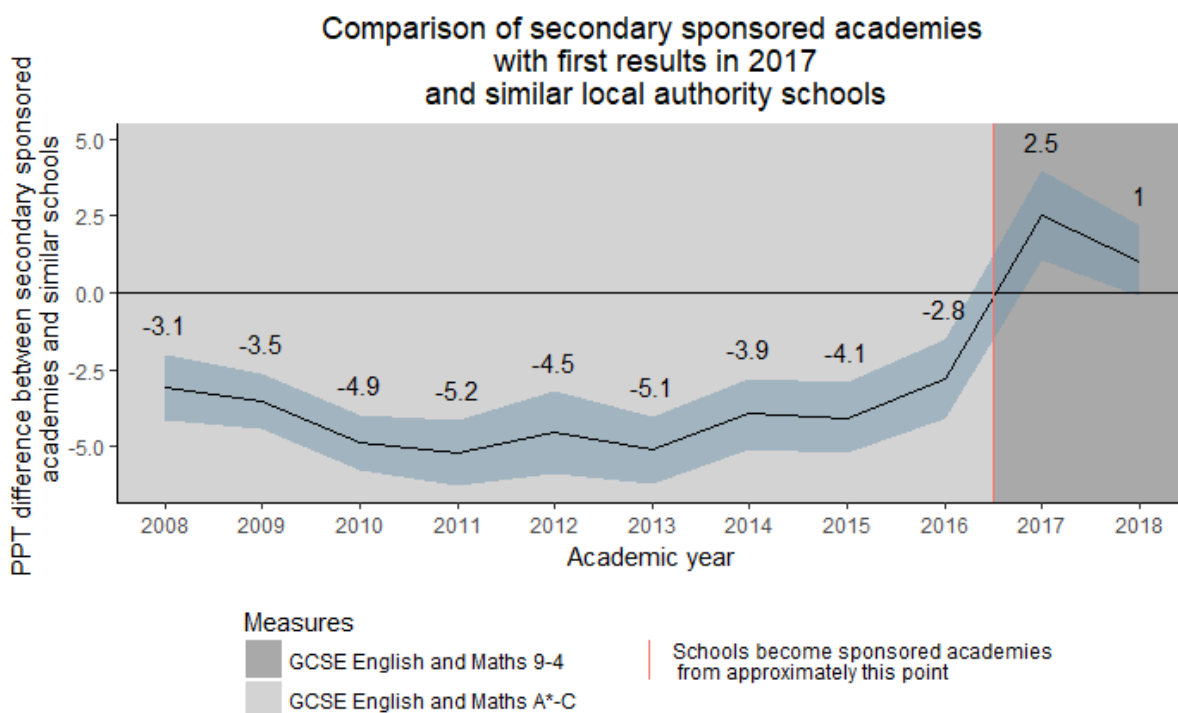


Secondary sponsored academies with first academy results in 2017

Figure 13 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2017 and similar schools. The results are also based on 25 schools in each group, which had an average of 872 pupils in each sponsored academy in 2018 (21,800 total). It shows that:

- In all years (in our analysis) prior to academisation the schools that became sponsored academies performed less well than their similar schools.
- The confidence region again allows an interpretation that suggests the performance of the set of schools that became sponsored academies was behind similar schools, but improving before becoming sponsored academies in comparison with the similar schools.
- In 2017, their results were significantly better than their similar schools, and in 2018 they were performing in line with the similar schools.

Figure 13:

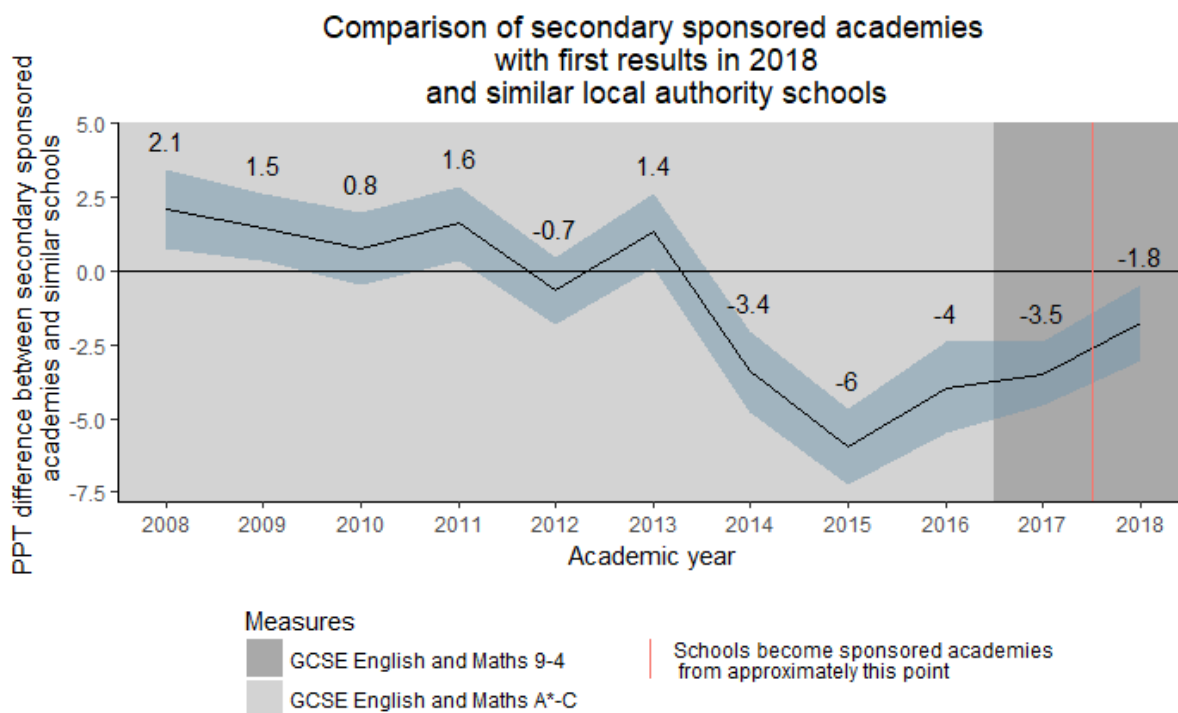


Secondary sponsored academies with first academy results in 2018

Figure 14 illustrates the difference in attainment between secondary sponsored academies with first academy results in 2018 and similar schools. The results are also based on 49 schools in each group, which had an average of 806 pupils in each sponsored academy in 2018 (39,500 total). It shows that:

