Department for Education

Early analysis of English Hubs phonics attainment: 2021/22 data

Research report

May 2023

Harry Shepherd & Ben Fortescue: Department for Education

Government Social Research

Contents

List of figures

List of tables

Introduction

Background

Funded by the Department for Education, the English Hubs Programme (EHP)[1] was established in 2018, with 34 Hubs selected for their expertise in teaching early reading. These hubs work with selected schools in their surrounding area to help them provide excellent teaching in phonics and early reading. They focus on supporting children making the slowest progress in reception and year 1 and help schools to ensure every child is successful, regardless of background, needs or abilities.

Although the programme is concentrated on three aspects of reading – phonics, reading for pleasure and early language – support to date has largely focused on phonics. The EEF identifies that phonics approaches have been consistently found to be effective in supporting younger pupils to master the basics of reading and that teaching phonics is more effective on average than other approaches to early reading (such as whole language), though it should be emphasised that effective phonics techniques are usually embedded in a rich literacy environment for early readers and are only one part of a successful literacy strategy.[2]

The 34 'hub' schools have offered intensive support to a select group of partner schools. Hubs have also provided less intensive support to a large number of other schools. This publication analyses intensively supported schools, known as 'partner schools', only. The terms 'intensively supported' and 'partner' schools are used interchangeably in this report.

As of Summer 2022, there were three waves of partner schools – waves 1, 2 and 3. The below table captures when each wave was recruited and when they graduated from the programme.

Table 1: Partner school waves over time

	2018/19	2019/20	2020/21	2021/22	2022/23
Wave 1		Supported	Supported	Supported	Graduated
Wave 2			Supported	Supported	Graduated
Wave 3				Supported	Supported
Wave 4					Supported

[1] http://www.englishhubs.org/

[2] https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/phonics

Purpose of publication

This publication uses year 1 phonics screening check (PSC) national attainment data from academic year 2021/22 and pre-pandemic (2016/17 to 2018/19) matched with management information data collected from the EHP.

The PSC is a statutory assessment for year 1 pupils (typically aged 6) that confirms whether they have met the expected standard (EXPs) in phonic decoding.[3] All state-funded schools with a year 1 cohort must administer the check. Teachers administer the check one-on-one with each pupil and record whether their response to each of 40 words is correct. Each pupil is awarded a mark between 0 and 40 to determine whether a pupil has met the expected standard, which in the years analysed was 32. The PSC has been administered since 2011/12 but was not administered as normal in 2019/20 and 2020/21, resulting in a data gap for these two years.[4]

This publication analyses how average attainment in the PSC in 2021/22 changed compared to a pre-pandemic baseline for EHP intensively supported schools (partner schools) relative to other schools that did not receive intensive support (non-partner schools), using a range of methods.

Summary of key findings

In 2021/22, compared to before the Covid-19 pandemic (2016/17 to 2018/19), the percentage of pupils meeting the expected standard in the PSC increased by around 1 percentage point amongst partner schools whereas it decreased by approximately 7 percentage points for non-partner schools. A one percentage point change is within the bounds of normal year-on-year fluctuation in attainment, as observed in pre-pandemic years, but the *relative change* of around 7 percentage points between partner schools and non-partner schools, over this time period, represents a substantial change (see Table 3 for figures).

Therefore, schools supported intensively by the EHP outperformed other schools by around 7 percentage points, when comparing the change in year 1 PSC results between pre-pandemic (2016/17 to 2018/19) and 2021/22. Therefore, partner schools appear to have therefore fared better than other schools, on average, in dealing with the negative impacts of the pandemic on PSC attainment. This is possibly due to the EHP.

By using a range of quasi-experimental methods, we suggest that a plausible estimate of the true or unbiased impact of the EHP on average year 1 PSC attainment could be

[3] https://www.gov.uk/government/collections/statistics-key-stage-1

[4] National PSC results are published annually. https://www.gov.uk/government/collections/statistics-key-stage-1

between 4 and 8 percentage points. This is based on the estimated relative change in PSC attainment between partner and non-partner school in our analysis (Table 2), with an adjustment, based on our assessment of the uncertainties around the proportion of the change in performance which could be attributed to EHP. To interpret these results as causal impacts we would have to assume that the results are not biased by variables that we do not or cannot control for. There are two particularly strong assumptions[5] that we made within the analysis.

- Firstly, that in the absence of the programme partner schools would have faced a similarly sharp drop in attainment that other schools experienced in 2021/22 relative to their pre-pandemic starting points.
- Secondly, that partner schools did not implement other initiatives that successfully improved phonics teaching. If this was the case then our estimates would then be capturing the effect of these initiatives, at least in part.

Whilst we cannot assess whether these assumptions hold true for all schools included in our analysis, the range of methods we used and the robustness checks we have undertaken, combined with the size of the estimates, suggest that EHP intensive support is likely to have had a positive impact on PSC attainment as of 2021/22, although we have less confidence in the precise magnitude as it varies across the different methodologies used. These results should be interpreted as early findings. We plan to assess how sensitive our findings are to new data when the 2022/23 PSC results are available.

[5] The term 'strong assumption(s)' is used in this report to mean firm assertions that we would have to make in the analysis to interpret the results as causal, despite not actually being able to evidence that these assertions are accurate with the data available at this time.

Figure 1: Estimated change in proportion of pupils meeting the PSC EXPs/ probability of meeting the PSC EXPs

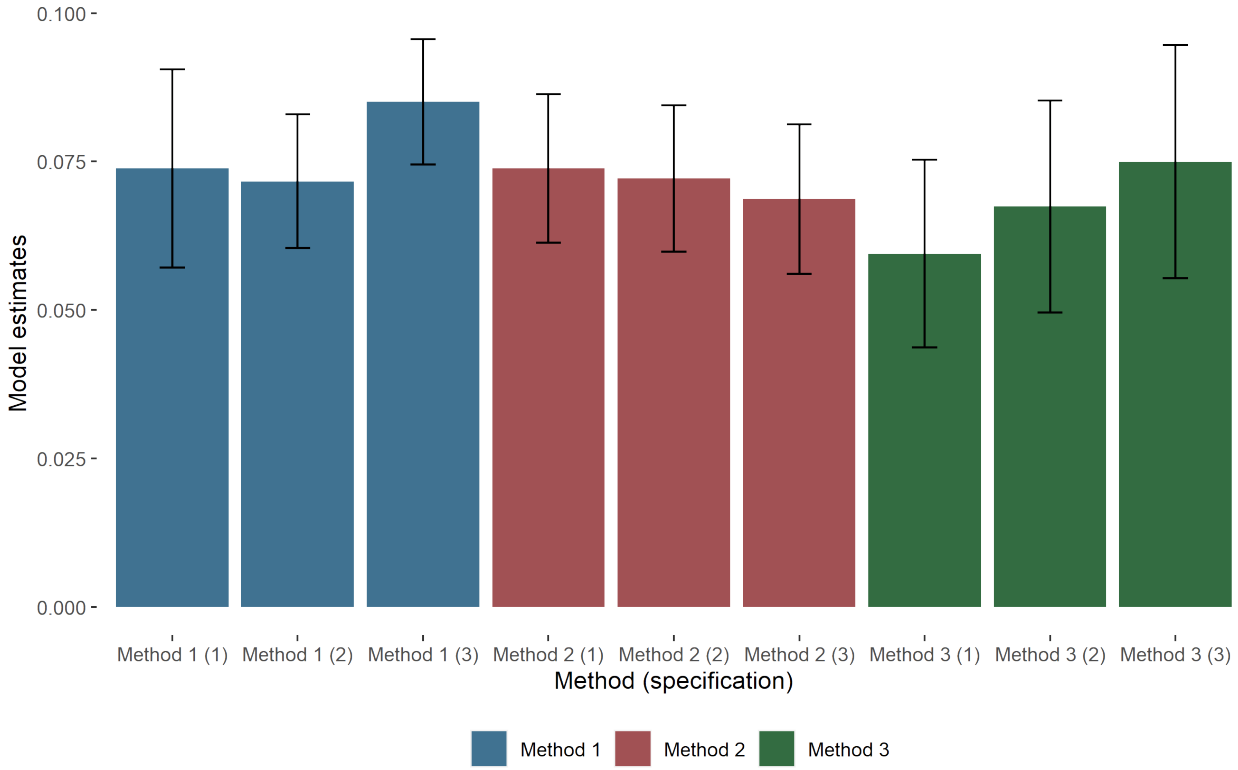


Table 2: Summary of estimated relative change in PSC attainment between partner and non-partner schools

	Measure	Results range (point estimates)
Method 1: Two-period difference-in-differences school-level regression model	Relative change in the percentage of pupils meeting PSC EXPs in year 1 between partner and non-partner schools since the start of EHP	+7.2 to +8.5 ppt
Method 2: School-level two-way fixed effects panel regression		+6.9 to +7.4 ppt
Method 3: Pupil-level difference-in-differences probability model	Relative change in probability of pupils meeting PSC EXPs in year 1 between partner and non-partner schools since the start of EHP	+5.9 to +7.5 ppt

Data sources and processing

The below figure captures the data collection and processing steps that were taken.

Figure 2: Data processing and analysis steps

Step	Description
Data collection & data cleaning	• EHP school level data was collected, so we know which schools received what support in which year. • Data cleaning undertaken, where possible.
Data matching process & cleaning	• School census data was matched with year 1 PSC national curriculum assessment results as well as data from Ofsted. • The EHP management information school level dataset was then matched to the above dataset. • Data cleaning was undertaken and choices of population restrictions made for the main analysis (see the discussion below).
Analysis	• All statistical analyses were performed using R Statistical Software (v4.1.2, R Core Team, 2021). • Because the EHP data has been collected via the programme at the school level, rather than the pupil level, the intention-to-treat (ITT) principle is used where all year 1 pupils in partner schools are assumed to have received support from the EHP.

'Partner schools' are schools in the EHP that have received intensive support from the EHP as of Summer 2022. Non-partner schools are all other schools in our dataset that did not receive intensive support. This group principally comprises schools that have not received any support from the programme, although thousands of schools in that group have received less intensive support from the EHP such support purchasing a validated systematic synthetic phonics programme.[6]

[6] https://www.gov.uk/government/publications/phonics-teaching-materials-core-criteria-and-self-assessment/validation-of-systematic-synthetic-phonics-programmes-supporting-documentation

There are over 1,000 partner schools as identified by EHP management information data in wave 1-3 of the programme. Of these, over 90% could be matched[7] to the PSC data returns in 2021/22 by their school unique reference number (URN).

For the main analysis, we elected for a more restrictive definition of which schools are included for our analysis. The following restrictions were imposed:

- have at least five year 1 pupils at the school;
- have been open continuously since 2016/17 and have not changed their URN over this period[8] and;
- all pupils could be matched to the school Spring census for each academic year.

This provides a balanced panel where all schools are observed in four academic years: 2016/17, 2017/18, 2018/19, and 2021/22. The table below captures how this restricts the number of schools and pupils used in the main analysis. The PSC attainment profiles are approximately the same across the various population definitions shown.

An alternative option would have been to match URNs across time in cases where schools have changed their URN to increase the number of schools that could be used in the analysis. Although it would have also been defensible to take this less restrictive definition, to maximise the number of schools being analysed, we chose the more restricted approach for the main analysis because there is a possibility these schools had changed in some way unrelated to the programme in a way which could create biases within our analysis.

[7] There are several reasons why schools could not be matched between the EHP management information and the PSC data, including incorrect URNs being recorded in the management information, schools changing URNs over time and schools having closed.

[8] See Ofsted's publication for a discussion of school identifiers - https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/666724/Consultation_document_-_changes_to_Ofsteds_statistical_final.pdf

Table 3: Population definitions and sample sensitivity

	2016/17		2017/18		2018/19		2021/22	
Population definition	**Partner school**	**Non-partner school**	**Partner school**	**Non-partner school**	**Partner school**	**Non-partner school**	**Partner school**	**Non-partner school**
No restriction applied								
Number of schools	893	15,468	979	15,420	1,053	15,365	1,018	15,448
Number of pupils	36,457	624,404	39911	620,738	41,686	603,033	37,776	593,331
% meeting PSC EXPs in year 1	75.2%	81.8%	74.6%	83.3%	73.1%	82.8%	75.0%	76.1%
Has 2021/22 & 2018/19 results								
Number of schools	810	12,640	897	13,473	897	13,447	897	13,473
Number of pupils	33,466	515,667	37,168	547,724	35,898	535,495	33,357	525,482
% meeting PSC EXPs in year 1	75.3%	82.3%	74.6%	83.6%	73.2%	83.0%	74.9%	76.1%
Balanced panel								
Number of schools	810	12,623	810	12,623	810	12,623	810	12,623
Number of pupils	33,466	515,599	33,430	515,228	32,538	503,674	30,129	493,867
% meeting PSC EXPs in year 1	75.3%	82.3%	74.7%	83.7%	73.1%	83.1%	74.6%	76.1%
Balanced panel and >=5 pupils								
Number of schools	788	12,096	788	12,096	788	12,096	788	12,096
Number of pupils	33,169	511,961	33,125	511,605	32,054	500,002	30,026	490,695
% meeting PSC EXPs in year 1	75.3%	82.4%	74.7%	83.9%	73.2%	83.2%	74.6%	76.3%

Descriptive analysis

Characteristics of partner schools

The table below illustrates the characteristics of partner schools compared to non-partner schools, restricted to schools that have been open continuously since 2016/17. The key EHP criteria for identifying schools in need of intensive support included low PSC attainment and a high percentage of free school meal (FSM) pupils. It is therefore reassuring that, compared to non-partner schools, partner schools had significantly lower PSC attainment and a higher proportion of FSM pupils before the programme started.