- The schools that became sponsored academies in this group had broadly similar, but slowly declining, performance compared to their set of similar schools between 2008 and 2013, before they began to decline rapidly from 2013 to 2015. This again coincides with changes to performance measures that mean only a pupil's first entry to a GCSE examination is counted rather than their best entry.
- These sponsored academies then performed significantly below the similar schools' average between 2014 and 2017, albeit slightly closing the gap prior to becoming sponsored academies.
- Since becoming sponsored academies, these schools remain significantly behind the similar schools, but continue to close the gap on the similar schools' performance.

Figure 14:



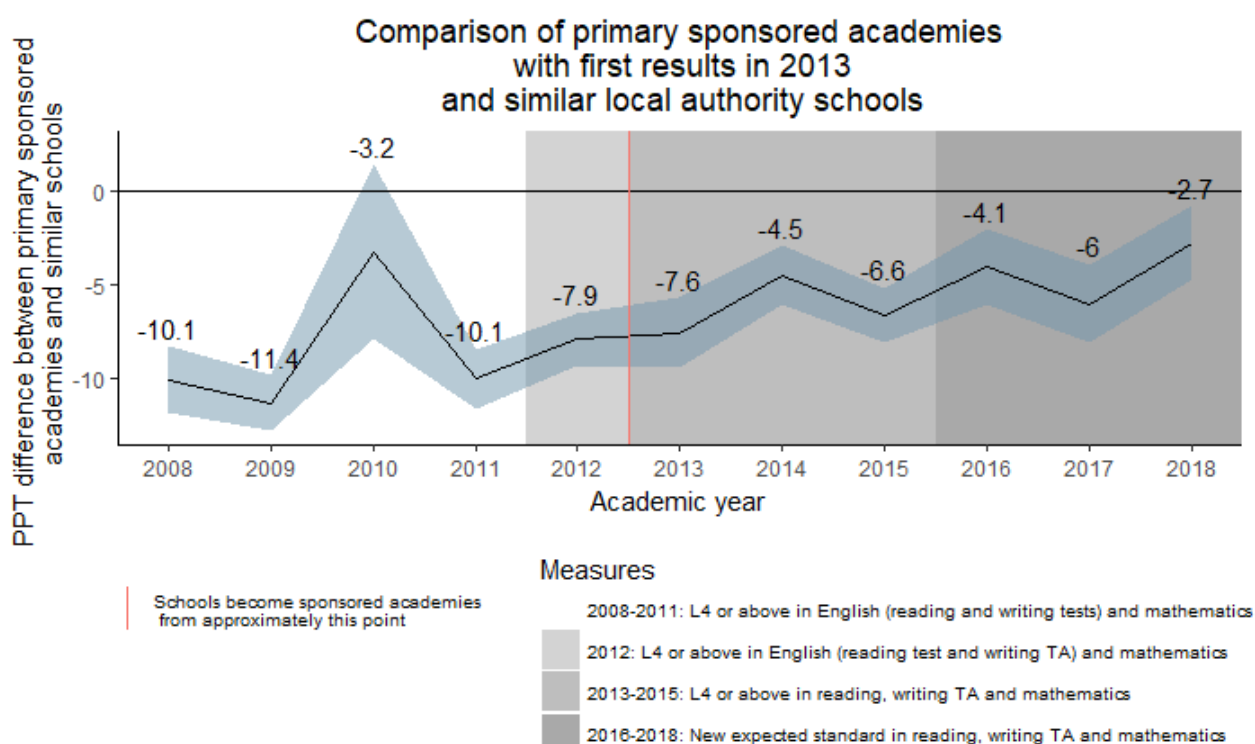
Detailed results: Primary sponsored academies (KS2)

Primary sponsored academies with first academy results in 2013

Figure 15 illustrates the difference in attainment between primary sponsored academies with first academy results in 2013 and similar schools. The results are based on 94 schools in each group which had an average of 346 pupils in each sponsored academy in 2018 (32,500 total). It shows that:

- Since becoming sponsored academies, the schools in this group show some evidence of closing the gap compared to the set of similar schools. The gap widened in both 2015 and 2017 suggesting that they may not be closing this gap consistently over time.
- The confidence region is wider in 2010 due to greater variation in our similar schools groupings. This is driven by missing data from schools that did not take part in the 2010 national curriculum tests.

Figure 15:

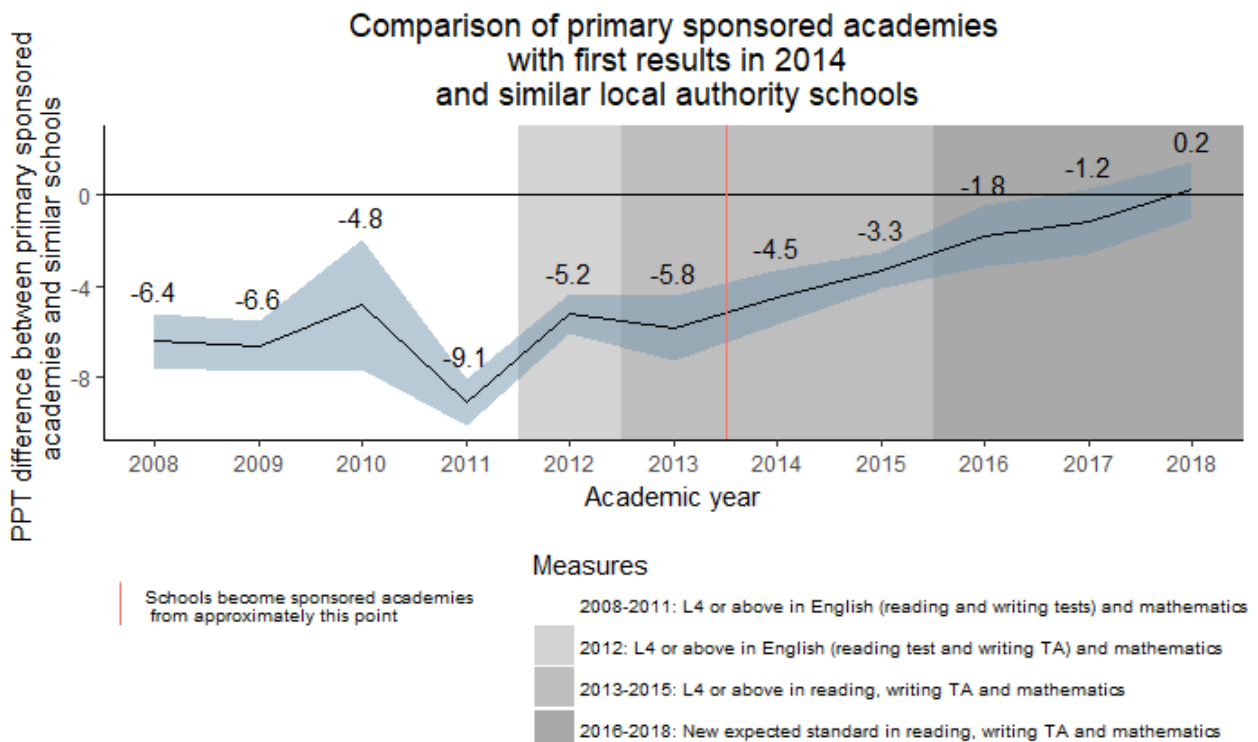


Primary sponsored academies with first academy results in 2014

Figure 16 illustrates the difference in attainment between primary sponsored academies with first academy results in 2014 and similar schools. The results are based on 177 schools in each group which had an average of 332 pupils in each sponsored academy in 2018 (58,700 total). It shows that:

- In all years of our analysis the performance of schools that became sponsored academies in this group remained significantly below the set of similar schools until 2017, when they were not significantly different.
- The gap in performance has reduced in each year between becoming sponsored academies and 2018.

Figure 16:

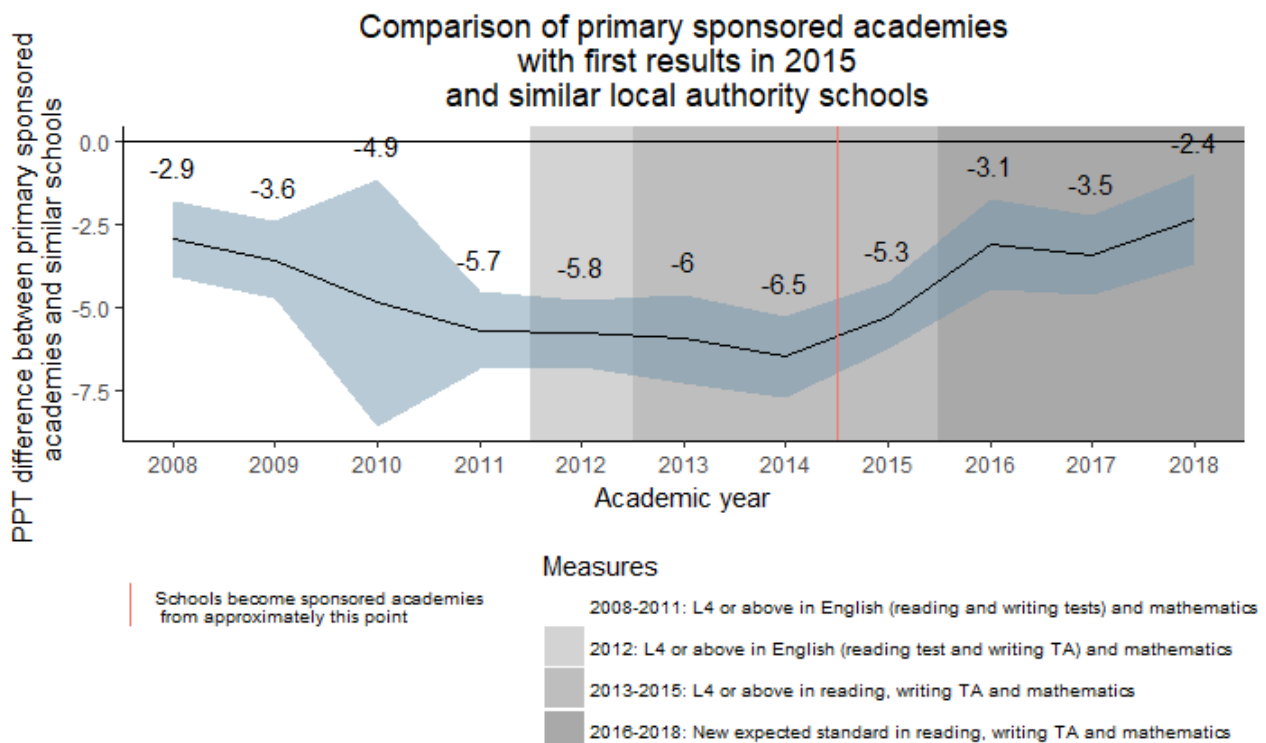


Primary sponsored academies with first academy results in 2015

Figure 17 illustrates the difference in attainment between primary sponsored academies with first academy results in 2015 and similar schools. The results are based on 143 schools in each group which had an average of 308 pupils in each sponsored academy in 2018 (44,000 total). It shows that:

- Prior to becoming sponsored academies, schools that became sponsored academies consistently performed significantly less well than the set of similar schools. On average, the results of the group were in decline from 2008 to 2014.
- Since becoming academies these schools have continued to perform significantly less well than the similar schools; however, there is evidence to suggest that there has been some narrowing of the performance gap, and on average in 2018 they were at their highest level relative to the similar schools since 2008.