Table 4: School level characteristics means and standard deviations (SD)

	2016/17		2017/18		2018/19		2021/22	
Characteristic	**Non partner**, N = 12,096	**Partner**, N = 788[1]	**Non partner**, N = 12,096[1]	**Partner**, N = 788[1]	**Non partner**, N = 12,096[1]	**Partner**, N = 788[1]	**Non partner**, N = 12,096[1]	**Partner**, N = 788[1]
Pupil numbers	42.3 (24.8)	42.1 (21.7)	42.3 (24.8)	42.0 (21.9)	41.3 (24.3)	40.7 (21.0)	40.6 (23.9)	38.1 (20.1)
% of pupils meeting PSC EXPs in year 1	81.1 (15.2)	75.5 (11.3)	82.5 (14.8)	74.4 (9.7)	81.8 (14.9)	72.7 (11.4)	75.1 (16.9)	74.9 (13.1)
% free school meals	13.2 (12.8)	21.1 (14.5)	13.0 (12.6)	21.4 (14.6)	15.3 (13.7)	24.8 (15.4)	21.4 (16.7)	33.1 (18.3)
% English as an additional language	16.0 (21.9)	18.5 (23.5)	16.2 (21.9)	18.8 (23.6)	16.1 (21.7)	19.2 (23.1)	16.2 (21.0)	19.7 (22.9)
% Special Educational Needs	13.5 (15.2)	15.6 (10.2)	13.5 (15.2)	16.3 (9.9)	13.8 (15.1)	16.6 (10.5)	14.7 (15.3)	16.2 (10.0)
% summer born	33.7 (9.2)	33.6 (8.3)	33.6 (9.1)	34.5 (9.2)	33.3 (9.2)	33.9 (8.8)	33.7 (9.3)	34.3 (8.7)
% male	51.5 (10.3)	51.6 (9.4)	51.4 (10.1)	51.4 (9.1)	51.5 (10.4)	51.6 (9.5)	51.4 (10.4)	50.8 (9.8)
% white British	72.4 (28.5)	70.7 (29.3)	71.9 (28.6)	70.0 (29.7)	71.1 (28.7)	69.4 (29.4)	69.2 (28.9)	67.7 (29.7)

[1]Mean (SD)

The table also shows that percentage of pupils meeting the PSC EXPs in year 1 marginally increased in 2021/22 compared to pre-pandemic for partner schools but decreased sharply for non-partner schools over this period.

Given that they had very different average attainment before the pandemic, to better understand the difference in performance between partner and non-partner schools we calculate the **relative change** or 'difference-in-differences' between them.[9] This is calculated using the following steps:

- **Step 1:** compare the attainment in the PSC between partner (treated) schools and non-partner (untreated) schools before the programme started, using an average of 2016/17 – 2018/19 PSC results (difference 1).
- **Step 2:** compare the attainment in the PSC between partner schools and non-partner schools in 2021/22 (difference 2).
- **Step 3:** calculate the difference between difference 1 and difference 2.

This process simply calculates the 2021/22 vs 2016/17 – 2018/19 average change in attainment for partner schools relative to non-partner schools. This calculation yields around a 7 percentage point change.

PSC distributions over time

As can be seen from the school level density plots below, before the Covid-19 pandemic, the attainment distribution and average attainment fluctuated slightly over time for partner schools before the programme started whilst the distribution of attainment for non-partner schools remained broadly similar. This is reflected in the changes to the averages and standard deviations of PSC attainment in Table 3. However, in 2021/22, there was substantial convergence in the distribution of attainment between partner schools and non-partner schools. For both partner and non-partner schools, there was also a higher degree of variance in attainment in 2021/22 as reflected by the change to the shape of the distributions and increases to their SDs.

By comparing average attainment overtime, we can also see that the attainment gap between partner schools and non-partner schools was significantly wider in 2018/19 compared to 2016/17. This may suggest that hubs placed more weight on schools' 2018/19 results when deciding which schools were most in need of support. This also demonstrates that the magnitude of the relative change figure in 2021/22 would be lower

[9] Cf. "Magenta Book Annex A analytical methods for use within an evaluation" for a discussion on difference-in-differences
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879418/Magenta_Book_Annex_A._Analytical_methods_for_use_within_an_evaluation.pdf

if we used 2016/17 as the definition of the pre-pandemic attainment gap, whilst it would be higher if the 2018/19 figure was used instead.

Overall, this analysis shows that partner schools fared better in their pandemic-affected PSC results in 2021/22 than non-partner schools relative to their pre-pandemic starting points.

Figure 3: Distribution of partner and non-partner schools by proportion of pupils meeting the PSC EXPs in year 1

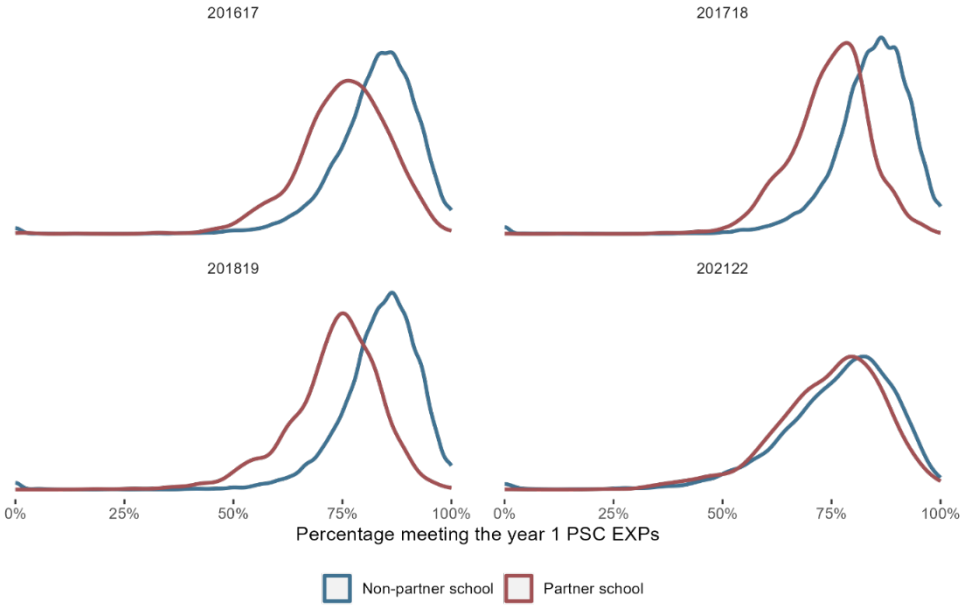


Figure 4: Distribution of partner and non-partner schools by proportion of pupils meeting the PSC EXPs in year 1, including quartile ranges