Figure 17:

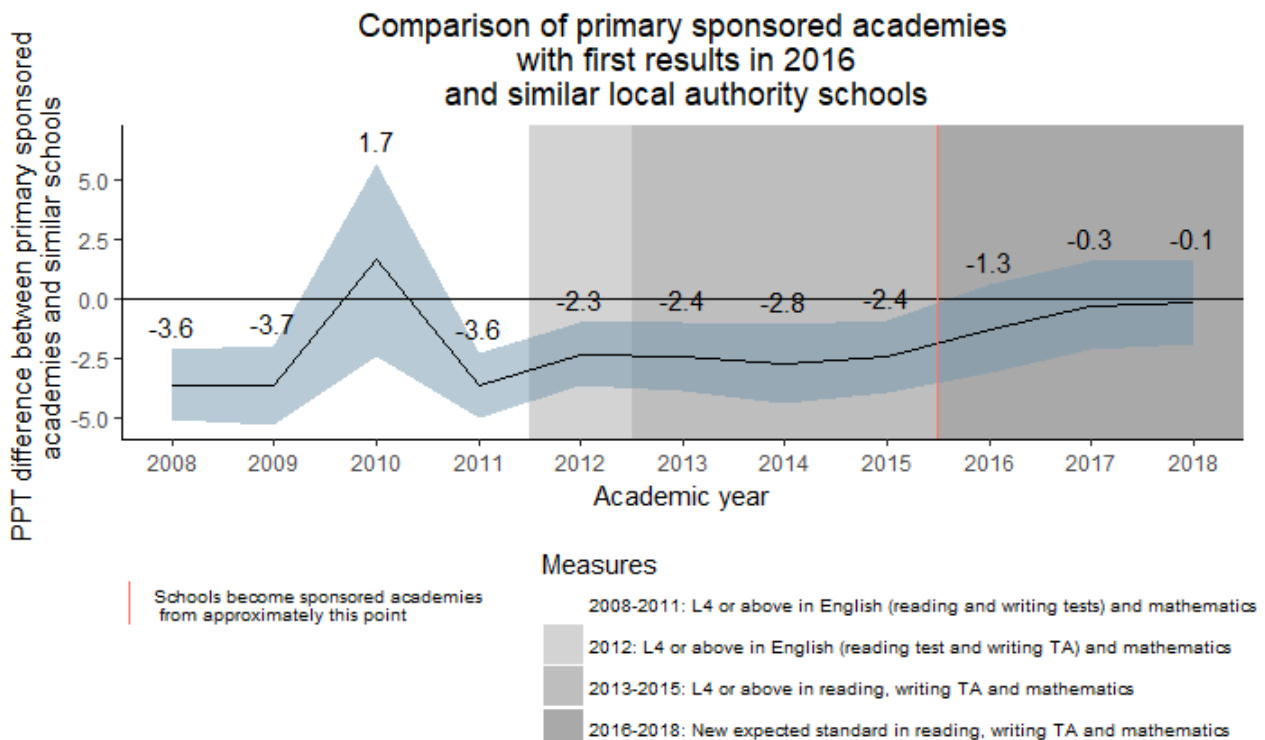


Primary sponsored academies with first academy results in 2016

Figure 18 illustrates the difference in attainment between primary sponsored academies with first academy results in 2016 and similar schools. The results are based on 106 schools in each group which had an average of 290 pupils in each sponsored academy in 2018 (30,800 total). It shows that:

- The point estimate in 2010 is unlikely to be representative of the true position as some schools boycotted key stage 2 tests that year. Other than this schools that are or were to become sponsored academies performed significantly less well than the set of similar schools before becoming academies (although were not as far behind the similar schools as some of the older groups of primary sponsored academies).
- Figure 18 gives some evidence to suggest that results have improved after academisation in comparison with their similar schools, and that since 2016, their results have not been significantly different to the similar schools.

Figure 18:

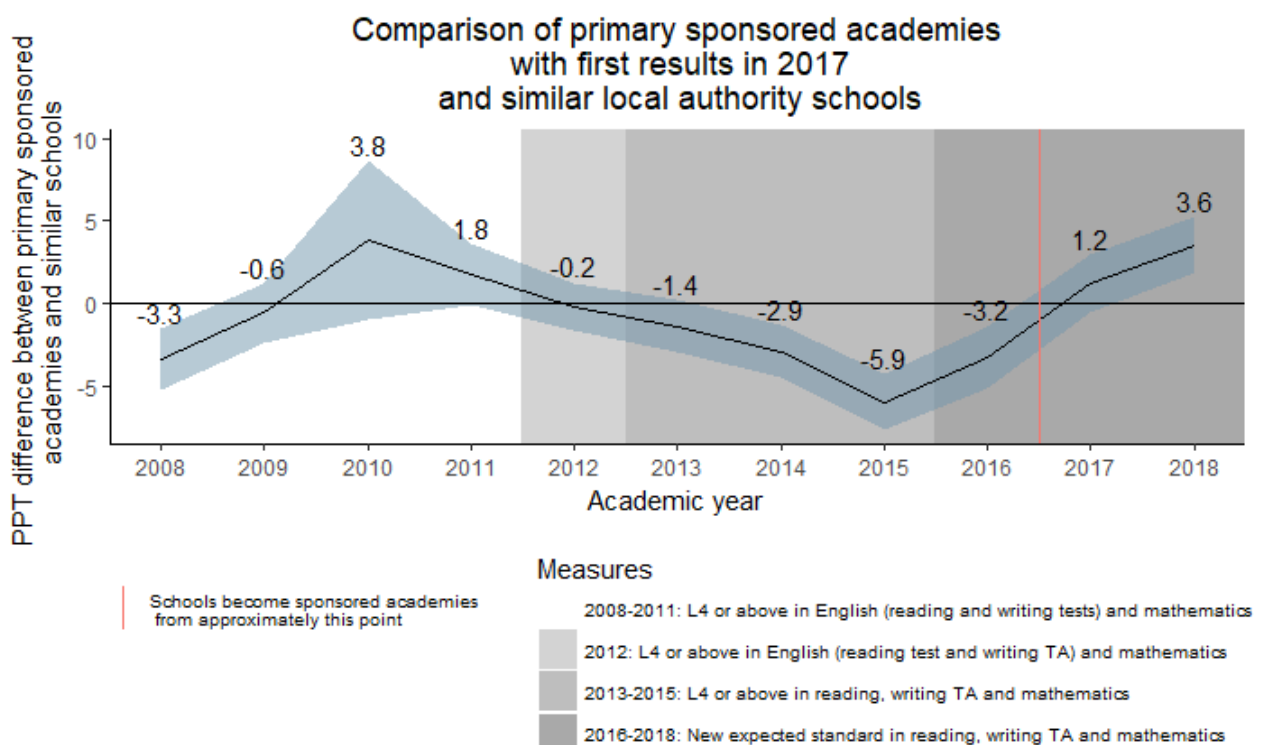


Primary sponsored academies with first academy results in 2017

Figure 19 illustrates the difference in attainment between primary sponsored academies with first academy results in 2017 and similar schools. The results are based on 82 schools in each group which had an average of 286 pupils in each school in 2018 (23,400 total). It shows that:

- Results in 2010 are unrepresentative of the true position due to the test boycott that year. In most of the years prior to becoming sponsored academies, these schools performed significantly less well than the set of similar schools.
- The schools that became sponsored academies began to improve prior to academisation between 2015 and 2016. Improvement has continued after academisation.
- In both years since becoming a sponsored academy (2017 and 2018), the sponsored academies have had performance at least as good as the similar schools, and in 2018 the performance of these schools was significantly above their similar schools.

Figure 19:

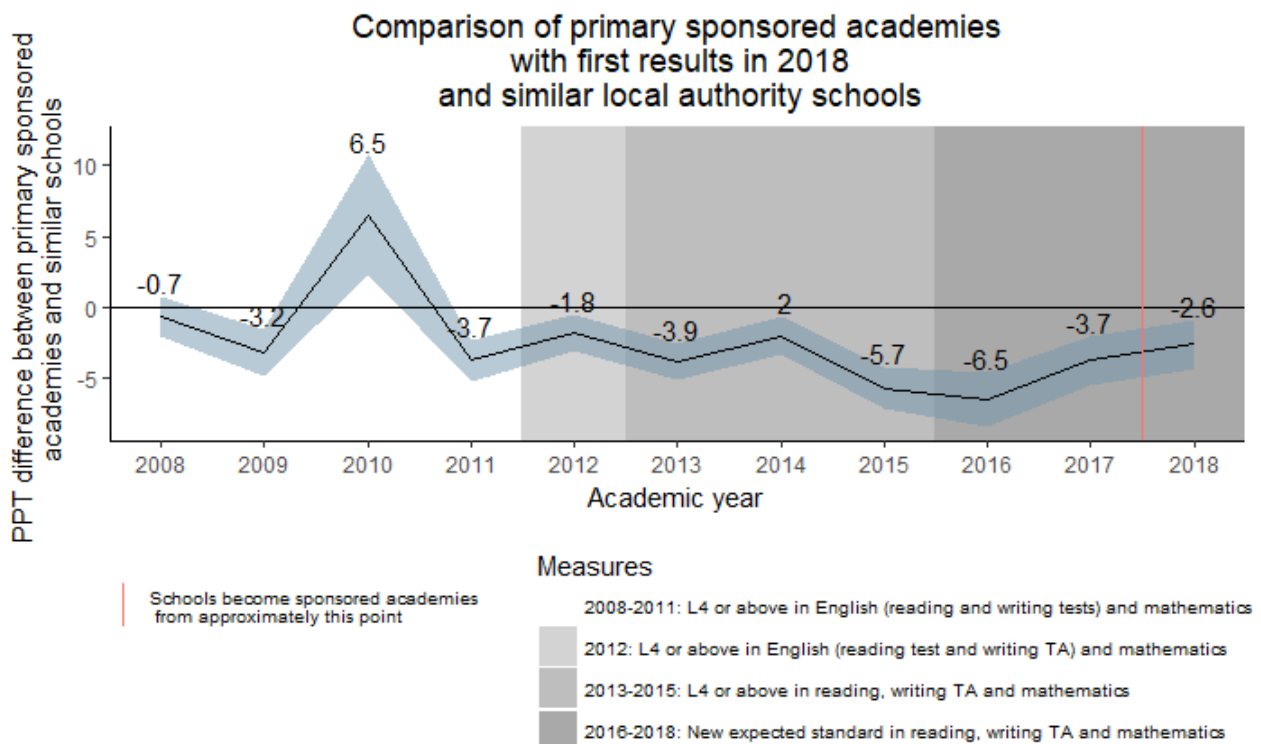


Primary sponsored academies with first academy results in 2018

Figure 20 illustrates the difference in attainment between primary sponsored academies with first academy results in 2018 and similar schools. The results are based on 145 schools in each group which had an average of 295 pupils in each school in 2018 (42,700 total). It shows that:

- Results in 2010 are likely to be unrepresentative of the true position due to the test boycott that year. In all other years prior to becoming sponsored academies, these schools performed significantly less well than the set of similar schools.
- In 2018 the sponsored academies remained behind the performance of the similar schools, but continued the trend prior to becoming academies of closing the gap to the smallest since 2014.

Figure 20:



Conclusions

The findings of this work reflect that typically the schools that became sponsored academies had many challenges in front of them. On some important measures these challenges were greater than for schools that remained in the local authority sector. This report shows evidence to suggest that sponsored academies have been open longer perform more in line with the average performance of state-funded (mainstream) schools. However, when looking at the distributions of performance of schools within these different groups, substantial variation is evident. There are academies, some open for several years, which continue to perform well below national average performance.

The outcomes of our work based on fair comparisons are more positive for secondary than for primary sponsored academies. All of the older groups of secondary academies have now caught up with their similar schools, and in some cases overtaken them, although these differences are not statistically significant. These findings are in line with other research such as Worth (2016)¹³.