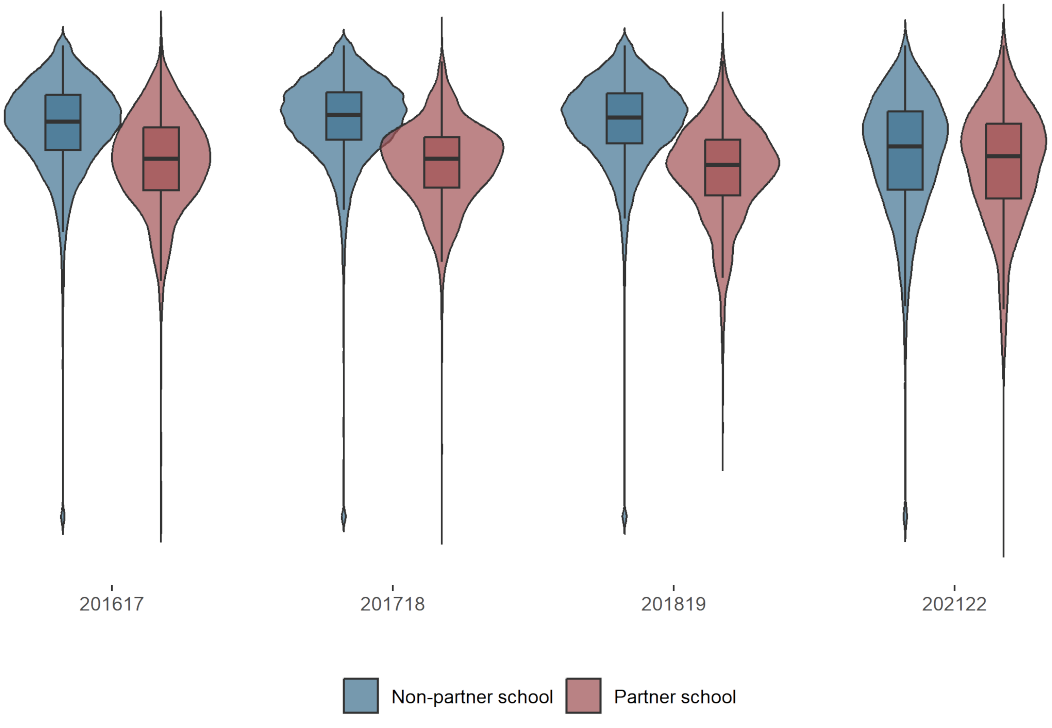


Figure 5: Cumulative distribution of partner and non-partner schools by proportion of pupils meeting the PSC EXPs in year 1

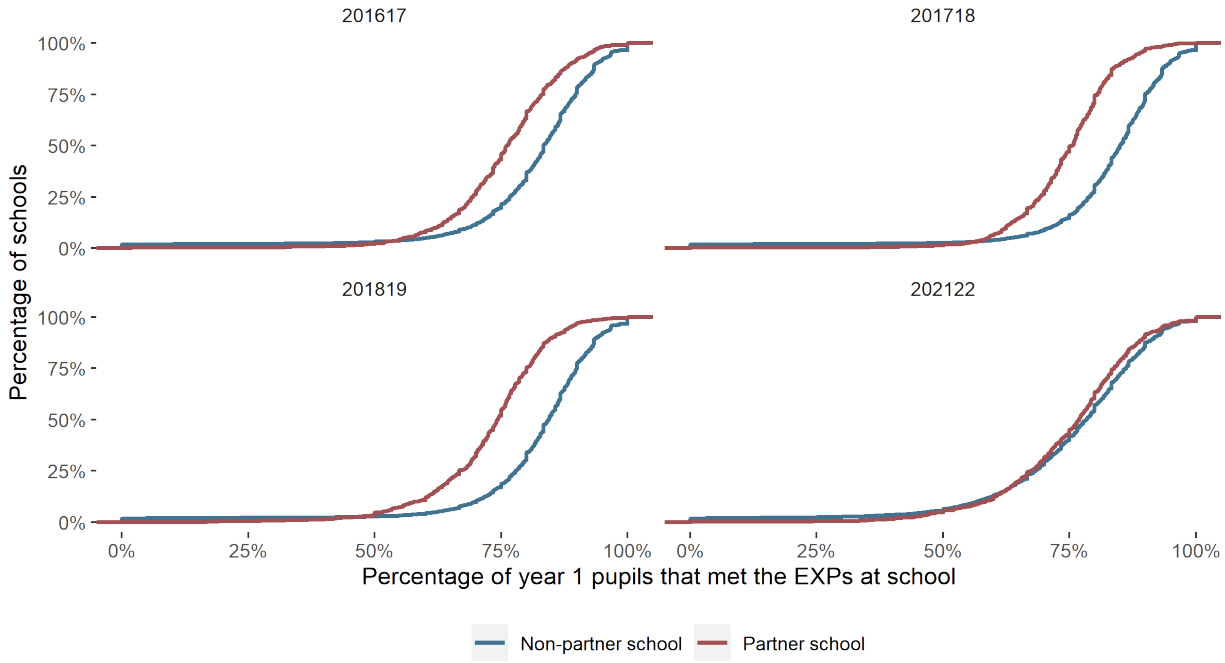
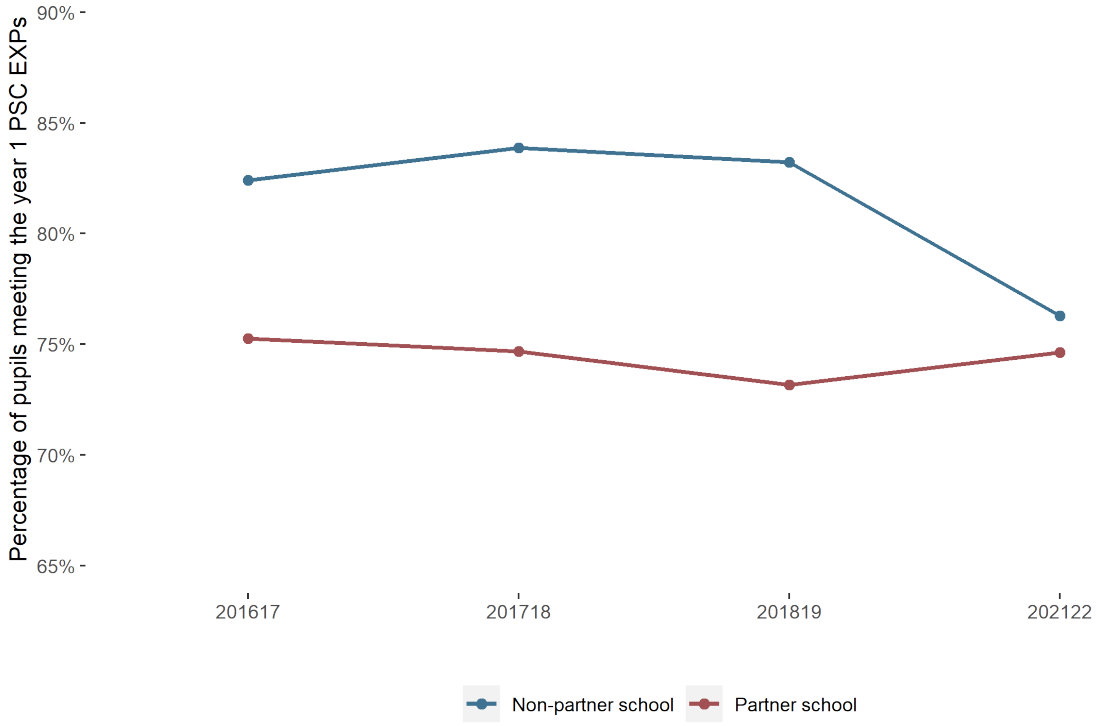


Figure 6: Percentage of pupils meeting the PSC EXPs in year 1 overtime



Quasi-experimental analysis

Purpose

The central goal of this section of the analysis is to estimate a plausible range for the impact of the EHP on PSC attainment, under some strong assumptions, once we control for key observable differences between partner schools and other schools that could be influencing the results.