These findings fill a previous evidence gap by examining outcomes for pupils in primary sponsored academies. However, the picture is complex. Most groups of primary sponsored academies showed signs of improvement towards the performance of similar schools with one group exceeding the performance of similar schools, two groups improving results to a point where they are in line with similar schools, and the three remaining groups continuing to perform below the level of their similar schools. The fact that these schools have typically improved over time is of benefit to the pupils that are attending them. Many primary academies are in their early years as academies and there is scope for continued school improvement. Some more recently opened primary sponsored academies are performing in line with their similar schools, but these schools had a smaller performance gap with similar schools prior to academisation.

Not all groups of sponsored academies have improved consistently over time and there is evidence to suggest that some have slipped back in their comparative performance (although it is possible that some of this slipping back may be explained by improved performance in the similar schools).

Although these results cannot *prove* that changes in performance of sponsored academies are fully caused by academisation (and there is scope for further improvements to our analytical approach), the findings in this report do demonstrate that, overall, pupil outcomes in sponsored academies have typically improved since their formation in comparison with sets of similar schools.

¹³ Worth, J. (2016). Analysis of Academy School Performance in 2015 (LGA Research Report). Slough: NFER.

Appendices

Data and methodology

Propensity Score Matching (PSM)

Sponsored academies are historically underperforming schools. It would be misleading to compare directly their attainment over time to all other state-funded schools. A fair comparison is comparing sponsored academies to other similar schools.

Propensity score matching enables the construction of these fairer counterfactual groups for comparison over time. This technique was used in a previous DfE publication (see page 6, footnote 1) and used by the National Audit Office in 2010 in their independent assessment of the Academies Programme¹⁴. It was also used in recent work to investigate the 'London effect'.¹⁵

PSM works by first, calculating the probability of each school becoming a sponsored academy and then, pairing (matching) sponsored academies with other similar schools that have not since converted based on those probabilities.

The probability of becoming a sponsored academy is estimated using key pupil and school characteristics that affect how likely a school is to become an academy. These are:

- Percentage of pupils eligible for free school meals (FSM),
- Prior attainment,
- Ofsted rating of the school,
- The percentage of children, who are in the black and ethnic minority (BAME) group
- The number of pupils in the school.

The probabilities are calculated for all schools, including the sponsored academies at the point before they converted. Models were chosen by ROC chart, AUC value, percentage accuracy of predictions and McFadden value^a. A range of models were considered

¹⁴ National Audit Office (2010) "Department for Education: The Academies Programme"

¹⁵ Allison (2018) "Academic performance of disadvantaged pupils in and out of London: an analysis". Department for Education, Sheffield.

^a See glossary in the Annex for detailed definitions of ROC charts, AUC values and McFadden values.

including a stepwise model of all parameters, interaction terms and squared terms, a stepwise model of all linear terms and a reduced model.

Sponsored academies are matched to non-academy schools that had a similar, modelled probability (propensity) to become a sponsored academy. This provides fair groups of comparator schools with which to compare performance over time. There are several different methods that can be used to perform the matching; different methods can give rise to slight variations in results and there are pros and cons depending on the method used. For this analysis, we considered the following methods:

- Nearest neighbour matching, including scenarios with replacement, without replacement and with or without a calliper. This method matches sponsored academies to other similar schools based on the closest propensity score. This method is relatively straightforward and quick for the software to process. It also typically results in a high number of matches. A drawback of this method is that it can potentially result in matches that are not as close as desired. To mitigate this problem a calliper can be used to restrict the maximum possible distance between matches. The use of a calliper can result in better quality matches but can reduce the overall number of matches. All the above can be performed with or without replacement after matches have been made. Allowing replacement could result in better quality matches. However, it could also result in one school being matched to many others. This could result in over-reliance on the results of that particular school. Matching without replacement overcomes this issue but could result in matches of a lower quality.
- Exact matching. This method matches sponsored academies to another similar school, which have exactly the same propensity score. This can also be conducted with or without replacement. The advantage of this method is that it results in perfect matches. The downside is that it could result in no or very few matches.

The results of propensity score matching can be sensitive to the underlying data and the order in which matching occurs. To mitigate the risk, our results are based on the average outcome of 200 iterations of our chosen methodology. We decided that this approach outweighed the drawback of increased processing time.

Our chosen methodology was nearest neighbour matching, without replacement, without calliper. This was selected by comparing the differences between the groups before and after matching under different methodologies and on balance was found to be the best approach, factoring in the differences before and after matching and the pros and cons of each variant mentioned above. A calliper was not used as it reduced how many schools in the matched groups and did not substantially change the results.

The tests are summarised over all iterations, which increases the risk of a type 1 error (a type 1 error is when the null hypothesis is rejected even though it is true, a false positive). We applied a Bonferroni correction to reduce the probability of a type 1 error, so each

individual test is compared to an adjusted value. To assess the suitability of the matches, and whether samples are still similar after a longer period in which the academy program has had chance to make a fuller impact, statistical tests are conducted to test that:

- the propensity score distribution of sponsored academies are similar before and after matching to ensure the matched sample still resembles the characteristics of the original sample. A Kolmogorov- Smirnov (KS) test with 5% significance and Bonferroni correction are used;
- propensity scores have similar distribution and mean between sponsored academies and their similar schools either before or after matching. A KS test and t-test with 5% significance and Bonferroni correction for post-match tests are used;
- similar means for the variables used to calculate the propensity score, between sponsored academies and their similar schools either before or after matching. A t-test with 5% significance and Bonferroni correction for post-match tests are used.

Detailed results: choosing a model

Before the sponsored academies can be matched to similar schools, each school must have a propensity score to be matched against (in this case, a probability of becoming a sponsored academy).

The first model considered was a stepwise model with all linear terms, squared terms and interactions (interactions were considered where the absolute correlation between factors was above 0.5). This then guided the choice of which variables to use, where most significant variables across all cohorts and both key stages were selected. Namely, the explanatory variables were percentage of pupils who were eligible for FSM, BAME, previous attainment of pupils, numbers of pupils and school Ofsted rating.

This model was chosen based on being both simple and sufficient in model assessment. When comparing ROC charts of a stepwise individual model, with the chosen model, the difference is negligible. The value for area under the ROC curve range goes from 0.86 and 0.95. Both these sets of values are close to one with one being 100% true positive rate. Accuracy predictions are limited, with the chosen model being less accurate.