The ideal 'textbook' methodology for achieving this would involve comparing the outcomes of a set of individuals who were randomly selected into the programme, but who are otherwise identical. This ideal design neither reflects the reality of how the EHP has been implemented, nor the considerable variation in how different classes and schools approach English teaching.

In practice, schools that received support were selected partly on the basis that they had relatively low PSC attainment and wanted to be supported. As a result, we chose to draw upon a range of quasi-experimental techniques to estimate the magnitude of impact, under some strong assumptions. The fundamental aim here is to construct a counterfactual which estimates what partner schools' results would have been in the absence of the programme.

We acknowledge upfront that none of the analysis can definitively conclude that there has been a causal impact of the programme on PSC attainment. However, by drawing on several techniques and evidence sources, we assess that on balance, our provisional results shows that the EHP is likely to have had a positive impact on PSC attainment for partner schools, although uncertainties remain around the exact magnitude of the impact.

Results summary and discussion

A detailed overview of the range of methodologies used is set out in Annex 1, which also summarises their assumptions.

Our results from method 1, a two-period difference-in-differences model, estimate that partner schools outperformed other schools by around 7 percentage points, when comparing the change in year 1 PSC results between pre-pandemic (2016/17 to 2018/19) and 2021/22. Before controlling for any characteristics, this method is the same as the descriptive analysis above and controlling for school characteristics does not substantially change the magnitude of the estimated impact in method 1. Controlling for lagged attainment (i.e., a school's prior attainment) increase the estimated impact to around 8 percentage points.

Our results from method 2, which uses a two-way fixed effects model, are very similar to method 1.

Method 3, which uses coarsened exact matching conducted at the pupil level rather than the school level, yields the most conservative estimates. Pupils at partner schools are estimated to have made approximately a 6 percentage point increase in the probability of meeting the PSC EXPs in year 1, when comparing the change in year 1 PSC results between pre-pandemic (2016/17) and 2021/22. This method is comparing observably similar pupils between partner schools and other schools *as well as* across time.

As set out in the limitations section, there are good reasons to think that these results may be being biased by other factors we can't control for although the size and exact direction of this bias is less obvious. Taking our results in the round, including the annexed sensitivity analysis, we suggest that a plausible range of impact on PSC attainment as of 2021/22 could be between 4 and 8 percentage points.

Limitations and assumptions

To interpret these results as genuine programme impacts rely on strong assumptions about what would have happened to partner schools' attainment in 2021/22 if they had not been enrolled into the programme. A causal interpretation also requires the assumption that the results are not being biased by variables we do not or cannot control for.

It is impossible to know for certain whether our results are biased and in what direction. On one hand, there are some schools that have received less intensive support from the EHP which are in the 'other schools' (control) group so, if these schools have had some impact, they will be supressing the size of our positive estimates. On the other hand, there are multiple unobserved factors that could be biasing our estimates positively. The way the programme was administered means that schools selected into the programme, which introduces possible omitted variable bias. For example, partner schools may have more motivation to improve their results compared to non-partner schools which may have resulted in a relative PSC attainment improvement in the absence of the EHP. We can see from the descriptive statistics presented that partner schools did worse on average than other schools prior to the Covid-19 pandemic. If partner schools were simultaneously taking part in other initiatives that successfully improved phonics teaching, then our estimates would be capturing the effect of these initiatives, at least in part. In addition, the negative impact of the Covid-19 pandemic on education means that further uncertainty is introduced. It is possible that pupils at partner schools may have faced disproportionately less education disruption from the pandemic, although it isn't obvious why this would be the case. Wave 4 schools, which have not been included in the main analysis, as they were not supported until 2022/23, followed a similarly large drop in attainment in 2021/22 as non-partner schools (Annex 2). This offers some reassurance that the assumption that wave 1-3 partner schools would have faced a similarly large drop in attainment in 2021/22 as other schools, in the absence of the EHP, is plausible.

Ultimately, we cannot definitively rule out that omitted variable bias is driving part or all of the estimates and therefore none of these results can be interpreted as causal. However, given the size of the coefficients across the various methodologies and sensitivity analysis carried out, it seems more likely than not that the EHP is having an impact on PSC results.

These results only capture average treatment effect estimates and hide variation by different waves of the programme, hubs, and subgroups such as FSM status. The results should be considered as early findings. When available, PSC 2023 results will enable us to assess how robust these early findings are to new data. In addition, we will review our methodologies used for this early analysis ahead of 2023 PSC results. This may lead to the selection of slightly different models, such as controlling for pupil absences, which could lead to different estimates simply due to methodology changes rather than data changes.

Annex 1 – quasi-experimental methodology overview

There are multiple ways to estimate programme impact. Different quasi-experimental approaches have various strengths and limitations, require different assumptions, and will yield different estimates.

As mentioned, difference-in-differences models estimate the effect of exposure by using changes over time in a treatment group relative to a control group and, broadly speaking, there is an argument that difference-in-differences is preferable to cross-sectional regression where it is feasible.[10] We therefore use a variation of this central approach for all the models presented. We estimate the average treatment effect of intensive EHP support using a range of methods, all of which ultimately fall within the umbrella methodology of difference-in-differences, and then suggest a plausible range the programme's impact on PSC attainment as of 2022, using this distribution of estimated effects.

We used the following controls (Table 5) and the use of these – which vary by methodology and specification – are made clear in the results tables. Robust standard errors were clustered at the school level.[11]

[10] https://educationendowmentfoundation.org.uk/public/files/Projects/Education_Data_Service_Pilot_Evaluation.pdf

[11] As a sense check for methodology 1, we assessed the sensitivity of our standard errors calculations by using a method suggested by Bertrand et al. (2014), where we simply aggregated the pre-treatment data to its average instead of having three years of pre-treatment data. Although the standard errors marginally increased for this calculation, the estimates remained strongly significant. See the following reference for a relevant discussion. Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. *The Quarterly journal of economics*, *119*(1), 249-275. https://doi.org/10.1162/003355304772839588

Table 5: List of control variables

School level controls	• School type • Whether situated in an urban area • Government Office Region • Latest Ofsted rating
Cohort level controls	• Proportion of pupils on free school meals (FSM) • Proportion of pupils whose first language was one other than English (EAL) • Proportion male • Proportion of pupils with any special educational needs (SEN) • Proportion of pupils who identified as of white British ethnicity • Proportion of pupils summer born[12] • Number of year 1 pupils
Lagged school attainment control[13]	• Mean proportion of pupils meeting the PSC EXPs in year 1 in the last (observable) national PSC.
Pupil level controls	• Whether FSM • Whether FSM at any point in the last six years • Whether receives SEN support • Whether in receipt of an EHCP • Whether EAL • Month of birth • Whether male

Method 1: Two-period difference-in-differences school level regression model

Method 1 overview

This two-period difference-in-differences model is estimated using the below equation. The main coefficient of interest is δ, where the interpretation is simply the 2021/22 vs pre-treatment (2016/17, 2017/18, 2018/19) difference in average attainment for treated (i.e.

[12] Defined as whether born in in the months of May, June, July or August. Cf. IFS (2013) for a discussion on summer born disadvantage. https://ifs.org.uk/publications/when-you-are-born-matters-evidence-england

[13] Not used for all specifications, which is transparent in the results tables. The 'lagged' attainment variable is defined using PSC attainment from the previous observable year. For 2021/22, this was the 2018/19 PSC results due to the PSC being cancelled for the previous two years (2020/21 and 2019/20). For 2018/19 and 2017/18 this lagged attainment is taken from 2017/18 and 2016/17 respectively.

intensively supported) vs untreated (i.e. other schools), after controlling for key sources of possibly confounding observable factors.

$$y_{it} = \alpha + \beta Treat_i + \gamma Post_t + \delta(Treat_i * Post_t) + \omega_{it} + \epsilon_{it}$$

$y =$	Attainment outcome
$\alpha =$	Constant term
$\beta =$	Treatment group specific effect (to account for average permanent differences between treatment and control)
$\gamma =$	Time trend common to control and treatment groups
$\delta =$	Estimated true treatment effect
$\omega =$	Vector of school level controls
$\epsilon =$	Error term (assumed to be independent of the other covariates in the model)

The main identifying assumption of difference-in-differences models is the common trends assumption, which is the assumption that no time-varying differences exist between the treatment and control groups. In the context of the EHP, this amounts to assuming that in the absence of the programme partner schools would have faced exactly the same sharp drop in attainment in 2021/22 as non-partner schools compared to pre-pandemic (pooled data from 2016/17, 2017/18 and 2018/19).

Method 1 results

As shown in the below results, this statistic is insensitive to controlling for school controls. Once we control for lagged attainment, our estimates become larger still.

Table 6: Method 1 regression results

	Mean proportion meeting the PSC EXPs in year 1		
	(1)	(2)	(3)
Post 2021/22	-0.067*** (0.001)	-0.055*** (0.001)	-0.058*** (0.001)
Partner school	-0.076*** (0.003)	-0.054*** (0.003)	-0.039*** (0.002)
Post 2021/22 * Partner school	**0.074*** (0.005)**	**0.072*** (0.005)**	**0.085*** (0.005)**
School cohort controls	No	Yes	Yes
School controls	No	Yes	Yes
Lagged attainment	No	No	Yes
Observations	51,536	51,536	50,794
Length of panel	4 years	4 years	4 years
Adjusted R^2	0.041	0.563	0.622
Note:			*p<0.1; **p<0.05; ***p<0.01

Method 2: School level two-way fixed effects panel regression

Method 2 overview

The two-way (year and school) fixed effects model (also known as a 'generalised difference-in-differences')[14] results are estimated using the following equation.

$$y_{it} = \alpha_i + \lambda_t + \delta Treat_{it} + \gamma X_{it} + \varepsilon_{it}$$

$y =$	Attainment outcome
$\alpha =$	School fixed effect
$\lambda =$	Year fixed effect
$\delta =$	Estimated treatment effect
$X =$	Vector of time varying control variables
$\epsilon =$	Error term (assumed to be independent of the other covariates in the model)

This model has fixed effects for school and time. Including school fixed effects in a linear regression is identical to removing school-specific time averages and then applying pooled ordinary least squares estimation. Including the time fixed effects then removes the temporal changes that have the same effects on all schools. One example of a temporal change that has affected schools is the Covid-19 pandemic (although there will be some variation across schools in practice). This method means that time invariant sources of bias within schools are controlled for, thereby limiting any sources of confounding factors that remain fixed overtime.[15] Possible confounding factors, that can stay fixed overtime *within* schools, include size of school site (i.e. school size), geographic location, and members of schools' leadership teams.[16]

Therefore, the assumption is that a school's attainment outcome before the change is a reasonable counterfactual for what a school's outcome would have been at a later time in the absence of the EHP. The interpretation of the coefficient of interest is the association between becoming a partner school and attainment.

Note that this model uses a different calculation to the first method. It is ultimately very similar to the normal difference-in-differences regression but offers a more flexible

https://www.annualreviews.org/doi/pdf/10.1146/annurev-publhealth-040617-013507

[15] cf. Wooldridge (2021) for a discussion on fixed effects models. https://economics.princeton.edu/wp-content/uploads/2021/08/two_way_mundlak-Wooldridge.pdf

[16] Note that this model does not account for any unobserved sources of bias that varies overtime. For example, if partner schools' leaders and teachers had increasingly higher motivation to improve PSC results since 2018/19, beyond anything induced by the EHP, this would affect our estimates (as with all the methodologies used).

estimation strategy that we intend to extend when we can use more data from 2023 onwards. From 2023 PSC data results, the EHP treatment structure becomes more complicated because there will be different schools receiving different treatments over different periods, and we can observe both 2022 and 2023 post-intervention attainment results. This model is well-suited for such treatment structures.

Method 2 results

These results are very similar to the method 1 results.

Table 7: Method 2 regression results

	Mean proportion meeting the PSC EXPs in year 1		
	(1)	(2)	(3)
Partner school	**0.074*** (0.005)**	**0.072*** (0.005)**	**0.069*** (0.005)**
School cohort controls	No	Yes	Yes
School controls	No	Yes	Yes
Lagged attainment	No	No	Yes
School fixed effect	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes
Weighted by school size	No	No	No
Observations	51,536	51,536	50,794
Length of panel	4 years	4 years	4 years
Adjusted R^2	0.677	0.692	0.697
Note:			*p<0.1; **p<0.05; ***p<0.01

Method 3: Pupil level difference-in-differences probability model (2016/17 – 2021/22 data)

Method 3 overview

This final method employs a difference-in-differences model using pupil level data, instead of school level data, where treatment varies only at the school level. The preferred specification estimates the equation below using a linear probability model (LPM). The coefficient of interest is δ which can be approximately interpreted as the percentage point change in the probability of meeting the year 1 PSC for partner schools relative to non-partner schools in 2022 vs pre-pandemic.

$$y_{it} = \alpha + \beta Treat_{ij} + \gamma Year_t + \delta\left(Treat_{ij} * Year_t\right) + \omega_{ijt} + \epsilon_{ijt}$$

$y =$	Attainment outcome
$\alpha =$	Constant term
$\beta =$	Treatment group specific effect (to account for average permanent differences between treatment and control)
$\gamma =$	Time trend common to control and treatment groups
$\delta =$	Estimated true treatment effect
$\omega =$	Vector of control variables
$\epsilon =$	Error term (assumed to be independent of the other covariates in the model)

In this methodology, we restrict the data to just academic years 2021/22 and 2016/17 and restrict it to schools where a lagged attainment measure is observable (i.e. were open in 2015/16). 2016/17 only was therefore chosen as the pre-intervention baseline difference between partner schools and non-partner schools. Schools were probably more likely to be prioritised becoming a partner school based on their more recent data available (2017/18 and 2018/19) and 2016/17 has the narrowest partner school vs other schools PSC attainment gap, when compared to the other pre-treatment years used in the analysis (2017/18 and 201819). Using 2016/17 as the baseline year therefore provides a more conservative estimate of impact.

We present results in Table 8 that show the sensitivity to using a trimmed sample and different controls (columns 1-3). The final column (4) uses a logistic regression instead. The trimmed sample was generated from the below steps.