McFadden values range from 22% to 50%. Therefore, the reduced model in theory explains slightly less variation but still a good proportion.

In light of the model being used to assign a propensity rather than make a solid prediction, with the ROC values and McFadden values also considered, the reduced model is still sufficient and simple.

Choosing how many iterations

Iterations of the matching process were taken to capture the different matching and therefore the different possible schools results. The following figures (A1 and A2) show error against number of iterations. Both KS2 and KS4 show a stable error from 20 iterations, so in the interest of minimising processing time 200 iterations seems sufficient.

Figure A1: KS2 error

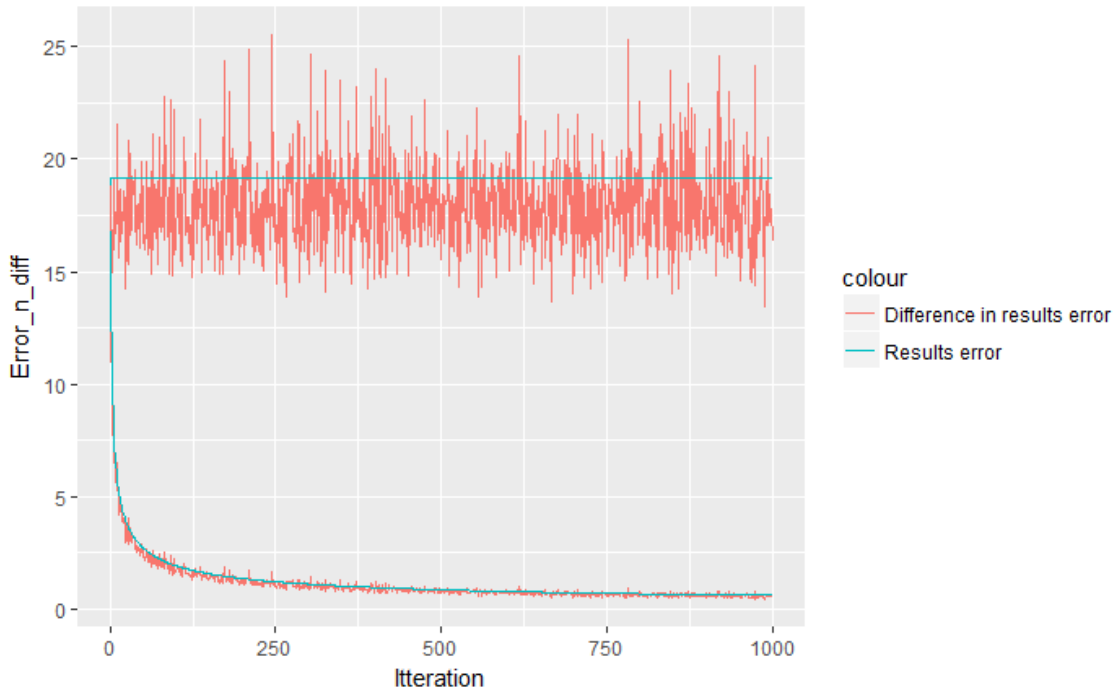
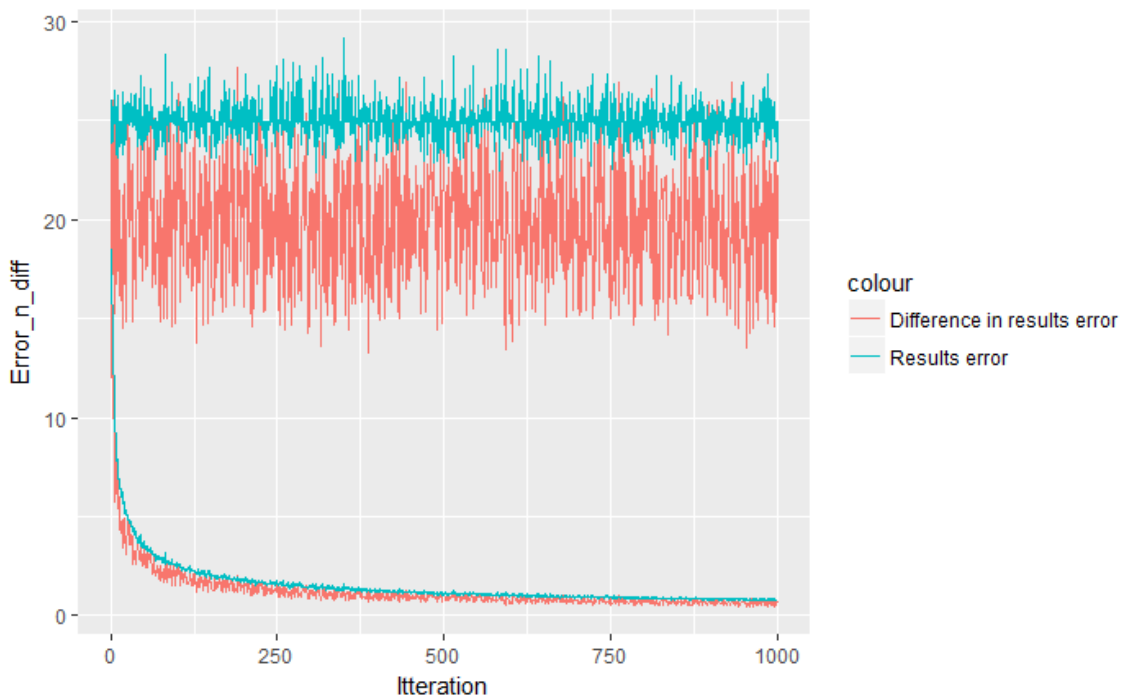


Figure A2: KS4 error



Percentage of sponsored academies matched

Percentage of secondary KS4 sponsored academies matched to a similar LA state funded school, by year of first results.

Year of first results	Number of sponsored academies matched	Number of sponsored academies	Percent matched (%)
2011	43	57	75
2012	31	42	74
2013	39	49	80
2014	40	64	63
2015	33	61	54
2016	40	59	68
2017	25	42	60
2018	49	67	73

Percentage of primary KS2 sponsored academies matched to a similar LA state funded school, by year of first results.