Step 1: a school level propensity score (i.e. the likelihood of being a partner school) was calculated based on 2016/17 data using variables, including:

- Proportion of FSM pupils
- Whether school is Ofsted requires improvement or inadequate

- Lagged school level PSC attainment

Step 2: using coarsened exact matching (CEM), pupils at partner schools were matched with observably similar pupils at non-partner schools by year (i.e. for 2016/17 and 2021/22 separately) using 1:1 matching. A range of pupil level and school level variables were used for matching. CEM coarsens the data by breaking continuous data into bins and then achieves exact matching on this coarsened data. Pupils at partner schools were discarded from the analysis where no comparison units could be found. This is the main drawback of this methodology as it trims the dataset.

Step 3: the average treatment effect was estimated using a difference-in-differences regression on the dataset generated from step 2.

The advantage of this procedure is that it reduces the observable (and possibly some unobservable) differences between the treatment and control group. The main disadvantage of this method is that, because the analysis is undertaken on a subsample of pupils and schools, the results may not be generalisable.[17]

Method 3 results

The population used in this analysis is restricted to all year 1 pupils in 2016/17 and 2021/22 at schools that have been open continuously at least since 2015/16 (as we needed to observe lagged attainment). The model specifications in columns (1) and (2) show that being at a partner school is associated with an estimated 6-7 percentage point higher probability of meeting the year 1 EXPs. The estimated impact becomes higher still once we include CEM weights (3).

Note that the non-linear logistic regression model (4) is also shown for transparency. This model produces very similar results once we convert the results into changes in the predicted probability of meeting the EXPs to (2) – known as average marginal effects (AME) on the treated (AME results not shown).[18]

[17] Note that the analysis on the trimmed population does not impose any restriction on whether the same schools are observed in both pre-pandemic and post-pandemic, so some of the schools will be observed in one year but not the other (due to lack of matches).

[18] The main advantages of using a LPM model compared to nonlinear binary response model is ease of interpretation and computational ease. It's computationally more time consuming to estimate the equivalent nonlinear binary response model *and then* present AME estimates in terms of changes to probabilities. AME provides an easier interpretation compared to the log odds functional form presented in column 4. See the following reference for a discussion on this: Horrace, W.C. and Oaxaca, R.L. (2006) 'Results on the bias and inconsistency of ordinary least squares for the linear probability model', *Economics Letters*, 90(3), pp. 321–327. doi:10.1016/j.econlet.2005.08.024.

Table 8: Method 3 regression results

	Whether achieved the PSC EXPs in year 1			
	Linear probability regression			***Logistic regression***
	(1)	(2)	(3)	*(4)*
Post 2021/22	-0.04*** (0.001)	-0.06*** (0.003)	-0.07*** (0.004)	*-0.45*** (0.013)*
Partner school	-0.04*** (0.004)	-0.030*** (0.005)	-0.028*** (0.005)	*-0.221*** (0.018)*
Post 2021/22 * Partner school	**0.059*** (0.006)**	**0.067*** (0.007)**	**0.075*** (0.008)**	***0.465*** (0.027)***
School cohort controls	Yes	Yes	Yes	*Yes*
School controls	Yes	Yes	Yes	*Yes*
Lagged attainment	No	No	No	*No*
Trimmed sample	No	Yes	Yes	*Yes*
Matching weights	No	No	Yes	*No*
Observations	970,088	237,788	237,788	237,788
Note:				*p<0.1; **p<0.05; ***p<0.01

Robustness and sensitivity checks

A number of sensitivity and robustness checks were undertaken on school level data, which are summarised below. Although these checks are non-exhaustive, they provide some reassurance in the headline findings.

Falsification check using wave 4 data. The line chart in Annex 2 illustrates that wave 4 partner schools faced a similarly large drop in attainment than non-partner schools. This is reassuring because wave 4 schools were not selected to be a partner school until after 2021/22 results. Because these schools have broadly similar pre-pandemic starting points than the wave 1-3 partner schools and they still faced a large attainment drop in 2021/22, this strengthens the parallel trends assumptions we make for the interpretation of the main analysis. It is also reassuring that wave 3 results (Annex 2) were slightly lower than wave 1 and wave 2 as these schools have had fewer years of support.

Sensitivity to a lagged dependent variable model. A model that uses 2021/22 data only with lagged attainment is perhaps a more robust model to estimate impact compared to a difference-in-differences model due to three reasons. Firstly, schools were selected into the programme, in part, due to their relatively low (on average) PSC results in 2018/19. Secondly, schools with low PSC attainment are likely to want to reverse this and therefore low PSC attainment can be seen as a proxy for motivation to improve. This motivation is an implicit source of omitted variable bias so controlling for lagged attainment can help to control for this. Finally, a school's average attainment in subsequent years is partly explained by prior attainment. Including a lagged variable

means that if two schools have the same previous average attainment then we can estimate the effect of being a partner school rather than a non-partner school in 2021/22. This model produces estimates within the range of the other models presented.

Sensitivity of the results by wave. For simplicity of presentation, the main results analysed partner schools as a single group rather than by wave of the programme. Extending methodology 1 by assessing the variation by wave shows that there was variation by programme wave. The impacts were broadly similar across different waves (the only statistically significant difference was a slightly smaller impact for wave 3, which is qualitatively consistent with these schools having spent less time in the programme as of 2022). The sensitivity to wave can be seen descriptively in Annex 2.

Assessing whether 'reversion to the mean'[19] could be biasing our results. Given that partner schools' attainment was, on average, lower than non-partner schools before the programme started then, in theory, they may have faced a lower drop in attainment than other schools in 2021/22 simply because they had unusually low attainment pre-pandemic and unusually low attainment at one time period may tend to be followed by attainment closer to the average (mean) in the next. This would undermine the assumption that partner schools would have faced the exact same large drop in attainment in 2021/22 than non-partner schools in the absence of the programme – i.e. the parallel trends assumption.

We assess whether this may be influencing our results descriptively. The bar chart below captures the 'year-on-year'[20] percentage point change in year 1 PSC attainment by quintile (i.e. five equally sized groups), where quintiles are defined by schools' lagged PSC attainment.[21] This is separately presented for the non-partner schools (left) and partner schools (right) populations. It is evident from the chart that regression to the mean is a widespread phenomenon amongst both non-partner schools and partner schools. Average attainment increases markedly year-on-year for those in the lowest quintile of attainment (quintile 1) compared to the middle quintile. Similarly, schools in the highest (quintile 5) face a disproportionately large year-on-year decrease in attainment. Although these results do not undermine the conclusions of the main results, as the patterns are different between partner schools and non-partner schools in 2021/22

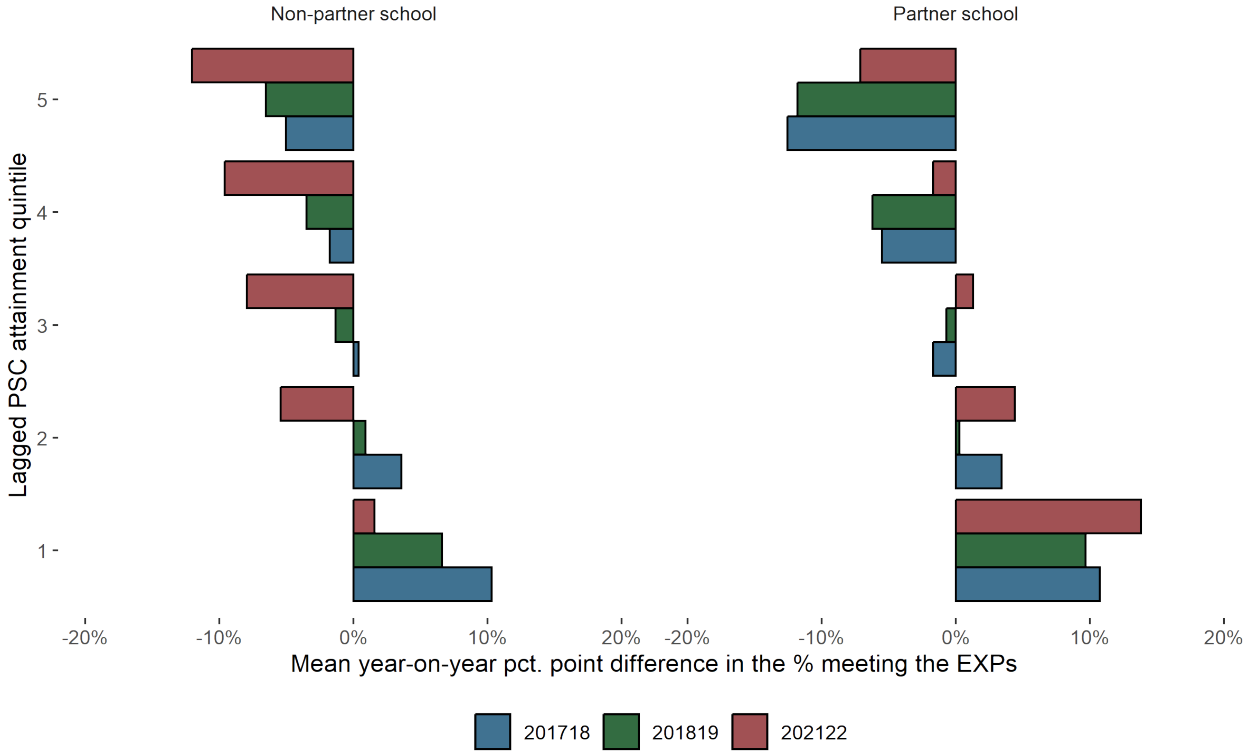
[19] Cf. Smith & Smith (2005) for a discussion of regression to the mean http://economics-files.pomona.edu/garysmith/papers/aveTestScores.pdf

[20] The term 'year-on-year' is used here for ease, but 2021/22 results is not actually a year-on-year interpretation – it compares 2021/22 to the last time periods where the PSC was conducted (2018/19).

[21] The 'lagged' attainment variable, used to construct the quintile categories, is defined using PSC attainment from the previous observable year. For 2021/22, this was the 2018/19 PSC results due to the PSC being cancelled for the previous two years (2020/21 and 2019/20). For 2018/19 and 2017/18 this lagged attainment is taken from 2017/18 and 2016/17 respectively.

compared to the other years, they identify that mean reversion could have some influence on the results.

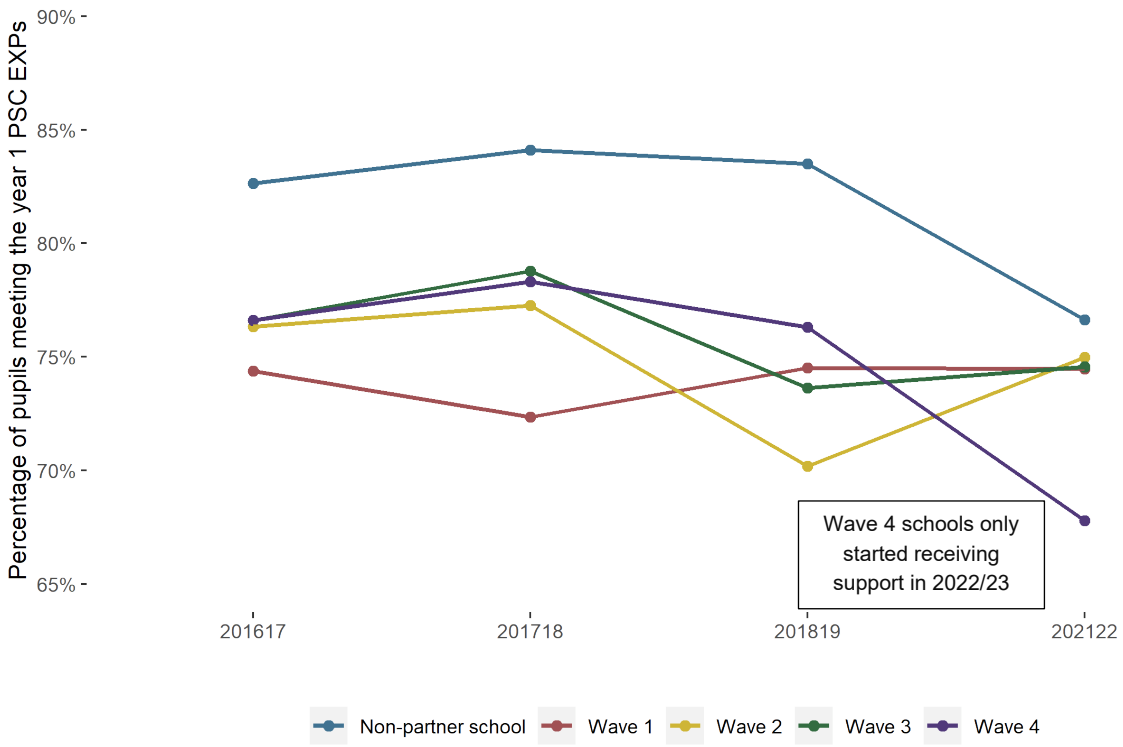
Figure 7: Year-on-year changes to PSC results by quintile of lagged attainment and partner school status



Annex 2 – variation by wave of the programme

The below figure shows variation by wave of the programme. We can see that there was significant variation in attainment by wave before the programme started. In 2021/22, non-partner schools saw a sharp drop in attainment compared to pre-pandemic. Wave 1's attainment remained broadly stable in 2021/22 compared to 2018/19. Wave 2 schools had higher attainment in 2021/22 versus 2018/19 but lower attainment when compared to 2016/17 and 2017/18. Wave 3's attainment followed a broadly similar pattern up to 2018/19, but there was a large increase in attainment in 2021/22 compared to 2018/19. Wave 4 schools – which did not get supported until 2022/23 so *after* the 2021/22 PSC assessment – saw a large drop in attainment between 2018/19 and 2021/22. This drop is roughly comparable in magnitude to the 2021/22 versus 2018/19 decreased faced by non-partner schools, which is reassuring for the main analysis.

Figure 8: Percentage of year 1 pupils meeting the PSC EXPs in year 1 by partner school wave (including wave 4)



Department for Education

About this publication:

enquiries www.education.gov.uk/contactus

download www.gov.uk/government/publications

ISBN: 978-1-83870-475-9

Follow us on Twitter: @educationgovuk

Like us on Facebook: facebook.com/educationgovuk