Year of first results	Number of sponsored academies matched	Number of sponsored academies	Percent matched (%)
2013	94	101	93
2014	177	267	66
2015	143	280	51
2016	106	210	50
2017	82	145	57
2018	145	206	70

Statistics tests

Statistics test summary KS2.

Number of tests rejected, for further detail see the glossary section of the annex.

PrePost	Key Stage	First Results Year	KS test for propensity score of sponsoreds pre/post	KS test for propensity score of similar LA schools pre/post	KS test for propensity score	t test for propensity score	t test for FSM	t test for BME	t test for KS1APS	t test for Ofsted rating	t test for pupil numbers
Post Match	2	2013	0	200	0	0	0	0	0	200	0
Post Match	2	2014	0	200	144	0	0	0	0	200	0
Post Match	2	2015	0	200	0	0	0	0	68	200	0
Post Match	2	2016	0	200	200	0	0	0	0	200	0
Post Match	2	2017	0	200	200	200	0	0	0	200	0
Post Match	2	2018	0	200	200	200	0	0	0	200	0
Pre Match	2	2013	-	-	1	0	0	0	0	1	0
Pre Match	2	2014	-	-	1	1	1	1	1	1	1
Pre Match	2	2015	-	-	1	1	1	1	1	1	1
Pre Match	2	2016	-	-	1	1	1	0	1	1	1
Pre Match	2	2017	-	-	1	1	1	0	1	1	0
Pre Match	2	2018	-	-	1	1	1	0	1	1	0

Statistics test summary KS2, 2 years after becoming a sponsored academy.

Number of tests rejected, for further detail see the glossary section of the annex.

Pre / Post	KS	First Results Year	KS test for propensity score of sponsoreds pre/post	KS test for propensity score of similar LA schools pre/post	KS test for propensity score	t test for propensity score	t test for FSM	t test for BME	t test for KS1 APS	t test for Ofsted rating	t test for pupil numbers
Post Match	2	2013	0	200	0	0	0	0	0	0	0
Post Match	2	2014	0	200	144	0	0	0	0	0	0
Post Match	2	2015	0	200	0	0	0	0	0	200	0
Post Match	2	2016	0	200	198	0	0	0	0	130	0
Pre Match	2	2013	-	-	1	0	0	0	0	1	0
Pre Match	2	2014	-	-	1	1	1	1	1	1	1
Pre Match	2	2015	-	-	1	1	1	1	1	1	1
Pre Match	2	2016	-	-	1	1	1	0	1	1	1

Stats test summary KS4.

Number of tests rejected, for further detail see the glossary section of the annex.

Pre / Post	KS	First Results Year	KS test for propensity score of sponsoreds pre/post	KS test for propensity score of similar LA schools pre/post	KS test for propensity score	t test for propensity score	t test for FSM	t test for BME	t test for KS2 APS	t test for Ofsted rating	t test for pupil numbers
Post Match	4	2011	0	200	0	0	0	0	0	0	0
Post Match	4	2012	0	200	0	0	0	0	0	0	0
Post Match	4	2013	0	200	0	0	0	0	0	0	0
Post Match	4	2014	0	200	0	0	0	0	0	0	0
Post Match	4	2015	0	200	0	0	0	0	0	200	0
Post Match	4	2016	0	200	0	0	0	0	0	199	0
Post Match	4	2017	0	200	0	0	0	0	0	0	0
Post Match	4	2018	0	200	0	0	0	0	0	200	0
Pre Match	4	2011	-	-	1	1	1	0	1	1	1
Pre Match	4	2012	-	-	1	1	1	1	1	1	1
Pre Match	4	2013	-	-	1	1	1	0	1	1	1
Pre Match	4	2014	-	-	1	1	1	0	1	1	1
Pre Match	4	2015	-	-	1	1	0	0	1	1	1
Pre Match	4	2016	-	-	1	1	1	1	1	1	1
Pre Match	4	2017	-	-	1	1	1	0	1	1	0
Pre Match	4	2018	-	-	1	1	1	1	1	1	1

Stats test summary KS4, 2 years after becoming a sponsored academy.

Number of tests rejected, for further detail see the glossary section of the annex.

Pre / Post	KS	First Results Year	KS test for propensity score of sponsoreds pre/post	KS test for propensity score of similar LA schools pre/post	KS test for propensity score	t test for propensity score	t test for FSM	t test for BME	t test for KS2 APS	t test for Ofsted rating	t test for pupil numbers
Post Match	4	2011	0	200	0	0	0	0	0	0	0
Post Match	4	2012	0	200	0	0	0	0	0	0	0
Post Match	4	2013	0	200	0	0	0	0	0	0	0
Post Match	4	2014	0	200	0	0	0	0	0	0	0
Post Match	4	2015	0	200	0	0	0	0	0	200	0
Post Match	4	2016	0	200	0	0	0	0	0	0	0
Pre Match	4	2011	-	-	1	1	1	0	1	1	1
Pre Match	4	2012	-	-	1	1	1	1	1	1	1
Pre Match	4	2013	-	-	1	1	1	0	1	1	1
Pre Match	4	2014	-	-	1	1	1	0	1	1	1
Pre Match	4	2015	-	-	1	1	0	0	1	1	1
Pre Match	4	2016	-	-	1	1	1	1	1	1	1

Glossary

ROC chart (receiver operating characteristic curve) shows the accuracy of the model. The ROC curve plots the true positive rate against the false positive rate.

AUC values are the area under the curve of the ROC chart where one is the ideal value and represents good accuracy.

McFadden values are a pseudo R^2 value (proportion of variation in data explained by the model), but must be interpreted with in context of the other measures.

KS test is a non-parametric test used to compare probability distribution of samples. The null hypothesis is that the samples being tested are similar.

Student t-test is a statistical test used to with null hypothesis that the sample being tested have a similar mean, with data following the Student's t distribution.



Department
for Education

© Crown copyright 2019

Reference: DFE-RR900

ISBN: 978-1-78105-988-3

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at:

adam.hatton@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications