



Department
for Education

English Hubs Programme phonics impact evaluation 2019 to 2025

Research report

February 2026

**Authors: James Trippett and Chris
Dixon**



Government Operational
Research Service



Government
Social Research

Contents

List of figures	3
List of tables	4
Acknowledgements	5
Executive Summary	6
Introduction	6
Methods	6
Key Findings	7
Discussion	8
Introduction	10
About the English Hubs Programme	10
Phonics pedagogy and assessment	12
Impact evaluation objectives and scope	13
Methods	15
Impact evaluation design	15
School filtering	16
Mitigating selection bias	17
Analysis by cohort	18
Comparison with other interventions	19
Results	20
Supporting analysis	20
Impact by cohort	24
Impact by time since support start	29
Discussion	33
Interpretation	34
Strengths and limitations	43
Conclusion	48
References	51
Annex A: Detailed methods	54
Introduction to quasi-experimental design methods	54
Application of methods to English Hubs Programme	55

Annex B: Further supporting analysis	60
Characteristics of partner schools and pupil populations	60
Distribution of school phonics screening check results over time	63
Annex C: Parallel trends tests	65
First cohort (wave 1)	65
Fifth cohort (AY 23/24)	66
Sixth cohort (AY 24/25)	68
Annex D: Matching assessments	69
Annex E: Impact estimates using Phonics Screening Check score	72
Methods	72
Results	73
Discussion	74
Annex F: Robustness checks	76
Annex G: Effect of methods change	77
Annex H: Criteria for hubs' selection of partner schools	78

List of figures

Figure 1: School PSC results over time for partner and non-partner schools	21
Figure 2: School PSC results over time for partner and non-partner schools by cohort	22
Figure 3: Impact estimates by cohort	25
Figure 4: School PSC results over time for wave 1 and comparison groups	27
Figure 5: Pupil PSC results over time for AY 23/24 cohort and comparison group after matching	28
Figure 6: School PSC results over time for AY 24/25 cohort and comparison group after matching	29
Figure 7: Impact over time on partner schools in AY 23/24 cohort	30
Figure 8: Estimated impact over time on partner schools in wave 1 cohort	31
Figure 9: School distribution of phonics screening check results – wave 1 cohort	63
Figure 10: School distribution of phonics screening check results – AY 23/24 cohort	64
Figure 11: School-level difference-in-differences impact estimates by year – wave 1 cohort	65
Figure 12: Pupil-level difference-in-differences impact estimates by year – wave 1 cohort	65
Figure 13: School-level difference-in-differences impact estimates by year – AY23/24 cohort	66
Figure 14: Pupil-level difference-in-differences impact estimates by year – AY 23/24 cohort	67
Figure 15: School-level difference-in-differences impact estimates by year – AY24/25 cohort	68
Figure 16: Pupil-level difference-in-differences impact estimates by year – AY 24/25 cohort	68
Figure 17: Wave 1 synthetic control weight distribution	71
Figure 18: Mean PSC score impact estimates by cohort	73

List of tables

Table 1: Maximum number of Literacy Specialist days offered to partner schools, yearly equivalent	12
Table 2: English Hubs programme cohort sizes	12
Table 3: School and pupil numbers before and after filtering and matching	23
Table 4: Impact estimates by cohort	25
Table 5: Impact over time on partner schools in AY 23/24 cohort.....	30
Table 6: Estimated impact over time on partner schools in wave 1 cohort	31
Table 7: Impact estimate results summary	38
Table 8: List of control characteristics.....	56
Table 9: List of control characteristics used in the first pre-matching stage of the pupil-level difference-in-differences model	57
Table 10: List of control characteristics used in the second pre-matching stage of the pupil-level difference-in-differences model.....	58
Table 11: Characteristics of pupil populations by cohort before filtering	60
Table 12: Characteristics of school-average pupil populations by cohort after filtering on schools open continuously.....	61
Table 13: Characteristics of school-average pupil populations by cohort after matching for school difference-in-differences method	62
Table 14: Wave 1 synthetic control weight balance table	70
Table 15: Wave 1 synthetic control weight distribution	71
Table 16: Mean PSC score impact estimates by cohort	74

Acknowledgements

We are grateful to Professor Peter Urwin for support with methods and feedback on the report.

With thanks also to Liam Maxwell, Chris Cook and Vikki McAuley for oversight, to Gwilym Owen, James Lawrence and other members of the Department for Education analytical community for support with methods, and to members of the English Hubs data team present and past for the management of programme data.

Executive Summary

Introduction

The English Hubs Programme (EHP) has so far aimed to improve phonics teaching, early language and reading for pleasure since it was established in 2018, particularly for pupils making the slowest progress in reading. This evaluation assesses the bespoke intensive support for phonics teaching provided to selected primary schools ('partner schools'), which has been the main focus of the EHP to date. The 34 English Hubs select partner schools according to criteria including a low percentage of pupils meeting the expected standard in the Phonics Screening Check (PSC) and a high percentage of disadvantaged pupils. The EHP begins support for a new cohort of partner schools each academic year. Intensive support is delivered by Literacy Specialists on 'Literacy Specialist' (LS) days over two to three years; this is in addition to non-intensive programme support including audits, bespoke action plans, resource funding, and Continuing Professional Development¹.

The aim of this impact evaluation is to estimate the effect of EHP intensive support on the percentage of year 1 pupils achieving the expected standard in the PSC. 80% of year 1 pupils nationally met the expected standard in the PSC in 2025, which is a slightly lower percentage than the pre-pandemic peak of 82% in 2018 and 2019².

Methods

This evaluation uses quasi-experimental design methods, which estimate impact by establishing the most appropriate comparison possible between partner schools and non-partner schools in the absence of randomisation. Three methods are used for robustness: difference-in-differences applied at both school and pupil level, and synthetic control. We complete analysis of each programme cohort separately.

There is a possibility that the way hubs select partner schools could lead to impact estimates being inaccurate because it means partner schools are different to non-partner schools in ways that could lead to a difference in PSC results. Hubs select partner schools ready to engage with the programme and partner schools have an unusually low PSC result one year before joining the programme. We control for this pattern of school PSC results in order to mitigate this potential selection bias. This mitigation is only possible for the first, fifth and sixth programme cohorts (labelled 'Wave 1', 'AY 23/24' and 'AY 24/25'³) because the PSC did not take place in 2020 or 2021 due to the COVID-19 pandemic, so we consider impact estimates for these cohorts to be the most reliable. Of

¹ The majority of other (non-partner) primary schools have received some non-intensive EHP support but not bespoke intensive support.

² [Phonics screening check attainment, Academic year 2024/25](#)

³ From the fifth cohort, each cohort is labelled according to the first academic year its partner schools are given support. Cohorts one to four are labelled 'Wave 1' to 'Wave 4'.

those, only the first and fifth cohorts (wave 1 and AY 23/24) have had full intensive support by June 2025.

Key Findings

EHP intensive support is estimated to increase the percentage of year 1 pupils meeting the expected standard in the PSC by three to five percentage points at the end of full intensive support. For a two-form entry year group of 60 pupils, this impact would be the equivalent of two to three additional pupils meeting the expected standard in the PSC. This is roughly equivalent to up to two months' additional progress for pupils, though there are notable limitations in this case with the standard conversion to months' progress which we use (consistent with the Education Endowment Foundation). The estimated impact of three to five percentage points is substantial in relation to the original PSC gap between partner and non-partner schools, which was six percentage points. Comparison with broadly similar teacher professional development interventions show positive findings as discussed below.

The impact estimates indicate that impact of EHP intensive support increases with each year of support and is partially (but not fully) sustained after support stops, with the impact reducing each year after support ends. However, we are not confident that impact reduces after support to the same extent in all cohorts of supported schools. Our clearest view of the extent to which impact is sustained post-support is from the wave 1 partner schools. These schools started support before the COVID-19 pandemic and initially recovered more quickly than non-partner schools when PSC results fell. The smaller impact we see in this cohort subsequently results mainly from comparable non-partner schools returning to pre-pandemic levels, narrowing the gap seen at the point EHP support for wave 1 schools ended. Analysis of future PSC results will show whether this pattern has continued for recent waves or whether this finding simply reflects the challenges of that period.

Looking at specific cohorts of supported schools, EHP intensive support is estimated to increase the percentage of year 1 pupils meeting the expected standard in wave 1⁴ partner schools by between two and three percentage points, measured from the final year of support to three years later. For a two-form entry year group of 60 pupils, this impact would be the equivalent of one to two additional pupils meeting the expected standard in the PSC. With simplified figures for illustration, we might expect around 56,000 out of 75,000 year 1 pupils in wave 1 partner schools taking the PSC between 2022 and 2025 to meet the expected standard without the EHP, and estimate around 1,500 to 2,250 additional pupils met the expected standard as a result of the EHP.

⁴ Partner schools in wave 1 received intensive support from September 2019 to July 2022.

EHP intensive support is estimated to increase the percentage of year 1 pupils meeting the expected standard in partner schools in the fifth cohort ('AY 23/24'⁵) by between three and four percentage points, measured over the first two years of support. That corresponds to approximately one pupil in every class of 30 pupils meeting the expected standard in the PSC who otherwise would not. This is roughly equivalent to up to two months' additional progress for pupils, though there are notable limitations in this case with the standard conversion to months' progress used. Again, with simplified figures for illustration, we might expect around 38,000 out of 50,000 year 1 pupils in AY 23/24 cohort partner schools taking the PSC in 2024 or 2025 to meet the expected standard without the EHP, and estimate around 1,500 to 2,000 additional pupils met the standard as a result of the EHP. As well as in the periods of programme support used for each impact estimate, the two cohorts differed in the amount of support received, partner school characteristics, pre-programme PSC scores, and the context in which support was given, which may have influenced the differences in impact observed. We consider the results of these two cohorts' representative of the impact of the EHP as a whole.

Our estimates of impact show that there has not been a decrease in impact between the start of the EHP and recent waves, since there has been a similar impact on PSC results at the end of full intensive support for the first and fifth cohorts (wave 1 and AY 23/24). This suggests the EHP has been effective both during the pandemic and during post-pandemic recovery. The difference between results of this impact evaluation and those of earlier published analysis are a consequence of a change in methods, mainly an improvement in controls used.

Discussion

Considering the EHP's estimated impact in the context of evidence from similar interventions supports interpretation of the programme's relative effectiveness. While comparisons between interventions are difficult, we see that the estimated impact of the EHP at the point full intensive support ends is similar to that found by a meta-analysis of trials of coaching interventions (Kraft, et al., 2018), and within the range of impacts found by meta-analyses of teacher professional development interventions more generally. Some more specific meta-analyses of professional development interventions find larger overall impacts than those estimated for the EHP, such as those focussed on 'managed professional development' (with precise materials) (Fryer, 2017) and small-scale coaching trials (Kraft, et al., 2018). However, the EHP has a considerably larger scale (2,995 schools⁶) than trials which are typically the basis of such syntheses, and researchers often find a 'scale-up penalty' in the impact of large-scale programmes due

⁵ Partner schools in the AY 23/24 cohort started receiving intensive support during academic year 2023 to 2024.

⁶ 2,995 current or graduated partner schools in the first six cohorts, included in this evaluation.

to recognised implementation challenges of scaling-up interventions; it is also worth considering the limitations of this evaluation noted below.

Since the estimated impact of the EHP is similar to that found by smaller-scale studies of professional development interventions, this suggests the programme maintains a certain level of quality of provision at scale and has some success in adapting support to local schools. Findings from the EHP process evaluation show hubs and partner schools appreciate aspects of the programme which enable support to be effectively adapted to local needs (Department for Education, 2024). The estimated impact of the EHP is comparable to the effect on pupil reading outcomes of having a teacher with 6 to 12 years of experience rather than a new graduate (Kini & Podolsky, 2016; Atteberry, et al., 2015). The strength of evidence for the effectiveness of phonics as a pedagogy suggests the impact of the EHP on pupils' broader literacy learning is likely to be significant.

The impact estimates represent the impact of the programme in the context of the COVID-19 pandemic and subsequent recovery, as measured by the PSC. The methods account to some extent for the effect on PSC results of differences between partner schools and non-partner schools (such as partner school cohorts having a greater proportion of disadvantaged pupils), by comparing PSC results before as well as after support, and controlling for relevant measurable characteristics including a pattern of school PSC results linked to the way hubs select partner schools which are ready to improve. This is demonstrated through parallel trends testing and by the consistency of results across three related but different methods, though some inaccuracy due to selection bias may remain.

The impact estimates do not capture the full benefits of non-intensive EHP support or benefits from teachers who move schools after receiving support. They also do not fully represent the benefits for the lower-attaining target group due to a combination of the impact measure used and that impact estimates reflect the overall impact on all pupils and do not control for pupil prior progress; they also do not reflect additional perceived benefits of EHP intensive support such as reading leads' leadership skills and career development, teachers' confidence and career progression, pupils' enjoyment and engagement with reading, and pupils' writing abilities. See the English Hubs process evaluation for further details (Department for Education, 2024).

The impact estimates correspond to the EHP intensive support as implemented so far. The process evaluation showed hubs and partner schools valued the support being bespoke to each partner school, provided by literacy specialists with particular skills and experience in schools similar to partner schools, and provided to partner schools which hubs consider ready for the programme (Department for Education, 2024). These aspects may be significant in achieving the estimated impact at scale and should be considered as the programme and its implementation develops.

Introduction

About the English Hubs Programme

Programme support

The English Hubs Programme (EHP) aims to improve the teaching of reading. The programme was developed by the Department for Education (DfE) to improve the phonics and early reading outcomes for schools with pupils who fall below the expected level of performance on Phonics Screening Check (PSC) scores or schools in areas of high deprivation.

The 34 English Hubs have provided support to improve teaching and encourage reading for pleasure since 2018. Selected primary schools, called partner schools, are given bespoke intensive support for phonics teaching by a local hub. Intensive support is provided by staff working within the education sector that are considered experts in literacy, called Literacy Specialists, on designated Literacy Support days (LS days) and work directly with reading/phonics leads at supported schools.

Additional support offered by English Hubs to improve phonics teaching has included school-based audits to understand current practice and areas for development, bespoke action plans to guide development and change, and funding for resources of up to £3000 matched funding (or up to £6000 non-matched funding if schools meet more stringent criteria). This is provided to some schools not receiving intensive support as well as partner schools. Hubs also offer Continuing Professional Development (CPD) events for phonics, early language and reading for pleasure, to a larger number of schools. The CPD provided is referred to as 'medium level support' (MLS), distinct from the 'intensive' Literacy Specialist support. More than 10,000 schools have received some form of MLS since September 2020, and approximately 10,000 schools have received funding from the EHP⁷.

The support provided by English Hubs expanded from September 2025 to additional curriculum areas, but this expanded support is not within the scope of this impact evaluation as it began after the most recent data used. The expanded support beginning from September 2025 includes CPD to enhance reading fluency in key stage 2, CPD to embed schools' understanding of the writing framework published 2025, CPD for supporting struggling readers in secondary schools and a pilot of intensive support for secondary schools.

⁷ This is significantly more than half of all state-funded primary schools. There were 16,743 state-funded primary schools in England in academic year 2024 to 2025, see [Schools, pupils and their characteristics, Academic year 2024/25](#)

Partner school selection

Schools are eligible for support if they are state funded, complete the PSC and have engaged in specified preliminary activities. Hubs select eligible partner schools based on prioritisation criteria suggested by DfE and their own judgement. Hubs use information from audits and in some cases from “self-assessment forms, or meetings with school senior leaders, LAs or MATs” (Department for Education, 2024, pp. 50-51) to do this. DfE’s prioritisation criteria include schools with a low percentage of pupils meeting the expected standard in the PSC and a high percentage of children eligible for pupil premium (used as an indication of disadvantage). For further details on DfE’s prioritisation criteria see Annex H: Criteria for hubs’ selection of partner schools, and for the eligibility criteria see Appendix 1 of (Department for Education, 2024). In addition, hubs consider factors such as “buy-in from senior leadership, stability of the school’s leadership team, and their perceived potential for impact ... and schools’ receptiveness to support” (Department for Education, 2024, p. 51).

Programme cohorts

Partner schools typically receive intensive support for two years, and a different group of partner schools (‘cohort’) joins the programme each academic year. Some partner schools in the third cohort (wave 3) onwards receive limited additional intensive support after ‘graduation’ at the end of the initial two-year period of support. Table 1 details the annual equivalent maximum number of LS days partner schools in each cohort can receive in each year.

Schools in the first four cohorts (waves 1 to 4) received support throughout the relevant academic years; from September 2023, new partner schools could join the programme at any point in the year and are each eligible for support for up to two calendar years. We label programme cohorts starting from September 2023 according to the first year they received support, to distinguish them from earlier cohorts; the fifth cohort is ‘AY 23/24’, referring to academic year 2023 to 2024. Schools may leave the programme before graduation (‘withdraw’) or have their programme support stopped early by the supporting hub (be ‘exited’).

Table 1: Maximum number of Literacy Specialist days offered to partner schools, yearly equivalent⁸

Cohort	2019/20	2020/21	2021/22	2022/23	2023/24	2024/25
Wave 1	3	3	3	0	0	0
Wave 2	-	4.5	4.5	0	0	0
Wave 3	-	-	4.5	4.5	3	0
Wave 4	-	-	-	4.5	7.5	3
AY 23/24 ⁹	-	-	-	-	6	6
AY 24/25	-	-	-	-	-	6
PSC took place	No	No	Yes	Yes	Yes	Yes

Table 2: English Hubs programme cohort sizes

	Wave 1	Wave 2	Wave 3	Wave 4	AY2324	AY24/25
Number of partner schools (current or graduated) ¹⁰	559	291	179	660	794	512

Source: English Hubs Programme data

Phonics pedagogy and assessment

Phonics is a way of teaching aspects of reading and writing. It involves developing pupils' understanding of the relationship between letters and sounds. The Rose Report identified phonics as the most effective method for teaching children to read (Rose, 2006) and the Education Endowment Foundation concluded (Education Endowment Foundation, 2021):

⁸ Orange indicates period of full intensive support; green is for the period following full intensive support, including in some cases limited support following graduation; red is where a cohort has a combination of schools which have graduated and others which have not.

⁹ Cohorts from AY 23/24 onwards are labelled differently as partner schools in these cohorts can start support at any point in the academic year.

¹⁰ Does not include schools which were withdrawn from or exited the programme.

Phonics approaches have been consistently found to be effective ... 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.

A number of independent organisations have published systematic synthetic phonics teaching programmes (SSP) which are resources for teaching phonics. Primary schools can use funding provided through the EHP to purchase SSP programmes which have been validated by the DfE¹¹.

The Phonics Screening Check (PSC), introduced in academic year 2011 to 2012, is a statutory assessment¹² for year 1 pupils¹³. It consists of 40 words and pseudo-words, and pupils are assessed one-on-one. The threshold score for meeting the expected standard is set annually, but has consistently been 32. Note that the PSC was not administered in 2020 or 2021 due to the COVID-19 pandemic.

Impact evaluation objectives and scope

This impact evaluation estimates the effect of the bespoke intensive phonics teaching support provided by the EHP on the percentage of year 1 pupils meeting the expected standard in the PSC, up to June 2025. The effect of other programme support, including strands offered from 2025, is not estimated.

This impact evaluation has the same scope as an early analysis of impact of the EHP published in May 2023 (Shepherd & Fortescue, 2023), but draws on PSC results from 2023, 2024 and 2025 in addition to 2022. The methods used in this report are closely related to those used in the early analysis, but develop and extend them. The methods used in this report are robust in application to outcomes for more than one year – see for instance (de Chaisemartin & D'Haultfoeuille, 2023), and more effectively control for the selection of partner schools. For this reason, results from this report are not directly comparable to those in the earlier analysis; for an appropriate comparison and further details see Annex G: Effect of methods change. The additional years of PSC results enable analysis of how the impact of the programme changes over time for supported schools as well as an estimate of the longer-term impact of the programme.

A process evaluation of the EHP (Department for Education, 2024)¹⁴ has findings on:

- how hubs are delivering intensive support,

¹¹ [Validation of systematic synthetic phonics programmes: supporting documentation](#)

¹² See [Assessment framework for the development of the year 1 phonics screening check](#) and [National curriculum assessments: past test materials](#)

¹³ Pupils not meeting the expected standard in year 1 retake the assessment in year 2.

¹⁴ [A process evaluation of the English Hubs Programme](#)

- what has worked well or not worked well from the perspective of Hub schools and partner schools.

Methods

Impact evaluation design

This impact evaluation was designed to estimate the impact of the EHP on pupils' phonics ability. A randomised controlled trial was not practical as the programme began in 2018. Instead, two quasi-experimental design methods were used. These are methods which aim to establish the best possible counterfactual outcomes (outcomes that would have been observed in a world in which the programme did not take place, but everything else is the same) from existing data. Selecting the most appropriate schools for comparison enables us to attribute differences in outcomes between partner and non-partner schools to an effect of the programme with greater confidence. This is particularly important since the programme targets schools with low PSC results and high levels of disadvantage.

In this evaluation we used difference-in-differences and synthetic control methods. Both these methods have the potential to account for some 'unobservable' differences between programme participants and others as well as 'observable'¹⁵ differences. As such they are the strongest methods available since no natural experiment is possible. Both methods rely on the assumption that partner schools' PSC results would have changed in the same way as non-partner schools' if the programme had not taken place. This is called the parallel trends assumption and if it holds, the methods mitigate the potential bias (effect which makes estimates inaccurate) from unobservable differences. Since it is not possible to establish whether this assumption holds directly, we test whether the assumption holds in the period before the programme. The synthetic control method is less sensitive to this assumption. We applied difference-in-differences methods separately at school level and pupil level (synthetic control can only be applied at school level) and used the combination of three methods for greater robustness, though these methods are related. See Annex A: Detailed methods for an introduction to each of these methods and further detail on our application of them, including a list of characteristics we control for (Table 8).

The outcome measure used is the percentage of year 1 pupils meeting the expected standard in the PSC, for school-level methods. For the pupil-level difference-in-differences method, the outcome measure is whether or not each pupil met the expected standard in the PSC in year 1. This evaluation focuses on the effect on pupils at schools participating in the programme (the 'ATT' or Average Treatment effect on the Treated) rather than the effect on the whole population.

We also applied the same evaluation methods using impact measures based on pupils' PSC score, to aid interpretation and compare impact estimates to those of other

¹⁵ Observable differences are characteristics which can be measured, unobservable differences cannot be measured.

evaluated interventions in education. Results and further detail are in Annex E: Impact estimates using Phonics Screening Check score. There are particular limitations of using pupils' PSC score as an impact measure, so our main interpretation and discussion draw on results using the percentage meeting the expected standard in the PSC.

School filtering

We placed restrictions on the schools we used in the impact evaluation to make the most appropriate comparison between partner and non-partner schools. The significant restrictions used are that schools:

- have at least five year 1 pupils at the school in each year (so that school results used were not unduly affected by differences between pupil cohorts year to year),
- have been open continuously since academic year 2015 to 2016 and have not changed their Unique Reference Number (URN) over this period
- have PSC results for the four assessed years prior to first receiving intensive support from the EHP

This means all the schools used have data for all the years used (from academic year 2015 to 2016 to academic year 2024 to 2025). It also means the evaluation is completed with a smaller number of schools than otherwise. The group of schools used in the analysis is broadly representative of the group of partner schools as a whole, though perhaps not fully representative of partner schools with the highest level of disadvantage in their cohorts; see an assessment in the Strengths and limitations section and detail in Annex B: Further supporting analysis.

Schools change their URN when they change some aspect of governance, such as becoming an academy. Excluding schools which change their URN avoids the potential bias from including schools which experienced a potentially significant change unrelated to the programme in the same period. We chose not to include schools which have changed their URN for this reason even though including them would increase the number of schools used.

We restricted schools used in the differences-in-differences models further so that all schools have PSC results for the four assessed years before support started. This is so that we can mitigate selection bias (see following section).

There are some special schools which received intensive support through the EHP. All school types receiving intensive support are included in the evaluation, and a variable for the broad type of school is included in the controls so that special schools have similar representation in the groups of partner and non-partner schools which are compared.

Most partner schools and many non-partner schools have received non-intensive support through the EHP (more than half of state-funded primary schools in total since

September 2020). We do not exclude non-partner schools from the analysis on the basis of whether they received non-intensive support from the EHP. The need to isolate the impact of intensive support from that of non-intensive support is less important than using an adequate number of schools, and the impact of intensive support is expected to be significantly greater than that of other kinds of support.

Schools which received intensive support for only part of the usual period (withdrawn or exited schools) are excluded from the analysis. This is so that the analysis compares partner schools which received support for the full period for their cohort with schools which did not receive any.

Mitigating selection bias

Partner schools have lower PSC results and a higher proportion of pupils eligible for free school meals than non-partner schools. This corresponds with the explicit prioritisation for the programme. They also have a slightly higher proportion of pupils with Special Educational Needs (SEN) and with English as an Additional Language (EAL). See Table 11 in Annex B: Further supporting analysis for full details of characteristics of partner schools and non-partner schools available in the data.

These differences between the group of partner schools and group of non-partner schools may result in a difference in outcome between the two groups even if the programme had not taken place. For this reason, impact estimates based on a comparison of the groups may be inaccurate unless these differences are accounted for (this is called selection bias). We controlled for the effect of these differences on our impact estimates by restricting our analysis to a selection of each group, chosen so that both resulting groups have a similar level of disadvantage (for example), and other relevant characteristics. This is called matching; a similar improvement can alternatively be achieved by weighting each school or pupil in the analysis.

The characteristics used for matching and weighting include level of disadvantage and SEN in pupil populations, school Ofsted overall effectiveness judgement and schools' average PSC results before the programme; the full list is in Table 8, Annex A: Detailed methods. For the pupil-level difference-in-differences model, we matched schools to establish an appropriate selection and then weighted pupils at selected schools. We used propensity score matching to match schools and entropy balancing (Hainmueller, 2012) to weight pupils. Further details are in Annex A: Detailed methods.

Even after matching on all relevant measurable school and cohort characteristics, parallel trends testing shows that schools participating in the programme had, on average, an unusually low PSC result (for the school) one year before starting support, with a return to a more normal PSC result immediately before joining the programme (see Figure 2). This pattern indicates selection bias. The PSC results one year prior to starting support would typically be the latest available at the time of selection for the programme. Further,

a recent process evaluation of the programme found hubs consistently expressed the importance “to the success of the EHP that they could decide whether schools are ready or in a good position to engage with the EHP” (Department for Education, 2024, p. 50). At the same time, “nearly half of the Partner Schools (49%, N = 85) reported that they were involved in other improvement initiatives alongside the EHP, and of those 85 schools, 51 (61%) were involved with another English focused initiative” (Department for Education, 2024, p. 40). This means some partner schools, considered ready to engage with the EHP, may have been more likely to engage with other improvement initiatives than non-partner schools.

For this reason, we include the year-to-year change in each school’s PSC results in the school characteristics we match on, one to three years prior to joining the programme. This is necessary in order to control for the effect of this difference between partner and non-partner schools on impact estimates. The parallel trends assumption only holds in the period before the programme (which is the only period we can test for) after controlling for school pre-programme PSC trajectory. Further detail in Annex A: Detailed methods.

We also used the synthetic control method for further robustness given the potential selection bias.¹⁶

Analysis by cohort

Both the difference-in-differences and synthetic control methods were applied separately to each programme cohort¹⁷. This is necessary as the pre-programme PSC results for partner schools follow a similar year-to-year pattern for each cohort, with the single-year drop in average PSC results taking place in a different academic year for each cohort (see Figure 2). Both methods establish an impact estimate for each academic year and combine those into a single estimate for each cohort.

We consider the results for the first and fifth (wave 1 and AY 23/24) cohorts the most reliable. Mitigating selection bias in the way described can only be done for schools which have PSC results both immediately before joining the programme and one year earlier. Since the PSC did not take place in 2020 or 2021 due to the COVID-19 pandemic, only the first, fifth and sixth cohorts (wave 1, AY 23/24 and AY 24/25) have the necessary pre-programme PSC results to fully control for the apparent selection bias. We applied the same methods without matching on trajectory as part of our analysis, for

¹⁶ The synthetic control method has the potential to address issues with differences-in-differences methods resulting from having a limited number of non-partner schools which are sufficiently similar to partner schools. Where difference-in-differences methods using pre-matching can be subject to extrapolation bias, synthetic control methods can instead be subject to interpolation bias. Using the combination mitigates the risk of a strong effect of either of these (Kellogg, et al., 2021).

¹⁷ We exclude partner schools from analysis of cohorts which they were not in.

robustness. The impact estimate results are greater when not matching on trajectory, and highest for the wave 2 and 3 cohorts.

Comparison with other interventions

To compare impact estimates of the EHP with those of other interventions, we also calculated the impact as an effect size. An effect size is a standardised measure of impact used to compare impacts between studies of different interventions which otherwise use different impact measures. There are a number of ways to calculate an effect size; the most commonly used in this context compares the impact with the typical spread of pupil outcomes. We calculated effect size using the Cox index for the pupil-level difference-in-differences model and Hedges' g for the school-level models (school-level difference-in-differences and synthetic control); this is consistent with the approach taken by the Education Endowment Foundation (Education Endowment Foundation, 2022)¹⁸. We use effect sizes from the pupil-level difference-in-differences model when comparing with other interventions, since effect sizes from the school-level models are greater as a result of reduced variation in outcomes at school level. For further detail of our approach see Annex A: Detailed methods.

We also converted the impact into a number of additional months' progress made by pupils, for comparison purposes. In order to do this, we applied the same evaluation methods using year 1 pupils' PSC score as an impact measure instead of the percentage of year 1 pupils meeting the expected standard. We converted results based on pupil PSC score into effect sizes (using Hedges' g) and additional months' progress in a way consistent with the approach taken by the Education Endowment Foundation, although there are limitations with that approach in this case. For further detail see Annex E: Impact estimates using Phonics Screening Check score.

¹⁸ See 'EEF statistical analysis guidance (2022)' [Evaluation design](#)

Results

Supporting analysis

Partner school characteristics

Partner schools tend to have a lower percentage of year 1 pupils meeting the expected standard in the PSC and cohorts with a higher level of disadvantage than non-partner schools. This is consistent with the programme prioritisation criteria. Partner schools also have a higher proportion of pupils with Special Educational Needs (SEN) overall, though a slightly lower proportion of pupils with an Education, Health and Care Plan (EHCP, provided to pupils with the greatest need)¹⁹.

These differences between partner schools and non-partner schools are the case for all cohorts; the differences are greater for earlier cohorts than more recent ones. Partner schools in earlier waves also have a slightly higher proportion of pupils with English as an Additional Language (EAL). See Table 11 in Annex B: Further supporting analysis for full details.

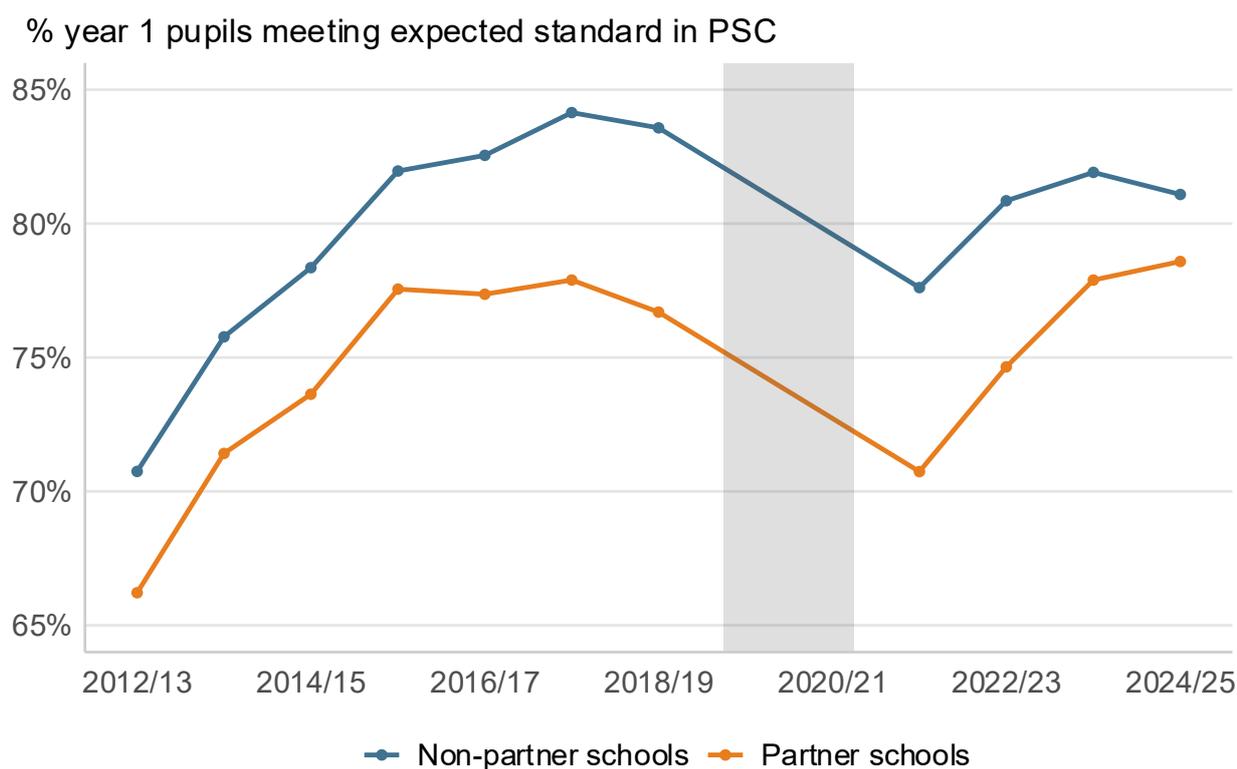
Phonics Screening Check results over time

The average of partner school PSC results increases in a very similar way to non-partner schools' from 2013 to 2016; between 2016 and 2019, partner schools' results are fairly stable while non-partner schools' continue to increase slightly, to a lesser extent than previously (see Figure 1); the changes in PSC results over time are more comparable between partner and non-partner schools between 2016 and 2019 after only including schools which were open continuously and did not change URN (as we do when estimating impact). This gives an indication that the parallel trends assumption broadly holds before EHP support starts.

Whilst the average PSC results for both groups overall fell by a similar amount during the pandemic (2019 to 2022), results for partner school cohorts receiving support improve in that period whilst cohorts not yet receiving support fall further than non-partner schools' (see Figure 2). There is a convergence of results of partner and non-partner schools in academic years 2023 to 2024 and 2024 to 2025; by this point the majority of partner schools have received at least some support.

¹⁹ This may result at least partly from inconsistencies in SEN identification between schools and Local Authorities. See, for instance: [Identifying pupils with special educational needs and disabilities - Education Policy Institute](#)

Figure 1: School PSC results over time for partner and non-partner schools



Note: grey area indicates period in which no PSC assessments took place during Covid-19 pandemic.

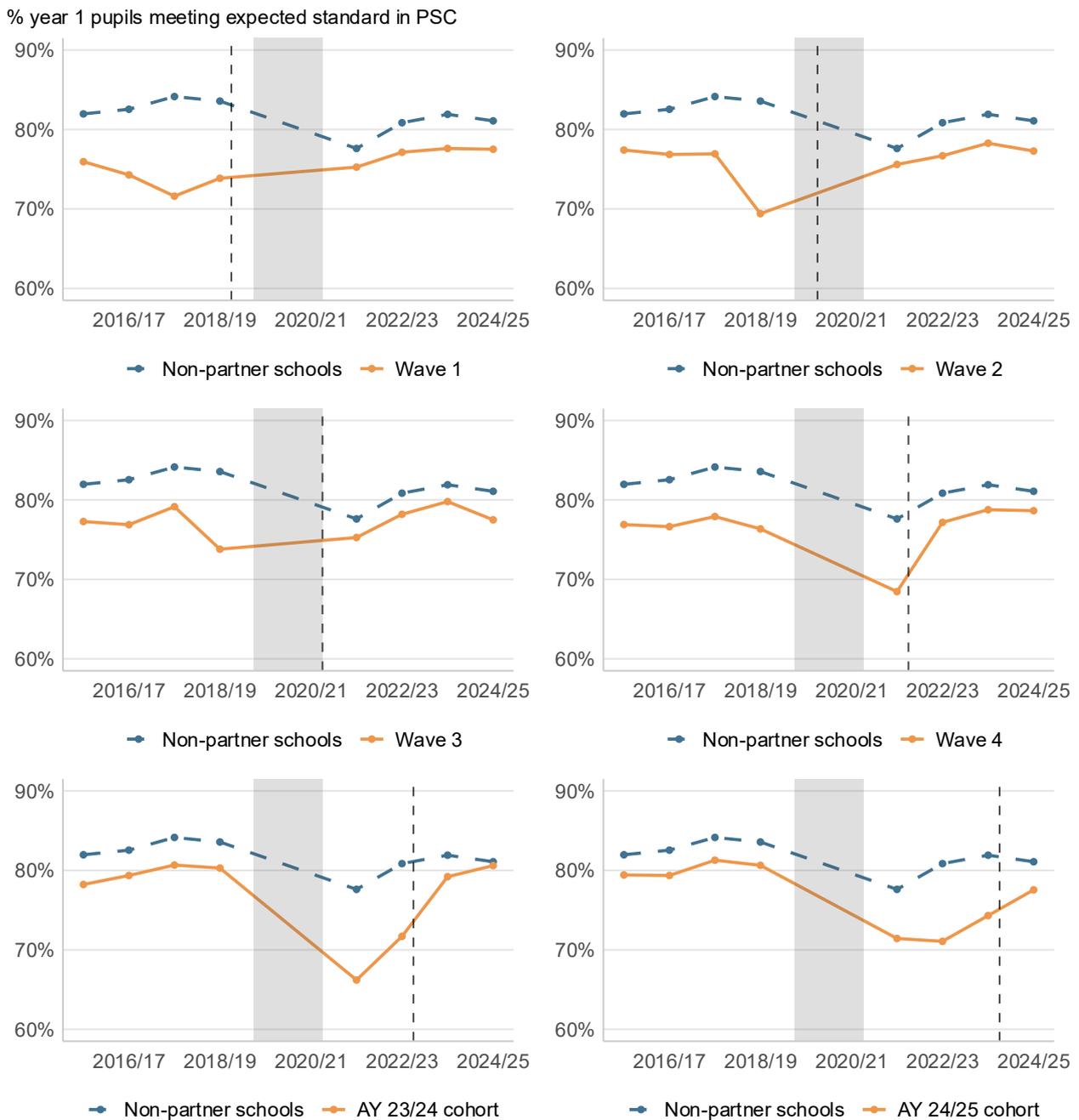
Source: Phonics Screening Check 2013 to 2025²⁰

There is some variation between different cohorts in the way partner schools' PSC results change over time, see Figure 2. The PSC results of most cohorts are lower one year before starting support, indicating the possibility partner schools are selected for recent underperformance relative to their own previous results. Partner schools' PSC results typically return to a typical level immediately before starting support. Steps to mitigate the effect of this apparent selection bias on impact estimates are noted in the Methods section.

²⁰ This figure shows results weighting each school's result equally regardless of school size and using only mainstream schools, since the inclusion of special schools affects the non-partner school average much more than the partner school average. This is in contrast to the impact estimates, which control for the different proportion of special schools making up each group without excluding them; this figure is also different to the published national PSC results which include alternative provision as well as mainstream schools and weights each pupil equally – see [Phonics screening check attainment, Academic year 2024/25](#).

Results are similar to those in the figure when only including schools with results for all years from 2016 to 2025 and at least five pupils, as in the impact analyses.

Figure 2: School PSC results over time for partner and non-partner schools by cohort



Note: grey area indicates period in which no PSC assessments took place during Covid-19 pandemic; dashed vertical line indicates start of programme support.

Source: Phonics Screening Check 2016 to 2025²¹

²¹ As for Figure 1, Figure 2 shows results weighting each school's PSC result equally regardless of school size and using only mainstream schools.

For further information on the distribution of partner and non-partner schools' PSC results by cohort, see Figure 9 and Figure 10 in Annex B: Further supporting analysis. Results of formal parallel trends tests for impact estimates for different cohorts (after controlling for partner schools' lower results one year before joining the programme) are in Annex C: Parallel trends tests.

Sample sizes

The number of schools and pupils we can use to estimate impact of the EHP is smaller after we restrict our analysis to schools with particular characteristics ('filtering', see School filtering section) and match partner and non-partner schools so that the two groups are similar in specified characteristics when estimating the impact of each cohort. The numbers at each stage are in Table 3; see also Table 2. Details of the characteristics of pupils and schools before and after filtering and matching are in Table 11, Table 12, and Table 13 in Annex B: Further supporting analysis.

Table 3: School and pupil numbers before and after filtering and matching²²

Group and cohort	Stage	Number of schools	Number of pupils	% year 1 pupils meeting expected standard in PSC ²³
Non-partner schools	Pre-filter	13,824	534,160	81.3
Non-partner schools	Balanced panel	8,933	373,445	81.9
Non-partner schools	Post-matching (wave 1 analysis)	755	27,566	76.6
Non-partner schools	Post-matching (AY 23/24 cohort analysis)	1,430	54,652	77.1
Non-partner schools	Post-matching (AY 24/25 cohort analysis)	1,000	36,666	77.8
Wave 1	Pre-filter	489	18,514	75.5
Wave 1	Balanced panel	308	12,431	75.3

²² Pre-filter figures are the mean of annual figures in the period 2016 to 2025, excluding 2020 and 2021 (this is why school numbers are different to those in Table 2).

Post-matching figures included are from the pupil-level difference-in-differences model; those from the school-level difference-in-differences model are very similar. School and pupil numbers used in the synthetic control model are the same as those for the balanced panel stage.

²³ % year 1 pupils meeting expected standard in PSC figures are mean of annual school-level percentages.

Group and cohort	Stage	Number of schools	Number of pupils	% year 1 pupils meeting expected standard in PSC ²³
Wave 1	Post-matching (wave 1 analysis)	151	5,504	75.6
AY 23/24	Pre-filter	678	27,491	77.0
AY 23/24	Balanced panel	468	19,860	77.2
AY 23/24	Post-matching (AY 23/24 cohort analysis)	286	10,731	76.5
AY 24/25	Pre-filter	443	17,170	76.9
AY 24/25	Balanced panel	303	12,660	77.0
AY 24/25	Post-matching (AY 24/25 cohort analysis)	200	7,261	77.4

Impact by cohort

We review the impact for each cohort of intensively-supported schools separately. The impact estimates for different cohorts reflect different periods of programme participation. Figure 3 and Table 4 have impact estimates with confidence intervals for the first, fifth and sixth cohorts (wave 1, AY 23/24 and AY 24/25). Results in this section are for the percentage of year 1 pupils meeting the expected standard in the PSC; impact estimates using pupil PSC score are in Annex E: Impact estimates using Phonics Screening Check score.

Figure 3: Impact estimates by cohort

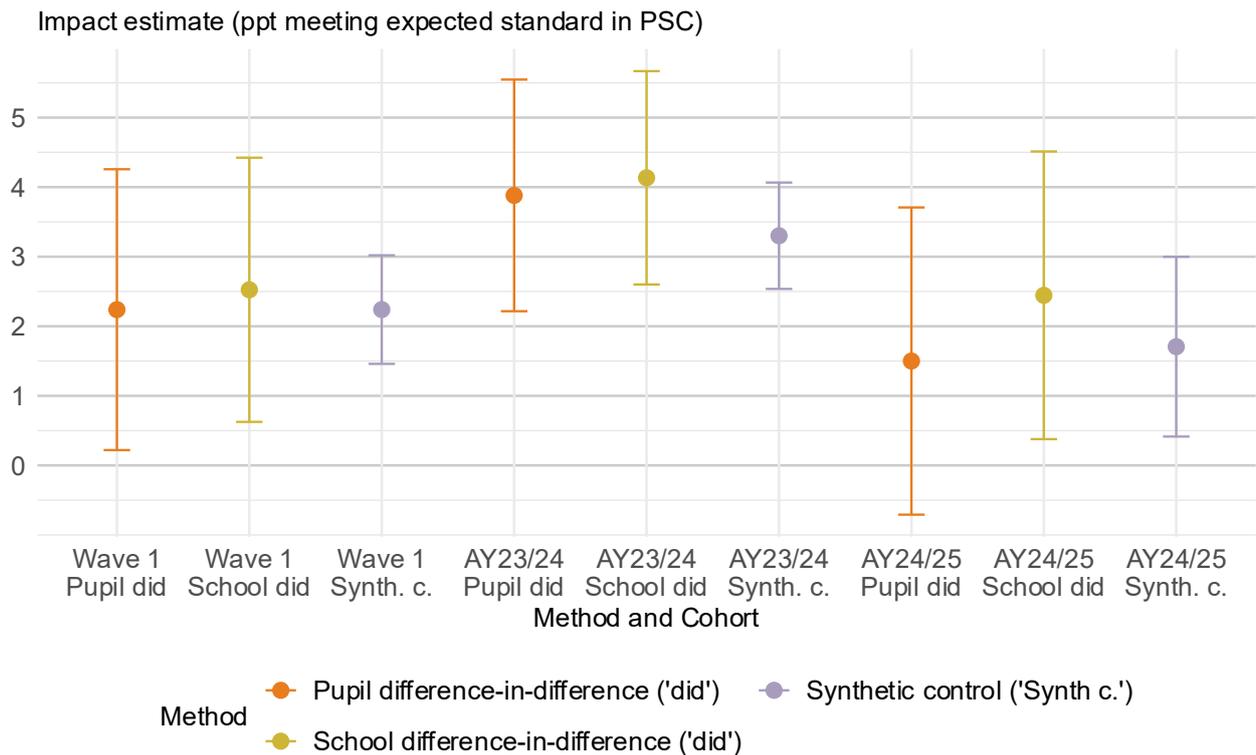


Table 4: Impact estimates by cohort

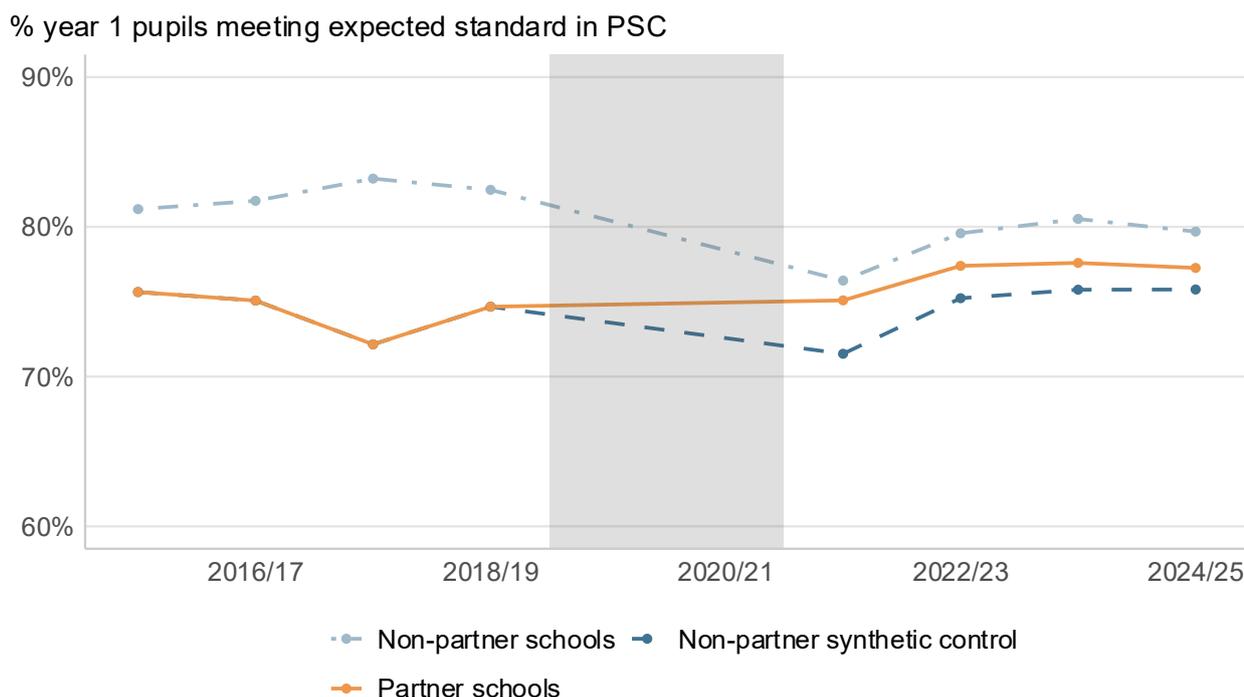
Cohort	Method	Central impact estimate (ppt)	Lower confidence interval (ppt)	Upper confidence interval (ppt)	Effect size (Hedges' g)	Effect size (Cox's Index)
Wave 1	Difference-in-differences (school-level)	2.52	0.63	4.42	0.159	-
Wave 1	Synthetic control	2.24	1.46	3.02	0.141	-
Wave 1	Difference-in-differences (pupil-level)	2.24	0.22	4.26	-	0.074
AY 23/24	Difference-in-differences (school-level)	4.13	2.60	5.67	0.264	-
AY 23/24	Synthetic control	3.30	2.54	4.07	0.211	-
AY 23/24	Difference-in-differences (pupil-level)	3.88	2.22	5.55	-	0.134
AY 24/25	Difference-in-differences (school-level)	2.44	0.38	4.51	0.155	-

Cohort	Method	Central impact estimate (ppt)	Lower confidence interval (ppt)	Upper confidence interval (ppt)	Effect size (Hedges' g)	Effect size (Cox's Index)
AY 24/25	Synthetic control	1.71	0.41	3.00	0.108	-
AY 24/25	Difference-in-differences (pupil-level)	1.50	- 0.71	3.71	-	0.051

First cohort (wave 1)

The estimates of impact of the EHP on the first cohort (wave 1) are 2.24 percentage points (95% confidence interval: 0.22 to 4.26, effect size: 0.074) from the pupil-level difference-in-differences model, 2.52 percentage points (95% confidence interval: 0.63 to 4.42, effect size: 0.159) from the school-level difference-in-differences model, and 2.24 percentage points (95% confidence interval: 1.46 to 3.02, effect size: 0.141) from the synthetic control model. Impact estimates from all models used are statistically significant at the 95% level. These impact estimates reflect impact seen in the period since the end of intensive support provided to this cohort (from June 2022 to June 2025), see Table 1. Figure 4 compares PSC results from wave 1 partner schools with non-partner schools and the synthetic control. Results from the difference-in-differences methods are similar (we show matched or weighted PSC results over time from a different method for each cohort for illustration).

Figure 4: School PSC results over time for wave 1 and comparison groups

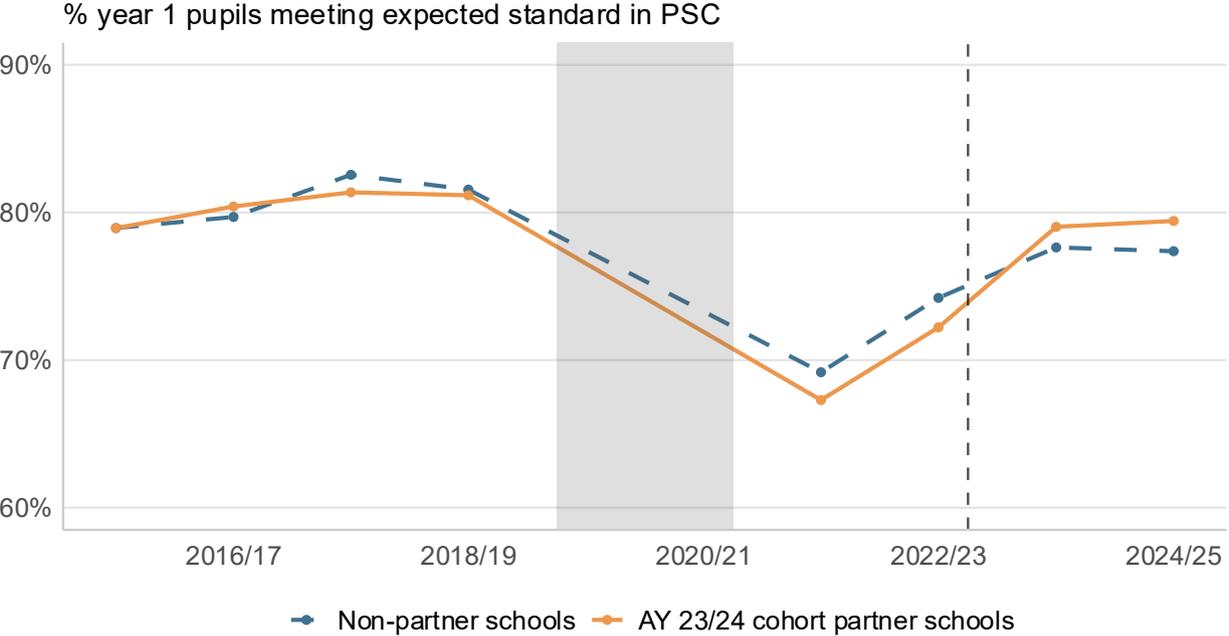


Note: grey area indicates period in which no PSC assessments took place during Covid-19 pandemic.

Fifth cohort (AY 23/24)

The estimates of impact of the EHP on the fifth cohort (AY 23/24) are 3.88 percentage points (95% confidence interval: 2.22 to 5.55, effect size: 0.134) from the pupil-level difference-in-differences model, 4.13 percentage points (95% confidence interval: 2.60 to 5.67, effect size: 0.264) from the school-level difference-in-differences model, and 3.30 percentage points (95% confidence interval: 2.54 to 4.07, effect size: 0.211) from the synthetic control model. Impact estimates from all models used are statistically significant at the 95% level. Figure 5 compares the percentage of year 1 pupils in partner schools and non-partner schools meeting the expected standard after pupil-level matching for the pupil-level difference-in-differences model; the results for other methods are comparable (see Figure 2 for PSC results before matching). Partner schools in the AY 23/24 cohort had received intensive support for between one and two academic years by June 2025, since schools in this cohort started receiving support at different times during academic year 2023 to 2024.

Figure 5: Pupil PSC results over time for AY 23/24 cohort and comparison group after matching

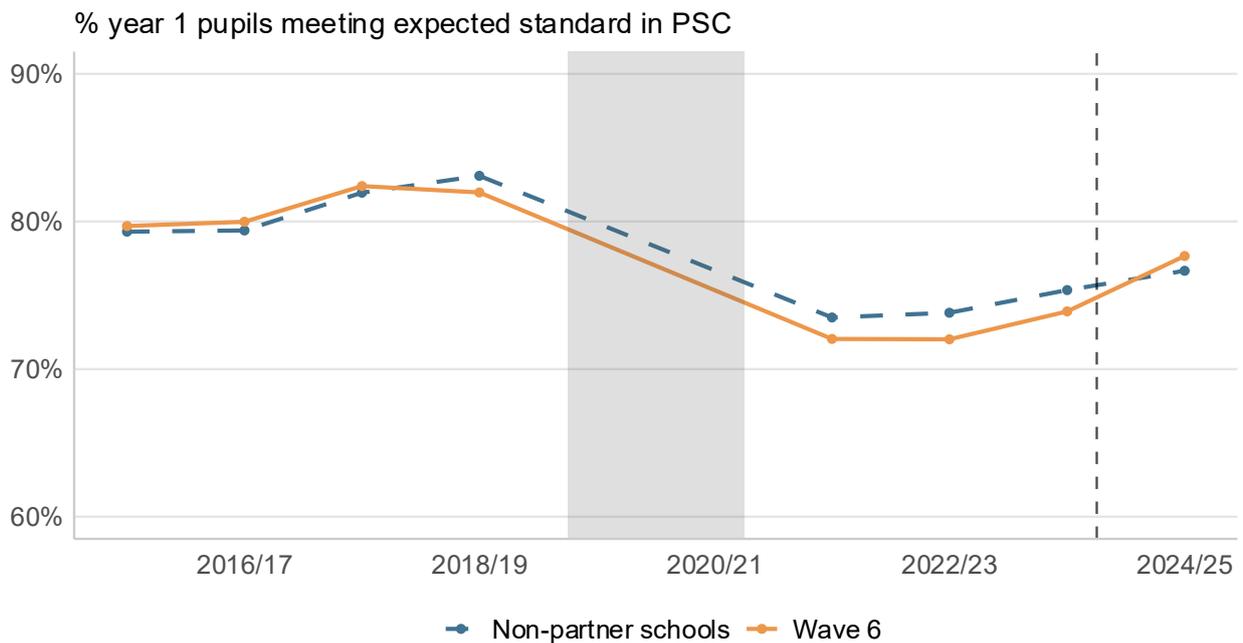


Note: grey area indicates period in which no PSC assessments took place during Covid-19 pandemic; dashed vertical line indicates start of programme support.

Sixth cohort (AY 24/25)

The estimates of impact of the EHP on the sixth cohort (AY 24/25) are 1.50 percentage points (95% confidence interval: -0.71 to 3.71, effect size: 0.051) from the pupil-level difference-in-differences model, 2.44 percentage points (95% confidence interval: 0.38 to 4.51, effect size: 0.155) from the school-level difference-in-differences model, and 1.71 percentage points (95% confidence interval: 0.41 to 3.00, effect size: 0.108) from the synthetic control model. Figure 6 shows PSC results for partner schools and non-partner schools after matching for the school-level difference-in-differences method; the results for other methods are comparable (see Figure 2 for PSC results before matching). Partner schools in the AY 24/25 cohort started support at different times during the academic year 2024 to 2025, and so the impact estimates reflect the impact of up to one academic year of intensive support.

Figure 6: School PSC results over time for AY 24/25 cohort and comparison group after matching



Note: grey area indicates period in which no PSC assessments took place during Covid-19 pandemic; dashed vertical line indicates start of programme support.

Impact by time since support start

We review impact estimates of the fifth cohort (AY 23/24) reflecting impact from the first and second years of support, then impact estimates from the wave 1 cohort reflecting the period from the final year of support onwards. The impact on partner schools in the AY 23/24 cohort increases between the end of the first and end of the second academic year of intensive support from the EHP. See Figure 7 and Table 5 for the impact over time and Table 1 for the timing of support for different cohorts.

Figure 7: Impact over time on partner schools in AY 23/24 cohort

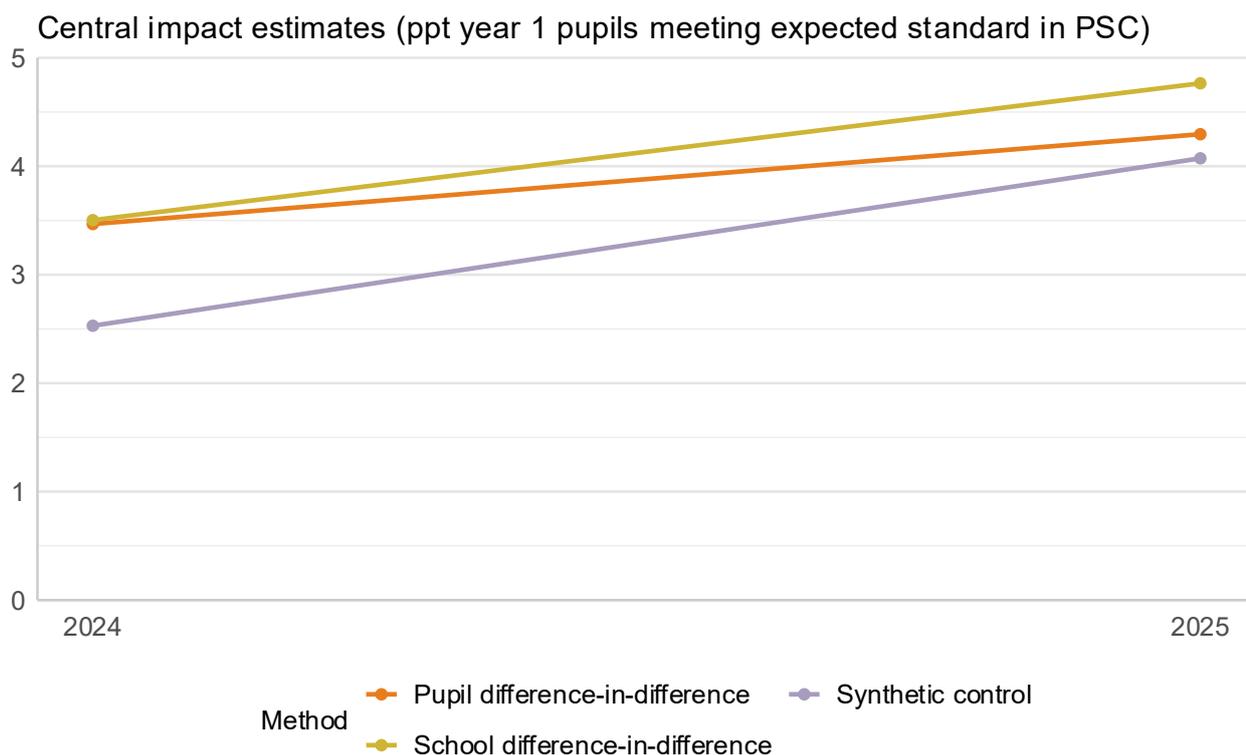


Table 5: Impact over time on partner schools in AY 23/24 cohort

Year	Pupil DiD – Central impact estimate (ppt)	Pupil DiD – Lower confidence interval (ppt)	Pupil DiD – Upper confidence interval (ppt)	School DiD – Central impact estimate (ppt)	School DiD – Lower confidence interval (ppt)	School DiD – Upper confidence interval (ppt)	Synthetic control – Central impact estimate (ppt)
2024	3.47	1.09	5.85	3.50	1.08	5.93	2.53
2025	4.30	1.70	6.89	4.76	2.14	7.39	4.07

The impact on wave 1 partner schools decreases gradually each year since 2022, when they stopped receiving intensive support from the EHP. See Figure 8 and Table 6 for the impact over time and Table 1 for the timing of support for different cohorts. Only the difference-in-differences impact estimates for 2022 are statistically significant on their own, but the impact estimates for each year nevertheless still give an indication of how the impact varies over time.

The reduction in impact between 2022 and 2025 does not necessarily mean there has been a deterioration in PSC results of partner schools in wave 1 in this period. In fact,

this reduction in estimated impact results from a greater improvement in the PSC results of the comparison partner schools in the same period, which are recovering from a much greater deterioration during the COVID-19 pandemic (see Figure 2).

Figure 8: Estimated impact over time on partner schools in wave 1 cohort

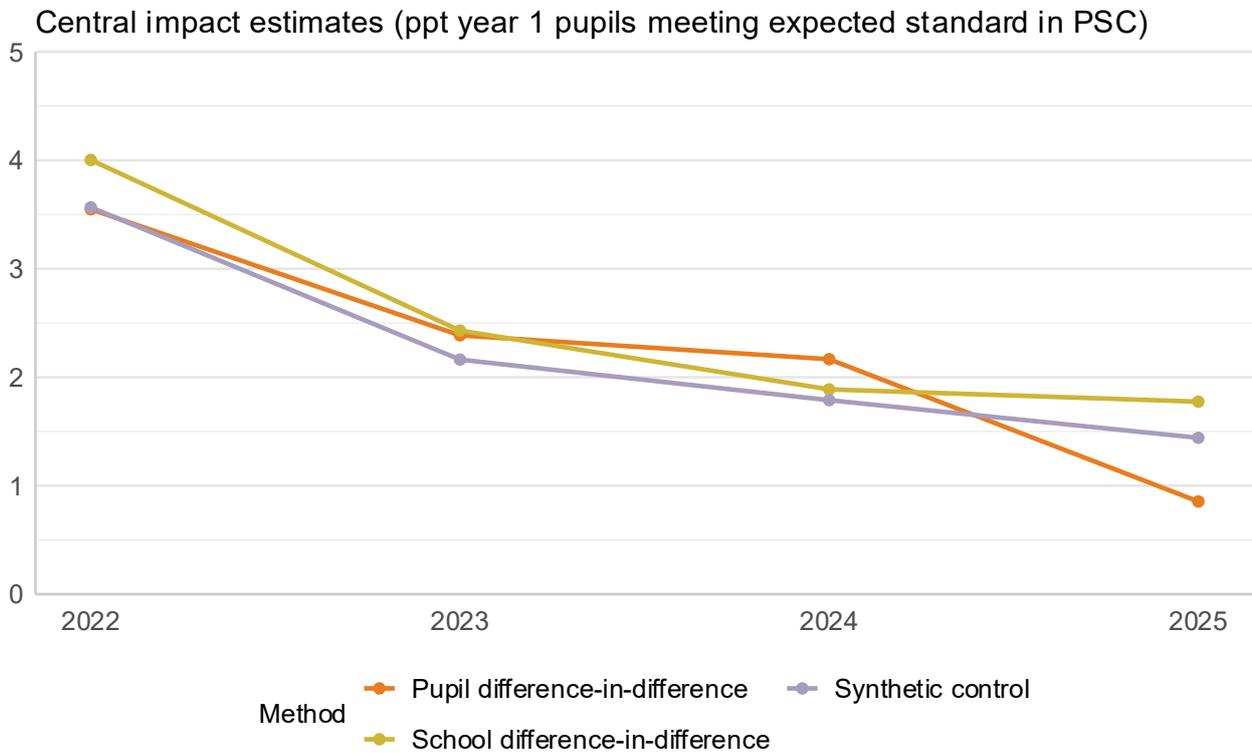


Table 6: Estimated impact over time on partner schools in wave 1 cohort

Year	Pupil DiD – Central impact estimate (ppt)	Pupil DiD – Lower confidence interval (ppt)	Pupil DiD – Upper confidence interval (ppt)	School DiD – Central impact estimate (ppt)	School DiD – Lower confidence interval (ppt)	School DiD – Upper confidence interval (ppt)	Synthetic control – Central impact estimate (ppt)
2022	3.55	0.12	6.98	4.00	0.65	7.36	3.57
2023	2.39	- 0.97	5.74	2.43	- 0.72	5.58	2.16
2024	2.17	- 1.51	5.84	1.89	- 1.49	5.27	1.79
2025	0.85	- 3.14	4.85	1.77	- 1.91	5.46	1.44

The range of central estimates across the three methods used of impact at the end of full intensive support on the percentage of year 1 pupils meeting the expected standard in the PSC is 3.6 to 4.0 percentage points for wave 1 partner schools (graduating in 2022) and 4.1 to 4.8 percentage points for AY 23/24 partner schools (graduating in 2025²⁴). All estimates of impact at the end of full intensive support from difference-in-differences models²⁵ are statistically significant at the 95% level. The corresponding effect sizes for impact at the end of full intensive support from the pupil-level difference-in-differences model are 0.111 for wave 1 partner schools and 0.151 for AY 23/24 partner schools.

²⁴ A small proportion of AY 23/24 partner schools will continue to receive a full level of intensive support after July 2025, which would mean the true impact at graduation may be higher than the estimates.

²⁵ Confidence intervals were not calculated for individual academic years for synthetic control impact estimates.

Discussion

Key findings

EHP intensive support is estimated to increase the percentage of year 1 pupils meeting the PSC expected standard by between three and five percentage points at the point support ends. For a two-form entry year group of 60 pupils, this impact would be the equivalent of two to three additional pupils meeting the expected standard in the PSC. The first and most recent graduating cohort have a similar impact at the point support ends.

EHP intensive support is estimated to increase the percentage of year 1 pupils meeting the expected standard in partner schools in the first cohort (wave 1) by between two and three percentage points, measured from the final year of support to three years later. For a two-form entry year group of 60 pupils, this impact would be the equivalent of one to two additional pupils meeting the expected standard in the PSC.

EHP intensive support is estimated to increase the percentage of year 1 pupils meeting the expected standard in partner schools in the fifth cohort (AY 23/24) by between three and four percentage points, measured over the first two years of support. This corresponds to approximately one pupil in every class of 30 pupils meeting the expected standard in the PSC who otherwise would not.

The impact on wave 1 partner schools is partially sustained and decreases gradually after support ends; it is unclear to what extent this may be the case for more recent cohorts. Wave 1 partner schools started support before the COVID-19 pandemic and initially recovered more quickly than non-partner schools when PSC results fell. The smaller impact we see in this cohort subsequently results mainly from comparable non-partner schools returning to pre-pandemic levels, narrowing the gap seen at the point EHP support for wave 1 schools ended.

A particular limitation is that the impact estimates are unlikely to fully capture the benefits for the lower-attaining target group due to the combination of the measure of impact used and that impact estimates reflect the overall impact on all pupils; they also do not reflect additional perceived benefits of the EHP intensive support such as reading leads' leadership skills and career development or teachers' confidence.

The estimated impact of the EHP is similar to that found by smaller-scale studies of professional development interventions. This suggests the programme maintains a certain level of quality of provision at scale and has some success in adapting support to local schools. The estimated impact of the EHP is comparable to the effect on pupil reading outcomes of having a teacher with 6 to 12 years of experience rather than a new graduate (Kini & Podolsky, 2016) (Atteberry, Loeb, & Wyckoff, 2015).

Interpretation

Impact of full period of intensive support

The estimated impact of the EHP at the end of full intensive support is that 3 to 5 percent of year 1 pupils in partner schools met the expected standard in the PSC who otherwise would not have done. This reflects the benefit of intensive support from the EHP at the point it has accumulated to its greatest extent. For a two-form entry year group of 60 pupils, this impact would be the equivalent of two to three additional pupils meeting the expected standard in the PSC. The range of estimates of impact at the end of full intensive support are similar for both the first and fifth cohorts (wave 1 and AY 23/24)²⁶. This suggests that the EHP intensive support has had a similar effect on PSC results for these two cohorts, and so the effect of EHP intensive support has been consistent between early and recent cohorts.

The range of effect sizes corresponding to these estimates of impact at the end of full intensive support is 0.11 to 0.15. This corresponds to approximately two additional months' progress made on average by each year 1 pupil. However, the method to convert to months' progress has substantial limitations and it is only included for comparison with other studies. Full details for how this is calculated by estimating impact using pupils' PSC scores are in Annex E: Impact estimates using Phonics Screening Check score.

Impact by cohort

First cohort (wave 1)

The estimated overall impact of the EHP on the wave 1 cohort is that 2 to 3 percent of year 1 pupils in partner schools met the expected standard in the PSC who otherwise would not have done. For a two-form entry year group of 60 pupils, this impact would be the equivalent of one to two additional pupils meeting the expected standard in the PSC. The effect size corresponding to this impact estimate for wave 1 overall²⁷ is 0.074. This corresponds to approximately up to one month's additional progress for year 1 pupils in wave 1 partner schools²⁸. However, the method to convert to months' progress has substantial limitations and it is only included for comparison with other studies. This estimate corresponds to the period from the final year of support to three years later (September 2021 to June 2025), so reflects the extent to which the cumulative benefit of intensive support is sustained over a period up to three years after support ends. This shows the impact of an early level of intensive support (up to 9 LS days per partner

²⁶ These are the two cohorts for which it is possible to produce reliable estimates

²⁷ The effect size for the pupil-level model is used as this is comparable to other evaluations.

²⁸ Conversion to months' progress uses impact estimates for pupils' PSC score, see Annex E: Impact estimates using Phonics Screening Check score

school) provided in the context of the COVID-19 pandemic to schools who were chosen to participate first (see school characteristics in Annex B: Further supporting analysis).

The gradual decrease in estimated impact between graduation and three years later seen for wave 1 partner schools shows that impact is sustained to some degree but not fully. Wave 1 partner schools started support before the COVID-19 pandemic and initially recovered more quickly than non-partner schools when PSC results fell (see Figure 4). The smaller impact we see in this cohort subsequently results mainly from comparable non-partner schools returning to pre-pandemic levels, narrowing the gap seen at the point EHP support for wave 1 schools ended. In addition, the PSC results of wave 1 partner schools increase beyond the highest pre-pandemic results (which were in 2016) after support from the EHP, whereas the highest synthetic control post-pandemic PSC results are not notably higher than the highest pre-pandemic results (in 2016). One interpretation is that intensive support from the EHP resulted in partner schools maintaining PSC results through the COVID-19 pandemic and even achieving some improvement in the following years, whereas comparable schools which did not receive intensive support from the EHP took more time to recover from the effect of the COVID-19 pandemic.

The gradual reduction in estimated impact after the end of support could indicate partner schools have some difficulty in maintaining the improvements made as a result of the intensive support. One factor in this could be staff turnover. High staff turnover was seen by hubs as a challenge to delivering the EHP: “Interviewed Hubs stated that the success of the EHP relies on upskilling and development of staff within the school, so progress halts or is lost if staff leave” (Department for Education, 2024, p. 37). Around 9% to 10% of qualified teachers left primary schools annually between academic years 2021 to 2022 and 2023 to 2024²⁹; in addition, movement of teachers from partner schools to non-partner schools, or between school phases, will also affect impact estimates.

Fifth cohort (AY 23/24)

The estimated overall impact of the EHP on the AY 23/24 cohort is that 3 to 4 percent of year 1 pupils in partner schools met the expected standard in the PSC who otherwise would not have done. This impact corresponds to approximately one pupil in every 30-pupil class meeting the expected standard in the PSC who otherwise would not. The effect size corresponding to this impact estimate for the AY 23/24 cohort overall³⁰ is 0.134. This corresponds to approximately up to two months’ additional progress for year 1 pupils in AY 23/24 cohort partner schools. However, the method to convert to months’ progress has substantial limitations and it is only included for comparison with other studies. These estimates reflect impact over the two academic years in which the majority of intensive support expected to be provided by the EHP to schools in this cohort

²⁹ This is the percentage of state funded primary school teachers who left the state funded system. [School workforce in England](#)

³⁰ The effect size for the pupil-level model is used as this is comparable to other evaluations.

took place³¹. This shows the impact of a recent level of support (up to 12 LS days per partner school by summer 2025) in a recent context including continued longer-term recovery from the COVID-19 pandemic, to schools selected more recently which have pre-pandemic PSC results closer to non-partner schools than earlier cohorts (see Annex B: Further supporting analysis).

The increase of impact between the first and second year of intensive support for this cohort indicates that the benefit of intensive support accumulates and possibly takes time to be fully realised.

Sixth cohort (AY 24/25)

The impact estimates for the AY 24/25 cohort are lower than estimates for the other cohorts, and not statistically significant for all models. This corresponds to the fact that partner schools in this cohort have had fewer LS days, provided over a shorter period, than other cohorts (up to June 2025). Intensive support for these schools will continue during academic year 2025 to 2026 and in some cases beyond.

Comparing impact on different cohorts

The effect of the EHP on wave 1 partner schools appears to largely be preventing a deterioration of PSC results during the COVID-19 pandemic, and the effect on the fifth cohort (AY 23/24) partner schools appears to be longer-term recovery from the COVID-19 pandemic. A summary of results and differences between cohorts is in Table 7.

As well as receiving support in different contexts, these two cohorts differ slightly in the amount of support they were given and the characteristics of their cohorts. Partner schools in the AY 23/24 cohort have received slightly more intensive support (up to 12 LS days per school rather than 9). AY 23/24 cohort partner schools have a lower proportion of pupils identified as disadvantaged, with SEN and with EAL, and higher pre-pandemic PSC results than those in earlier cohorts. Impact estimates without controlling for the pre-programme mean PSC are significantly higher than estimates of full methods presented here, which suggests that non-partner schools with pre-programme PSC results comparable to partner schools' find more improvement in the programme period than non-partner schools with higher pre-programme PSC results, though this may be related to the context of pandemic recovery.

We conclude that the effectiveness of the EHP intensive support has not decreased between the first years of the programme and recent waves. Since each cohort's impact estimates correspond to a different period of support, the only directly comparable results for the different cohorts are those of the first and fifth cohorts (wave 1 and AY 23/24) at

³¹ Some partner schools in the AY 23/24 cohort who started support after September 2023 will continue to receive intensive support at the same level for part of the academic year 2025 to 2026 and the impact of that remaining intensive support will not be reflected in these impact estimates; similarly the impact of support offered to partner schools in the year following the first two years of intensive support is not reflected in these estimates

the point the period of full intensive support ends. The range of central impact estimates for each is similar; the range for AY 23/24 is slightly higher, though it is difficult to conclude this difference is significant given the confidence intervals and limitations. Whilst initial estimates of impact in an earlier report using only 2022 PSC results (Shepherd & Fortescue, 2023) were higher than those presented in this report, this is largely due to a change in methods used here, see Annex G: Effect of methods change.

We consider the impact estimates of the wave 1 and AY 23/24 cohorts to be representative of the programme as a whole. It is not possible to control for the apparent selection bias for partner schools' underperformance one year prior to joining the programme when applying the methods used to waves 2, 3 and 4. When we apply the same methods without controlling for schools' pre-programme PSC trajectory (or pre-programme mean PSC result), the results for waves 2, 3 and 4 are comparable to or higher than those for the wave 1 and AY 23/24 cohorts, and the impact estimates of all cohorts are higher than those resulting from the full methods presented here.

Table 7: Impact estimate results summary

Cohort	Range of overall central impact estimates (ppt)	Effect size (pupil-level model)	Maximum total number of LS days per partner school (up to summer 2025)	No of years over which support offered (to summer 2025)	Period of PSC results used to estimate impact, relative to support provision	Partner school characteristics	Context
Wave 1	2.2 to 2.5	0.074	9	3	Final year of support and following three years.	Highest level of disadvantage and SEN (relative to other cohorts). Lowest pre-pandemic PSC results	Support during COVID-19 pandemic.
AY 23/24	3.3 to 4.1	0.134	12	2	First two years of support. Schools started support at different times in the first year.	Lower level of disadvantage and SEN (relative to other cohorts). Pre-pandemic PSC results similar to non-partner schools.	Support during longer-term pandemic recovery.
AY 24/25	1.5 to 2.4	0.051	6	1	First year of support. Schools started support at different times in that year.	Lower level of disadvantage and SEN (relative to other cohorts). Pre-pandemic PSC results similar to non-partner schools.	Support during longer-term pandemic recovery.

Programme impact estimates in context

The English Hubs programme provides professional development support and resources for established teachers teaching an existing part of the curriculum. The EHP impact estimates therefore reflect the combination of:

1. How effective the EHP professional development approach is
2. How effectively the EHP has implemented this professional development at scale (1000s of schools)
3. The extent to which any improvement in teaching due to EHP support has resulted in improvement in pupil attainment for all pupils

In summary, the estimated impact of the EHP at the point full intensive support ends, an effect size of 0.11 to 0.15, is similar to that found by a meta-analysis of trials of coaching interventions using standardised pupil assessments (0.12) (Kraft, et al., 2018). The EHP has a considerably larger scale than these trials (2,995 current or graduated partner schools in cohorts in this evaluation), which suggests the programme maintains a certain level of quality of provision and has some success in adapting support to local schools, overcoming at least to some extent recognised implementation challenges of scaling-up interventions; this is supported by process evaluation findings. The estimated impact of the EHP is comparable to the effect on pupil reading outcomes of having a teacher with 6 to 12 years of experience rather than a new graduate (Kini & Podolsky, 2016; Figure 1 in Atteberry, et al., 2015).

Some approaches in the Education Endowment Foundation (EEF) teaching and learning toolkit³² (which summarises education evidence) have a greater impact than that estimated for the EHP, but it is worth considering that in many cases these are pedagogies whose impact is established in comparison with not using that pedagogy. Some researchers conclude that teacher professional development has the greatest consistent impact of all kinds of interventions in education that were reviewed (Fryer, 2017). The overall effect sizes found by meta-analyses of professional development interventions do include some larger than those for the EHP, such as those of 'managed professional development' (with precise materials, effect size of 0.403) (Fryer, 2017) and small-scale coaching trials (effect size 0.28) (Kraft, et al., 2018). It is worth considering that the estimated impact of the EHP is an aggregate of the impact on all pupils, rather than on those who are the target of the EHP.

Further details behind these conclusions below, and further limitations of this evaluation follow.

³² [Teaching and Learning Toolkit, EEF](#)

Potential impact of teacher professional development

Recent research synthesising studies on the effect of professional development on pupil outcomes find average effect sizes of between 0.05 and 0.12 (Sims, et al., 2021; Blank & de las Alas, 2009; Fletcher-Wood & Zuccollo, 2020; Kahmann, et al., 2022; Lindvall, et al., 2025), and between 0.18 and 0.225 when focussed on studies on the effect of professional development on pupil reading outcomes (Basma & Savage, 2017; Didion, et al., 2019; Rice, et al., 2024).

Researchers have found that effect sizes vary by the type of professional development. One particularly relevant meta-analysis found a summary effect size of 0.403 on pupils' reading achievement from 'managed professional development' approaches, that is including "precise training and curriculum materials that schools and districts can implement to increase teacher effectiveness." (Fryer, 2017). Another meta-analysis focussed on teacher coaching interventions³³ found a summary effect size of 0.18 on pupil attainment (largely from prekindergarten and elementary schools in the US, most using reading outcomes) (Kraft, et al., 2018), although the average effect size was 0.12 when restricted to studies using standardised pupil assessments.

The impact of teacher professional development can be compared with other kinds of intervention in education. One meta-analysis of studies of interventions in education found that 'managed professional development' (as above) had the largest overall effect size of any category of school-based intervention (0.403). This is larger than for high dosage tutoring (overall effect size 0.229) and other categories with overall effect sizes lower than the EHP including pupil incentives, low dosage tutoring, teacher certification, teacher incentives, general professional development, data-driven, extended time, school choice/vouchers, charters, and no excuse charters (Fryer, 2017).

The studies used in these meta-analyses are mostly substantially smaller in scale than the EHP (there are 2,995 current or graduated partner schools in EHP cohorts in this evaluation), and researchers consistently find impacts are lower for larger interventions. There are measurable differences in aggregated impact even when comparing smaller and larger scale studies. In one meta-analysis, studies on professional development with more than 50 randomised units had half the effect size of those with fewer than 50 (Sims, et al., 2021) and in another, studies on coaching with more than 100 teachers had an effect size of 0.10 on pupil outcomes compared with 0.28 for studies with fewer than 100 teachers (Kraft, et al., 2018). An example study is given of a US Florida-wide mandatory coaching intervention with 2,300 reading coaches which found effect sizes on pupil reading over four years of only 0.00 to 0.06, with not all years having statistically

³³ Coaching interventions are defined as "in-service PD [professional development] programs where coaches or peers observe teachers' instruction and provide feedback to help them improve... coaching is intended to be individualized, time-intensive, sustained over the course of a semester or year, context-specific, and focused on discrete skills"

significant impacts (Lockwood, et al., 2010)³⁴. Example reviews quantifying the ‘scale-up penalty’ outside education include scaled-up obesity interventions having 75% or less of the impact found in efficacy trials (McCrabb, et al., 2019), and scale-up penalties from 0 to 50% in preventive interventions with criminological outcomes (Yohros & Welsh, 2019).

Programme implementation

There are particular challenges of maintaining the impact of professional development approaches while implementing at a larger scale. EHP intensive support includes support similar to coaching interventions; for coaching interventions, the challenges of maintaining impact at a larger scale include establishing a large group of coaches with the skills to support the range of needs of teachers and getting engagement of teachers (Kraft, et al., 2018). More generally, there is the dual challenge of maintaining quality at greater scale and additionally maintaining quality while adapting support to local needs. One of the ways the EHP implements support at scale is that literacy specialists work more with partner school reading leads, who in turn work with school teachers.

For implementing teacher professional development interventions more generally, aspects found consistently important across many studies include facilitating teacher buy-in including by demonstrating value to teachers, whether the intervention was accessible and worthwhile for pupils, flexibility to make interventions feasible, providing resources which met teachers’ needs “simply and effectively”, making limited demands of teachers, securing and maintaining leadership support, and aligning interventions with schools’ existing priorities and needs, current practice and aspirations (Sims, et al., 2021).

Many findings from the EHP process evaluation are consistent with these aspects and additionally suggest the EHP was implemented in a way which adapts support to local needs. For instance, hubs tailor support to the needs of the school; drawing on experienced local teachers with a diverse range of skills enables hubs to maintain quality support which meets the diverse needs of different schools; hubs select partner schools ready to engage with the programme, recognising the importance of senior school staff dedication to the EHP (Department for Education, 2024).

Further analysis of how EHP impact varies between different kinds of pupils may help identify the extent to which EHP support is effective for all intended pupils. The impact estimates reported here reflect the average impact on all year 1 pupils in partner schools. As the majority of year 1 pupils meet the expected standard in the PSC (in both partner and non-partner schools), it is likely there is a minority of pupils who benefit more than most from the EHP and this is supported by initial exploratory analysis.

³⁴ A contrasting recent example is the Nuffield Early Language Intervention, for which an effectiveness trial (with 193 randomised schools) and an evaluation of wave two (which had 4,422 registered schools) found similar effect sizes (0.26 and 0.297 respectively) (Dimova, et al., 2020; Smith, et al., 2023)

Comparison with alternative interventions and measures

To aid comparison with other interventions, we also calculated EHP impact estimates using pupil PSC score (in addition to the percentage of year 1 pupils meeting the expected standard), which we can convert to an equivalent of up to two additional months' progress made on average by each pupil resulting from the period of full intensive EHP support (see Annex E: Impact estimates using Phonics Screening Check score). There is a notable limitation of this conversion relating to the way many pupils achieve the maximum PSC score which means it should not be used as a direct representation of impact in itself. We include it, nevertheless, so that the EHP programme impacts can be compared to those of other interventions such as in the EEF teaching and learning toolkit³⁵.

There are three particular aspects of the EHP to consider when comparing the impact with other interventions in the EEF toolkit, which is aimed at practitioners. Many approaches in the EEF toolkit are pedagogies whose impact is established in comparison with not using that pedagogy. The EHP, in contrast, aims to improve the teaching of a pedagogy (phonics) which is to some extent already established. Secondly, the EHP is a large-scale intervention (more than 3000 schools) so effective implementation is dependent on overcoming challenges of scaling up noted in the previous section; this is not the case for many of the research studies drawn on in EEF's toolkit which are smaller-scale trials. Thirdly, the EHP is most beneficial for pupils who would not otherwise meet the expected standard in the PSC, but the impact estimates are an average for the impact on all pupils; again, some effectiveness trials focus on the impact of a specific target subgroup. The Nuffield Early Language Intervention, for example (effect size 0.297, four months' additional pupil progress), was implemented and evaluated for pupils with the weakest language skills in their school cohort (Smith, et al., 2023).

A final comparison is with the measurable impact of teacher experience. One review found the impact of a teacher's first six years of experience on pupil outcomes in English, language and arts (ELA) varies between an effect size of 0.05 and 0.18 between different studies (Figure 1 in (Atteberry, et al., 2015)). Another found studies with pupil outcome effect sizes of 0.08 in ELA for teachers with 12 years' experience (much greater effect in maths subjects), of 0.27 for teachers with 27 years' experience, and of 0.16 for teachers with 15 to 24 years' experience³⁶ (Kini & Podolsky, 2016). Therefore, the estimated impact of the EHP is comparable to the effect on pupil reading outcomes of having a teacher with 6 to 12 years of experience rather than a new graduate.

³⁵ [Teaching and Learning Toolkit, EEF](#)

³⁶ Authors note this study is based on a slightly less preferred study specification.

Strengths and limitations

Accuracy and reliability of impact estimates

Unobservable characteristics and parallel trends assumption

The methods we use are quasi-experimental design methods without any kind of randomisation. As such they may be subject to bias from characteristics which cannot be measured ('unobservable' characteristics) since it is not possible to control for any possible effect on outcomes resulting from differences in these characteristics between partner and non-partner schools. For example, pupils' phonics learning is affected by aspects of themselves or their circumstances other than phonics teaching they receive. We have controlled for potential bias of impact estimates from characteristics which we have data on³⁷ ('observable' characteristics, see Annex A: Detailed methods). Unobservable characteristics which are consistent over time may also be accounted for to some extent by the methods used.

The main assessment of the extent to which results of these methods may be biased by unobservable characteristics is whether partner schools' and non-partner schools' PSC results change in a similar way before the programme started (a test of the parallel trends assumption). We see that PSC results of the two groups do change in a similar way over the long-term, including as PSC results change substantially over the first few years in which it was introduced (Figure 1); the similarity is greater when comparing only schools most appropriate for the analysis. More formal parallel trends tests show that there is a reasonable basis for the parallel trends assumption in the separate application of each difference-in-differences model to each cohort. There is one exception which further analysis indicates is due to random volatility and which would bias the impact estimate downwards if at all, see Annex C: Parallel trends tests.

In addition to the parallel trends test, the consistency of results across the three methods also provides some assurance that the limitations are not substantial, as each method is affected slightly differently by the identified limitations. See Annex F: Robustness checks for details on robustness checks carried out.

There are nevertheless three particular ways unobservable characteristics may still be biasing our impact estimates. These are the selection of partner schools (using characteristics we do not have data on), the possibility partner school pupils' progress before any school literacy teaching is different to non-partner school pupils', and the possibility that the COVID-19 pandemic has affected partner and non-partner schools differently.

³⁷ Pupil characteristics are controlled for at a cohort level in the school-level methods and at a pupil-level in the pupil-level methods.

Selection bias

The reliability of impact estimates from the methods we use can be affected by the way in which partner schools are chosen (selection bias). This is because partner schools are different to non-partner schools in ways which may affect their PSC results (such as an attitude towards improvement or quality of leadership), and the methods used to estimate impact do not include randomisation which would ensure the comparison group of non-partner schools used is similar to the group of partner schools in the ways that matter.

Mitigating selection bias is a particular challenge in these impact estimates. We observed that partner schools tended to have a lower PSC result one year before joining the programme compared to their own earlier PSC results (with a return to a more typical PSC result just before joining the programme), and the process evaluation found that hubs considered it important that they were able to select partner schools who were “ready or in a good position to engage with the EHP” (Department for Education, 2024, p. 50) and also that some partner schools were involved in other improvement initiatives, including English-focused initiatives. Impact estimates which do not control for this selection may be inflated by any improvement in PSC results which comes from the readiness of schools to improve or any response schools typically make to a downturn in PSC results, including implementing improvement initiatives. In addition, the programme targets schools with lower PSC results and cohorts with higher levels of disadvantage.

The methods used do overcome some of the recognised selection bias. This is shown by the fact that the difference-in-differences models used do not find an apparent impact of intensive support in the three-year period before support started in which we control for PSC trajectory (i.e. they pass the parallel trends test for this period), see Annex C: Parallel trends tests. In addition, we have only used results from cohorts where we can test and control for the observed pre-programme PSC trajectory (wave 1, AY 23/24 and AY 24/25 cohorts). The synthetic control method is also robust to the selection bias as far as it is apparent in pre-programme PSC results as it controls exactly for the pre-programme PSC results.

However, it is likely that some effect of selection bias remains which is not possible to overcome, as there will be some characteristics which we cannot control for as we do not have data on them. For instance, if the effectiveness of schools’ leadership differs between the two groups and, most importantly, this is associated with differing trends over time.

Variation in pupil prior learning

Estimates of impact of teacher professional development would be better measured by pupils’ progress rather than attainment since progress corresponds more closely to teaching effectiveness. This would be achieved by controlling for pupils’ prior attainment but that was not practical for this evaluation due to data controls relating to the agreed purpose of relevant pupil data. If there is a substantial difference in the progress made by

pupils prior to reception year between pupils in partner schools and non-partner schools, then this could affect the accuracy of the impact estimates. This potential bias is mitigated to some extent by the controls for pupil characteristics (full list of characteristics used for controls in Annex A: Detailed methods).

Uncertain effect of COVID-19 pandemic

It is possible that the COVID-19 pandemic affected partner schools and non-partner schools differently. If this is the case to a significant extent, then impact estimates for partner schools receiving support during the pandemic (including in wave 1) are not reliable since the parallel trends assumption does not hold. We are limited in how we establish the extent of any difference in the effect of the COVID-19 pandemic on partner schools and non-partner schools because the PSC did not take place during it. The PSC results of wave 4 likely give the most reliable indication, and they fell by a similar amount to PSC results of non-partner schools (see Figure 2). We consider wave 4 the most reliable indicator because this is the only cohort which started support after 2022 (the first year after the pandemic when the PSC took place) for which partner schools' selection would not have been based on post-pandemic PSC results. Whilst the PSC results for the AY 23/24 and AY 24/25 cohort partner schools fell further than those of non-partner schools between 2019 and 2022, this may be a result of partner schools being selected for programme support *because* their PSC results reduced further than typical during the pandemic. However, the disadvantage gap for PSC results widened between 2019 and 2022³⁸, which would suggest that partner schools may have been affected more by the pandemic since partner schools have a higher proportion of disadvantaged pupils. If partner schools were affected more by the pandemic than non-partner schools, then the wave 1 impact estimates will likely be biased downwards (i.e. underestimating the true impact, if there is no other bias).

Validity of impact estimates for all schools

The number of partner and non-partner schools we can use in the analysis is reduced as a result of the way we control for selection bias and only include schools open continuously with the same URN since 2015. This is a limitation, which particularly affects the difference-in-differences method. The sample sizes used for wave 1 are an example: 559 partner schools received the full period of EHP intensive support; of those 308 are used in the synthetic control method and 151 are used for both difference-in-differences methods (see Table 2, and Table 3). The impact estimates for each cohort overall are nevertheless still statistically significant.

Using a restricted selection of partner and non-partner schools means there is the possibility the impact estimates may be valid for the selected schools but not necessarily valid for other schools. For this reason, we compare the characteristics of the groups of partner and non-partner schools before and after selecting a smaller group for analysis.

³⁸ [Phonics screening check attainment, Academic year 2024/25](#)

The measurable characteristics of partner schools are broadly similar before and after filtering, and between filtering and matching (see Table 11, Table 12, and Table 13 in Annex B: Further supporting analysis). The largest difference is between wave 1 partner schools before and after matching. Wave 1 partner schools have the highest level of disadvantage and lowest PSC results of any cohort; the average percentage of year 1 pupils known to be disadvantaged in wave 1 partner schools is three percentage points lower, and the percentage of year 1 pupils with any SEN is one percentage point lower, after matching than before matching.

Since there are only specific differences between partner schools before and after matching, and only in wave 1 partner schools, we can consider the impact estimates to be broadly valid for partner schools in the EHP as a whole. However, it is possible that the true impact on schools with the very highest proportion of disadvantaged pupils may be different to our estimates.

Regression model used in difference-in-differences methods

The regression model used in the difference-in-differences methods is linear, which has limitations when applied to percentages (as we have done here). Some of the effect of this is mitigated by controlling for pre-programme mean PSC results, since the regression models are making comparisons of schools with similar PSC results.

Teacher turnover

There may be an impact on phonics learning from the programme support which is not captured by these impact estimates, as a result of teachers who benefitted from EHP intensive support moving to schools which are not partner schools. Any sustained improvement in teaching by these teachers would be experienced by pupils at their new school and not captured as an impact in these results (i.e. the impact estimates would be biased downwards). Hubs and partner schools reported that high turnover of staff in schools was a challenge in the English Hubs process evaluation (Department for Education, 2024).

Scope and impact measure

Effect of EHP non-intensive support

This impact evaluation estimates the impact on partner schools of intensive support. Partner schools and many other schools have received support through the EHP other than intensive support, including audits, CPD and resource funding. The impact estimates presented include the benefits of intensive support and some but not all of the benefits of this additional support, since some schools in the comparison groups used in each method have not received any kind of support from the EHP. The impact of additional support is likely to be included in our impact estimates to only a limited extent since more than half of all state-funded primary schools have received some kind of support from the programme.

Limitations of the Phonics Screening Check as an impact measure

The impact measure used is whether pupils meet the expected standard in the PSC. The impact estimates therefore capture the benefits for children who would be near the expected standard when the PSC takes place, and do not necessarily fully reflect benefits for pupils who make progress more slowly or already reach the expected standard earlier. In particular, our impact estimates do not necessarily fully capture impact on pupils who still do not meet the expected standard even after benefitting from improved teaching as a result of the EHP intensive support, or those who would meet the expected standard in the PSC without the effect of the programme support. This is pertinent given the programme prioritisation criteria include schools which have low attainment in the bottom 20% of children, and a high proportion of groups considered hard to reach, in addition to low PSC results and high levels of disadvantage (see Annex H: Criteria for hubs' selection of partner schools).

The process evaluation (Department for Education, 2024) found additional perceived benefits of the EHP intensive support which would not be directly reflected in PSC results. These include reading leads' leadership skills and career development, teachers' confidence and career progression, pupils' enjoyment and engagement with reading, and pupils' writing abilities.

Relevance to future implementation

The impact estimates show the effectiveness of the implementation of the EHP in achieving the aim of improving phonics learning in reception and year 1, since the impact measure used, the PSC, relates directly to phonics learning. These estimates correspond to the EHP as it has been implemented up to June 2025. Aspects of the implementation considered significant by hubs and partner schools reported in a 2024 process evaluation of the EHP include, amongst others, the bespoke nature of support including coaching, modelling and observations; that support is provided by literacy specialists with the most important skills and experience in schools similar to partner schools; and hubs' ability to "decide whether schools are ready or in a good position to engage with the EHP" (Department for Education, 2024, p. 50). The increasing number of SEND learners in mainstream schools was reported as a challenge for interviewed partner schools.

The impact estimates correspond to the impact on the partner schools that have participated so far. This is because we have controlled quite closely for relevant characteristics of schools and pupil populations. However, the impact estimates are for cohorts from both early and recent programme implementation, and the measurable characteristics of partner schools in the recently supported AY 23/24 cohort are in-between those of the wave 1 partner schools and non-partner schools. This means there is a substantial range of schools for which these impact estimates are valid for, though there is still some uncertainty about the extent to which the EHP intensive support would have the same impact on all non-partner schools.

The impact estimates reflect the impact on partner schools receiving support either during the COVID-19 pandemic or while returning to pre-pandemic PSC results. As more schools' PSC results return to pre-pandemic results, it is possible that the impact of a similar implementation of EHP intensive support may be different. There are, however, some cohorts of the EHP whose partner schools have reached an average PSC result higher since the pandemic than the highest pre-pandemic result, suggesting that the EHP intensive support may help schools to improve in a way that goes beyond reducing the effect of the pandemic.

Conclusion

We estimate that the EHP intensive support has resulted in an increase in the percentage of year 1 pupils meeting the expected standard in the PSC by three to five percentage points at the end of full intensive support. The impact estimates indicate that impact of EHP intensive support increases with each year of support and is partially but not fully sustained after support stops, with the impact reducing each year after support ends. The gradual reduction of impact following the end of support may be related to the context of longer-term COVID-19 pandemic recovery for the wave 1 cohort used to estimate longer-term impact, in which wave 1 partner schools recovered more quickly from the pandemic than non-partner schools whose PSC results improved later (from 2022).

We estimate the overall impact of EHP intensive support on partner schools in wave 1³⁹, measured from the final year of support to three years later, is between two and three percentage points; we estimate the overall impact on partner schools in the fifth cohort (AY 23/24⁴⁰), measured over the first two years of support, is between three and four percentage points.

The impact of the EHP at the end of intensive support has an effect size of 0.11 to 0.15. This is roughly equivalent to two months' additional progress for pupils, though the conversion to months' progress has notable limitations (see Annex E: Impact estimates using Phonics Screening Check score). These results are similar to those found by a meta-analysis of trials of teacher coaching using standardised pupil assessments (effect size of 0.12) (Kraft, et al., 2018) even though the EHP is at a much greater scale; this suggests the programme maintains a certain level of quality of provision and has some success in adapting support to local schools, overcoming at least to some extent recognised implementation challenges of scaling-up interventions. The EHP impact estimates are comparable to the effect on pupil reading outcomes of having a teacher with 6 to 12 years of experience rather than a new graduate (Atteberry, et al., 2015; Kini & Podolsky, 2016). There are some findings of meta-analyses of professional

³⁹ Partner schools in wave 1 received intensive support from September 2019 to July 2022.

⁴⁰ Partner schools in the AY 23/24 cohort started receiving intensive support during academic year 2023 to 2024.

development studies with slightly higher effect sizes. When comparing EHP impact estimates with those and e.g. the EEF teaching and learning toolkit, it is worth considering any differences in intervention type and scale, and that the estimated impact of the EHP is an aggregate of the impact on all pupils rather than on those who are the target of the EHP. The strength of evidence for the effectiveness of phonics as a pedagogy suggests the impact of the EHP on pupils' broader literacy learning is likely to be significant.

The difference in estimated impact between the first and fifth cohorts (wave 1 and AY 23/24) reflects that different periods of programme support are used in the analysis for each; other differences include that the wave 1 cohort received less support, had partner schools with higher levels of disadvantage and SEN and lower pre-pandemic PSC results, and received support during the COVID-19 pandemic rather than in a period of longer-term recovery from the pandemic.

Our estimates of impact suggest that there has not been a decrease in impact between the start of the EHP and recent waves. This is because there has been a similar impact on PSC results at the point of graduation for the first and fifth cohorts (wave 1 and AY 23/24). The consistency of impact between the wave 1 and AY 23/24 cohorts indicates that the EHP has been effective both during the pandemic and during post-pandemic recovery.

We consider these impact estimates reasonably strong evidence that the programme has had a positive effect. The methods account to some extent for the effect on PSC results of differences between partner schools and non-partner schools (such as partner school cohorts having a greater proportion of disadvantaged pupils), by comparing PSC results before as well as after support, and controlling for relevant measurable characteristics including lower PSC results, higher levels of disadvantage in pupil cohorts, and a pattern of school PSC results linked to the way hubs select partner schools which are ready to improve. Tests of the parallel trends assumption, which underpins the methods used, indicate it holds as far as we can assess it, and results are broadly consistent across all three methods used. The impact estimates here correspond to level three on the Maryland scale, which is a scale from one to five used to indicate the quality of evidence of intervention impact. Stronger research approaches would need to involve planned or accidental randomness in the selection of schools to receive support.

There are nevertheless notable limitations of the methods used, which may not fully account for all unobservable differences between partner schools and non-partner schools. There will likely be some remaining bias of our impact estimates resulting from the way in which partner schools are selected, including for a readiness to improve. The methods used mitigate the effect of this as far as possible with the available data, but there may be some effect of characteristics which we cannot control for. In this evaluation we were not able to directly control for pupil prior progress, so any difference in the prior learning between pupils at partner schools and non-partner schools may bias the impact

estimates somewhat. It also remains uncertain whether the COVID-19 pandemic may have affected partner schools and non-partner schools differently, although indications are that any difference is not substantial and would bias our impact estimates downwards (i.e. estimates would underestimate the true impact). The number of partner schools we are able to use for this analysis is substantially restricted due to the need to mitigate selection bias, but impact estimates are still statistically significant and the partner schools used in the analysis are broadly representative of all partner schools.

There are also a number of ways these impact estimates do not capture the full benefits of the programme. The likely movement of some teachers in partner schools to work in non-partner schools will bias the estimates downwards (i.e. estimates would underestimate the true impact). The impact estimates reflect the impact of EHP intensive support but only some limited part of the impact of non-intensive support. In addition, the impact estimates capture the benefits for year 1 children who would be near the expected standard when the PSC takes place, and do not necessarily fully reflect benefits for pupils who make progress more slowly or already reach the expected standard earlier; this is notable since prioritisation criteria for the selection of partner schools includes low attainment in the bottom 20% of pupils and a high proportion of groups considered hard to reach. Further, hubs and partner schools perceived benefits of the EHP intensive support which are not captured by our impact measure such as reading leads' leadership skills and career development, teachers' confidence and career progression, pupils' enjoyment and engagement with reading, and pupils' writing abilities (Department for Education, 2024).

The impact estimates correspond to the EHP intensive support as implemented so far. Aspects of this considered significant by hubs and partner schools include that the support is bespoke to each partner school, provided by literacy specialists with particular skills and experience in schools similar to partner schools, provided to partner schools which hubs consider ready for the programme. See the English Hubs process evaluation for further details (Department for Education, 2024).

There are reasons why the impact of future support provided by the programme could be different to the impact estimated of programme support so far. Schools which have not yet received intensive support and their pupil populations may be different to partner schools supported so far, and the impact of intensive support may be different as more schools return to PSC results they were achieving before the pandemic. However, there are indications across the cohorts so far that EHP intensive support has impact on partner schools with a range of characteristics, and has supported some improvement in partner schools beyond returning to pre-pandemic results.

References

- Abadie, A., 2021. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature*, pp. 391-425.
- Atteberry, A., Loeb, S. & Wyckoff, J., 2015. Do First Impressions Matter? Predicting Early Career Teacher Effectiveness. *AERA Open*.
- Basma, B. & Savage, R., 2017. *Teacher Professional Development and Student Literacy Growth: a Systematic Review and Meta-analysis*, s.l.: Educational Psychology Review.
- Blank, R. K. & de las Alas, N., 2009. *The Effects of Teacher Professional Development on Gains in Student Achievement: How Meta Analysis Provides Scientific Evidence Useful to Education Leaders*, s.l.: Council of Chief State School Officers.
- Callaway, B. & Sant-Anna, P. H. C., 2021. Difference-in-Differences with multiple time periods. *Journal of Econometrics*, pp. 200-230.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. s.l.:Routledge.
- de Chaisemartin, C. & D'Haultfoeuille, X., 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*.
- de Chaisemartin, C. & D'Haultfoeuille, X., 2023. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, pp. C1-C30.
- Department for Education, 2024. *A Process Evaluation of the English Hubs Programme*, s.l.: Department for Education.
- Didion, L., Toste, J. R. & Filderman, M. J., 2019. Teacher Professional Development and Student Reading Achievement: A Meta-Analytic Review of the Effects. *Journal of Research on Educational Effectiveness*.
- Dimova, S. et al., 2020. *The Nuffield Early Language Intervention Evaluation Report*, s.l.: Education Endowment Foundation.
- Education Endowment Foundation, 2021. *Phonics*. [Online]
Available at: <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/phonics#nav-what-is-it>
- Education Endowment Foundation, 2022. *Statistical analysis guidance for EEF evaluations*. [Online]
Available at: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design>

- Fletcher-Wood, H. & Zuccollo, J., 2020. *The effects of high-quality professional development on teachers and students*, s.l.: Education Policy Institute.
- Fritz, C., Morris, P. & Richler, J., 2012. Effect Size Estimates: Current Use, Calculations, and Interpretation. *Journal of Experimental Psychology*.
- Fryer, R. J., 2017. Chapter 2 - The Production of Human Capital in Developed Countries: Evidence From 196 Randomized Field Experiments. In: *Handbook of Economic Field Experiments*. s.l.:Elsevier, pp. 95-322.
- Hainmueller, J., 2012. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, pp. pp. 25-46.
- Higgins, S. et al., 2013. *The Sutton Trust - Education Endowment Foundation Teaching and Learning Toolkit*. [Online]
Available at: <https://durham-repository.worktribe.com/output/1608325>
- Institute of Education Sciences, 2022. *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0*, s.l.: National Center for Education Evaluation at IES.
- Kahmann, R., Droop, M. & Lazonder, A. W., 2022. Meta-analysis of professional development programs in differentiated instruction.
- Kellogg, M., Mogstad, M., Pouliot, G. A. & Torgovitsky, A., 2021. Combining Matching and Synthetic Control to Tradeoff Biases From Extrapolation and Interpolation. *Journal of the American Statistical Association*, pp. 1804-1816.
- Kini, T. & Podolsky, A., 2016. *Does Teaching Experience Increase Teacher Effectiveness? A Review of the Research*, s.l.: Learning Policy Institute.
- Kraft, M. A., Blazar, D. & Hogan, D., 2018. The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Review of Educational Research*.
- Kreif, N. et al., 2016. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. *Health Economics*, pp. 1514-1528.
- Lindvall, J., Kirsten, N., Ryve, A. & Gustafsson, J.-E., 2025. A critical review and meta-analysis of studies investigating the effects of the professional development teachers typically receive.
- Lockwood, J. R., McCombs, J. S. & Marsh, J., 2010. Linking Reading Coaches and Student Achievement: Evidence from Florida Middle Schools. *Educational Evaluation and Policy Analysis*, pp. 372-388.

Lynch, K., Gonzalez, K., Merritt, R. & Hill, H., 2025. A Meta-Analysis of the Experimental Evidence Linking Mathematics and Science Professional Development Interventions to Teacher Knowledge, Classroom Instruction, and Student Achievement. *AERA Open*.

McCrabb, S. et al., 2019. Scaling-up evidence-based obesity interventions: A systematic review assessing intervention adaptations and effectiveness and quantifying the scale-up penalty. *Obesity Reviews*.

Rice, M., Lambright, K. & Wijekumar, K., 2024. Professional Development in Reading Comprehension: A Meta-analysis of the Effects on Teachers and Students. *Reading Research Quarterly*.

Robbins, M., Saunders, J. & Kilmer, B., 2017. A Framework for Synthetic Control Methods With High-Dimensional, Micro-Level Data: Evaluating a Neighborhood-Specific Crime Intervention. *Journal of the American Statistical Association*, pp. 109-126.

Rose, J., 2006. *Independent review of the teaching of early reading*, s.l.: Department for Education and Skills.

Roth, J., Sant-Anna, P. H. C., Bilinski, A. & Poe, J., 2023. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*.

Sant'Anna, P. H. C. & Zhao, J., 2020. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, pp. 101-122.

Shepherd, H. & Fortescue, B., 2023. *Early analysis of English Hubs phonics attainment: 2021/22 data*, s.l.: Department for Education.

Sims, S. et al., 2021. *What are the Characteristics of Teacher Professional Development that increase Pupil Achievement? A systematic review and meta-analysis*, London: Education Endowment Foundation.

Sims, S. et al., 2023. Effective Teacher Professional Development: New Theory and a Meta-Analytic Test. *Review of Educational Research*.

Smith, A., Staunton, R., Sahasranaman, A. & Worth, J., 2023. *Impact Evaluation of Nuffield Early Language Intervention (NELI) Wave Two*, s.l.: Education Endowment Foundation.

van den Brande, J. Z. J., 2021. *The effects of high-quality professional development on teachers and student: A cost-benefit analysis*, s.l.: Education Policy Institute.

Yohros, A. & Welsh, B., 2019. Understanding and Quantifying the Scale-Up Penalty: a Systematic Review of Early Developmental Preventive Interventions with Criminological Outcomes. *Journal of Developmental and Life-Course Criminology*, pp. 481-497.

Annex A: Detailed methods

Introduction to quasi-experimental design methods

Difference-in-differences

Applying difference-in-differences to the English Hubs programme means comparing the change in the percentage of pupils meeting the expected standard in the PSC in partner schools before and after programme support with the change in the percentage of pupils meeting the expected standard in the PSC in non-partner schools before and after programme support. The difference between these two changes is the estimate of impact of the programme. For instance, if PSC results in partner schools increase more than those in non-partner schools during the time of their participation in the programme then this would indicate a positive impact of the programme.

We use regression to implement the difference-in-differences model as it offers advantages over a simple calculation. Using regression, we can use multiple years of results within the same model, calculate statistics to indicate confidence level, run assumption tests, and improve the comparison of outcomes between groups beyond averages for each group. This is a typical approach for impact evaluations using difference-in-differences.

Synthetic control

Synthetic control (SC) is a statistical method used to estimate causal effects in observational studies where treatment allocation is not randomised. It works by creating a synthetic control which closely approximates the time trend of the treatment group (schools participating in the programme) prior to the onset of intervention. Synthetic control is therefore advantageous when the parallel trends assumption does not hold (Abadie, 2021). We applied this method to school-level PSC outcomes⁴¹.

The motivating factor behind synthetic control is that a weighted combination of control units may provide a better counterfactual to the treatment group than any single unit in the control group. In contrast to matching methods such as Propensity Score Matching, synthetic control weights each untreated unit according to its similarity to the treatment group and therefore uses all available data as opposed to only a subset of units matched on similar characteristics. This weighting process is applied to the pre-intervention period only, resulting in closely matching PSC trajectories for the synthetic control and treatment groups up until the onset of support. Weights are constrained to be nonnegative and to sum to 1. While every untreated unit contributes to the synthetic control, weights are

⁴¹ Unlike difference-in-differences, there is no equivalent method to synthetic control which can be applied at a pupil-level. This is because the synthetic control method works by establishing a single synthetic control which is used for all years, and year 1 pupils are different each year.

sparse such that only a small number of units have high weights while the majority have low or very low weights.

Although the original formulation of synthetic control involves a single aggregated treated unit (e.g. a state or country), recent work has extended this to study designs with multiple treated units (Robbins, et al., 2017), as in the case of the multiple partner schools in the EHP. The use of multiple treated units allows the synthetic control to more closely approximate the trajectory of the treated units, and results in higher statistical power to detect an effect (Kreif, et al., 2016) . Use of such disaggregated units or ‘micro’ data in this context is known as microsynthetic control.

The treatment effect under synthetic control is the difference between the summed weights of the treatment and control group in the post-intervention period. The statistical significance of this effect is determined through a permutation test.

Matching

Propensity Score Matching calculates a probability for each school of it being a partner school based on a range of specified known characteristics of partner schools and non-partner schools. It then includes schools in the selection for each group for the difference-in-differences model if they can be paired with a school with a sufficiently similar probability from the other group. For the pupil-level analysis we used entropy balancing (Hainmueller, 2012) to establish an appropriate weighting of pupils using the chosen characteristics.

Application of methods to English Hubs Programme

In all methods, the period of PSC results used is from 2016 to 2025, inclusive. We control for the following characteristics, by pre-matching for the difference-in-differences models, and by weighting within the synthetic control model.

Table 8: List of control characteristics

<p>School level controls [all models]</p>	<ul style="list-style-type: none"> • Summary school type (mainstream, special, other) • Whether school in an urban area • Government Region • Whether school’s latest Ofsted judgement is either ‘requires improvement’ or ‘inadequate’ • Number of year 1 pupils
<p>School PSC controls [difference-in-differences models only]</p>	<ul style="list-style-type: none"> • Mean pre-programme PSC result • Change in schools’ PSC result from previous year, for each of three assessed years immediately prior to starting support
<p>Cohort level controls [school-level models]</p>	<ul style="list-style-type: none"> • Proportion of year 1 pupils who have been eligible for free school meals at any point in the past six years • Proportion of year 1 pupils with any identified Special Educational Need (SEN Support or an Education, Health and Care Plan) • Proportion of year 1 pupils whose first language was one other than English (English as an Additional Language, EAL) • Proportion of year 1 pupils born in summer • Proportion of year 1 pupils who are male • Proportion of year 1 pupils of white British ethnicity
<p>Pupil level controls [pupil-level model]</p>	<ul style="list-style-type: none"> • Whether eligible for free school meals at any point in the past six years • Whether identified with any Special Educational Need • Whether has a first language other than English (EAL) • Whether born in summer • Whether male • Whether of white British ethnicity

Difference-in-differences

There is a need to use methods robust to heterogeneous impacts since the programme has multiple cohorts starting support in different years and multiple years with PSC results since support started (de Chaisemartin & D’Haultfoeuille, 2020; Roth, et al., 2023). We applied the ‘did’ R package (Callaway & Sant-Anna, 2021) separately to each cohort after pre-matching. The ‘did’ package uses linear regression, which has limitations when applied to an outcome which is a percentage (as we have done); this limitation is

mitigated by controlling for school pre-programme mean PSC in both difference-in-differences models.

We used errors clustered at the school level (for both school-level and pupil-level difference-in-differences). This may mean the uncertainty in our pupil-level impact estimates is overestimated.

Matching for the school-level difference-in-differences model

We used nearest neighbour propensity score matching with a 5:1 ratio. The variable we used when matching to control for school-level PSC ‘trajectory’ is the difference in the percentage of year 1 pupils meeting the expected standard in a school between one year and the previous year (i.e. the percentage point change since the previous year). These variables are used for school PSC results one, two and three years prior to support starting.

Matching and weighting for the pupil-level difference-in-differences model

We completed pre-matching for the pupil-level difference-in-differences model in two stages. We first used nearest-neighbour propensity score matching with a 5:1 ratio at school-level, in the same way as for the school-level difference-in-differences model, but with the variables in Table 9:

Table 9: List of control characteristics used in the first pre-matching stage of the pupil-level difference-in-differences model

School level controls	<ul style="list-style-type: none"> • Summary school type (mainstream, special, other) • Whether school in an urban area • Government Region • Whether school’s latest Ofsted judgement is either ‘requires improvement’ or ‘inadequate’ • Number of year 1 pupils
School PSC controls	<ul style="list-style-type: none"> • Mean pre-programme PSC result • Change in schools’ PSC result from previous year, for each of three assessed years immediately prior to starting support
Cohort level controls	<ul style="list-style-type: none"> • Proportion of year 1 pupils who have been eligible for free school meals at any point in the past six years

We then used entropy balancing at the pupil level on year 1 pupils in schools matched in the first pre-matching stage. Entropy balancing was completed on year 1 cohorts in each academic year separately, with the variables in Table 10:

Table 10: List of control characteristics used in the second pre-matching stage of the pupil-level difference-in-differences model

School level controls	<ul style="list-style-type: none"> • Summary school type (mainstream, special, other) • Whether school in an urban area • Government Region • Whether school’s latest Ofsted judgement is either ‘requires improvement’ or ‘inadequate’ • Number of year 1 pupils
School PSC controls	<ul style="list-style-type: none"> • Mean pre-programme PSC result
Cohort level controls	<ul style="list-style-type: none"> • Proportion of year 1 pupils who have been eligible for free school meals at any point in the past six years
Pupil level controls	<ul style="list-style-type: none"> • Whether eligible for free school meals at any point in the past six years • Whether identified with any Special Educational Need • Whether has a first language other than English (EAL) • Whether born in summer • Whether male • Whether of white British ethnicity

We explored alternative matching strategies in the difference-in-differences models, including coarsened exact matching both at school-level and pupil-level in combination with other matching and weighting approaches, but found when using this approach that the final selection of schools or pupils was too small (and year-to-year variation of PSC results of the final selection of schools or pupils too great).

Synthetic control

The ‘microsynth’ R package (Robbins et al., 2017) was used to carry out synthetic control analysis. Since the method does not allow for time-varying covariates, we computed average values for model covariates across the pre-treatment period (i.e. for Wave 1, average values for each school were computed between the academic years 2015 to 2016 and 2018 to 2019 for EAL, SEND, FSM, and so on).

P-values for estimates were derived using permutation tests, with the number of tests set to 250 for each wave. All p-values were under 0.001.

Weight distributions were inspected visually, with all adhering to the expectation of sparseness (only a small number of high weights). Manual checks were also carried out

to compare covariate values between partner and non-partner schools, and between schools with high and low weights. This comparison consistently met the expectation that schools with high weights had more similar covariate values to partner schools than schools with low weights.

Effect size calculations

We calculated effect sizes to show the impact of the programme in terms of a standardised measure. We used the Cox index to calculate an effect size for the pupil-level difference-in-differences model (Institute of Education Sciences, 2022) since the impact measure for that model (whether or not each year 1 pupil met the expected standard in the PSC) is a dichotomous variable. This was calculated using the percentage of year 1 pupils meeting the expected standard in the PSC in partner schools in the period since programme support started (or the year full intensive support ends), together with the same percentage minus the estimated percentage point impact.

We used Hedges' *g* to calculate effect sizes for the school-level models (school-level difference-in-differences and synthetic control models). This was applied by calculating the standard deviation for the group of non-partner schools (after filtering but before matching) in the period since programme support started. This accounts for the effect on outcome distribution of the COVID-19 pandemic and does not account for the effect on outcome distribution of the EHP, which is appropriate. Non-partner schools only are used as we expect there is a difference between the true standard deviations of the group of partner schools and non-partner schools due to the way partner schools are selected for low PSC results; this is consistent with the approach taken by the Education Endowment Foundation (Education Endowment Foundation, 2022).

For impact estimates using the pupil PSC score we used Hedges' *g* to calculate effect sizes (as described for school-level models) and to convert to a number of additional months' progress made by pupils.

Annex B: Further supporting analysis

Characteristics of partner schools and pupil populations

This information is shared to show to what extent the schools used in the analysis are representative of all schools, since only a selection are used after filtering and matching.

Table 11: Characteristics of pupil populations by cohort before filtering

Characteristic	Non-partner, N = 4,273,300	Wave 1, N = 148,115	Wave 2, N = 78,211	Wave 3, N = 45,201	Wave 4, N = 173,958	AY23/24, N = 219,934	AY24/25, N = 137,364
% disadvantaged	19.2	30.6	30.1	28.8	28.2	24.9	24.2
% SEN Support	11.1	14.8	14.3	14.6	14.1	13.0	13.0
% EHCP	2.8	2.3	2.3	2.5	2.2	2.1	2.2
% Special Educational Needs	13.9	17.1	16.6	17.1	16.3	15.0	15.2
% English as an additional	20.4	23.4	24.2	24.2	22.9	20.4	21.2
% summer born	33.7	33.9	33.9	33.8	33.9	33.8	33.9
% male	51.2	51.2	51.3	51.3	51.3	51.1	51.2
% white British	64.4	63.1	62.4	61.7	62.8	65.5	64.4

Source: School Census, EHP programme data 2016 to 2025⁴²

⁴² Mean of available data in period 2016 – 2025.

Table 12: Characteristics of school-average pupil populations by cohort after filtering on schools open continuously⁴³

Characteristic	Non-partner, N = 8,933	Wave 1, N = 308	Wave 2, N = 170	Wave 3, N = 94	Wave 4, N = 358	AY23/24, N = 468	AY24/25, N = 303
% disadvantaged	18.7 (14.0)	30.7 (15.2)	29.5 (15.1)	29.2 (16.2)	27.9 (14.6)	24.0 (14.2)	23.5 (14.9)
% SEN Support	10.8 (5.8)	14.5 (6.3)	13.9 (5.6)	14.5 (6.8)	14.1 (6.7)	13.0 (6.0)	13.1 (6.5)
% EHCP	3.8 (13.7)	2.5 (2.7)	2.3 (2.6)	2.7 (2.1)	2.2 (1.9)	2.2 (4.8)	2.2 (1.9)
% Special Educational Needs	14.6 (13.7)	16.9 (6.7)	16.2 (6.3)	17.2 (7.2)	16.2 (7.0)	15.2 (7.5)	15.4 (6.9)
% English as an additional language	17.0 (20.7)	21.0 (22.5)	20.7 (22.1)	21.8 (24.6)	20.1 (22.1)	17.3 (20.4)	16.9 (21.3)
% summer born	33.7 (3.3)	34.0 (3.0)	33.6 (3.8)	34.2 (3.3)	33.8 (3.4)	33.7 (3.3)	33.8 (3.2)
% male	51.4 (4.8)	51.3 (3.1)	51.2 (3.6)	51.4 (3.6)	51.3 (3.4)	51.2 (3.6)	51.0 (3.7)
% white British	69.1 (28.2)	65.6 (29.7)	66.8 (28.1)	64.5 (32.0)	66.5 (29.5)	69.6 (28.1)	70.9 (28.6)
% Ofsted 'requires improvement' or 'inadequate'	1,687 (19%)	140 (45%)	78 (46%)	45 (48%)	143 (40%)	180 (38%)	101 (33%)
Pupil numbers	41.8 (23.5)	40.4 (18.4)	39.3 (19.6)	39.7 (19.9)	40.6 (20.4)	42.4 (21.8)	41.8 (24.9)

Source: School Census, EHP programme data 2016 to 2025⁴⁴

⁴³ Figures are a mean (with standard deviation in brackets), except for Ofsted 'requires improvement' or 'inadequate', which are a count with the proportion of schools in brackets.

⁴⁴ Mean of available data in period 2016 to 2025. Calculated for schools open continuously in that time without changing URN and with at least five pupils.

Table 13: Characteristics of school-average pupil populations by cohort after matching for school difference-in-differences method⁴⁵

Characteristic	Wave 1 analysis - Non-partner, N = 755	Wave 1 analysis – partner, N = 151	AY 23/24 analysis - Non-partner, N = 1,445	AY 23/24 analysis - partner, N = 2891	AY 24/25 analysis - Non-partner, N = 1,015	AY 24/25 analysis – partner, N = 203
% disadvantaged	28.0 (16.5)	27.7 (14.6)	22.6 (15.0)	22.7 (14.5)	22.8 (15.3)	22.9 (15.3)
% SEN Support	13.8 (6.5)	13.6 (6.3)	12.4 (6.2)	12.7 (6.1)	13.3 (6.7)	13.0 (6.8)
% EHCP	2.1 (2.2)	2.3 (2.7)	2.5 (7.0)	2.3 (6.0)	2.0 (1.9)	2.3 (2.1)
% Special Educational Needs	15.9 (7.1)	15.8 (6.6)	14.8 (8.9)	14.9 (8.2)	15.4 (7.2)	15.3 (7.2)
% English as an additional	19.7 (22.7)	18.2 (21.4)	15.5 (19.8)	15.0 (19.4)	13.7 (19.3)	14.1 (19.2)
% summer born	33.8 (3.4)	34.0 (3.4)	33.8 (3.4)	33.7 (3.6)	33.6 (3.4)	33.7 (3.5)
% male	50.9 (3.8)	51.0 (3.1)	51.2 (4.0)	51.3 (4.1)	51.1 (4.0)	51.1 (4.0)
% white British	66.2 (30.9)	68.6 (29.5)	72.1 (26.9)	72.6 (27.0)	75.4 (25.5)	74.7 (26.4)
N (%) Ofsted 'requires improvement' or 'inadequate'	317 (42%)	63 (42%)	496 (34%)	104 (36%)	320 (32%)	63 (31%)
Pupil numbers	37.7 (20.6)	36.4 (17.4)	37.9 (21.6)	37.6 (21.3)	36.7 (21.0)	36.6 (19.9)

Source: School Census, EHP programme data 2016 to 2025⁴⁶

⁴⁵ Figures are a mean (with standard deviation in brackets), except for Ofsted 'requires improvement' or 'inadequate', which are the number with the proportion of schools in brackets.

⁴⁶ Mean of available data in period 2016 to 2025. Calculated for schools matched on specified characteristics.

Distribution of school phonics screening check results over time

Figure 9: School distribution of phonics screening check results – wave 1 cohort

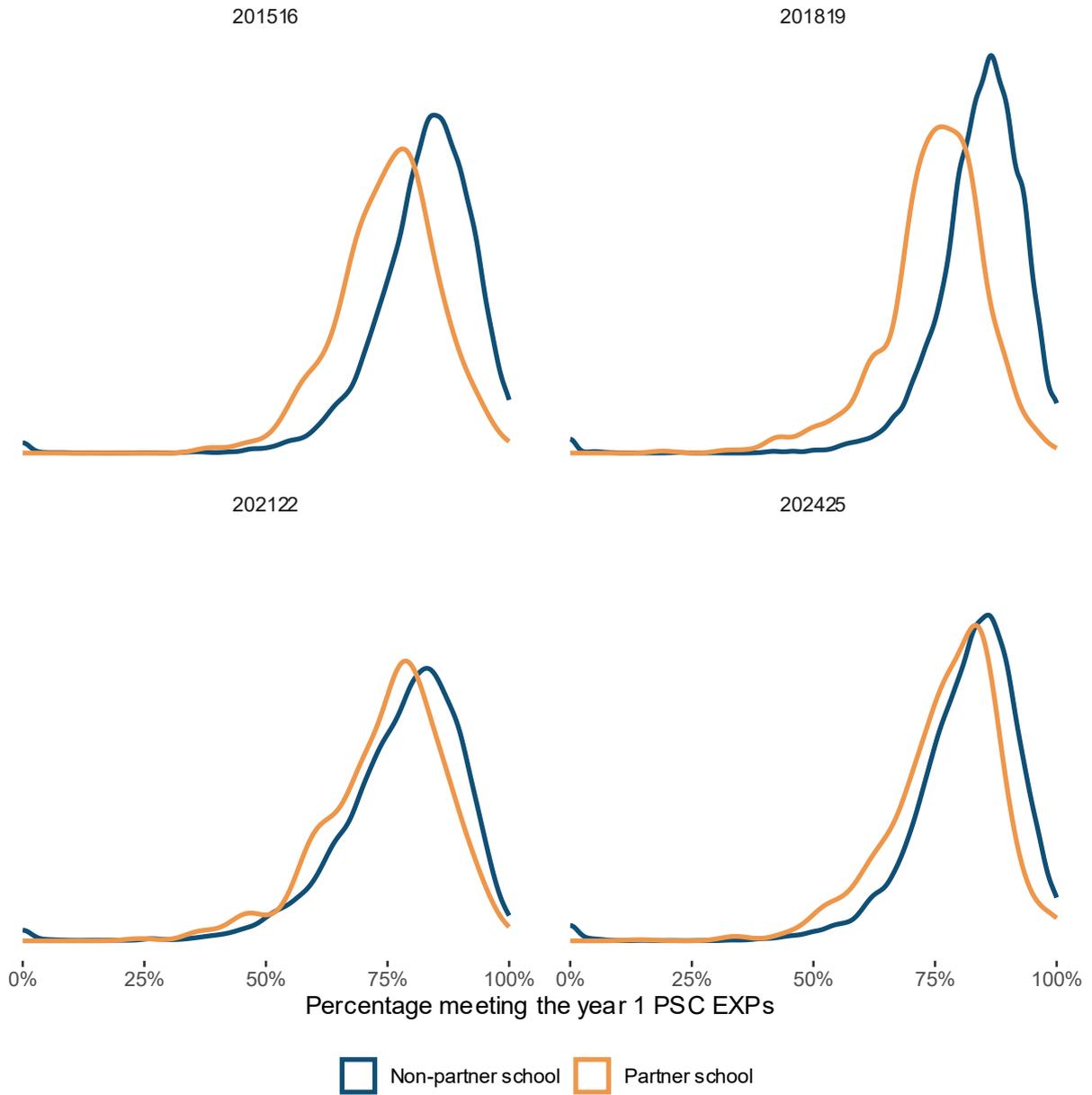
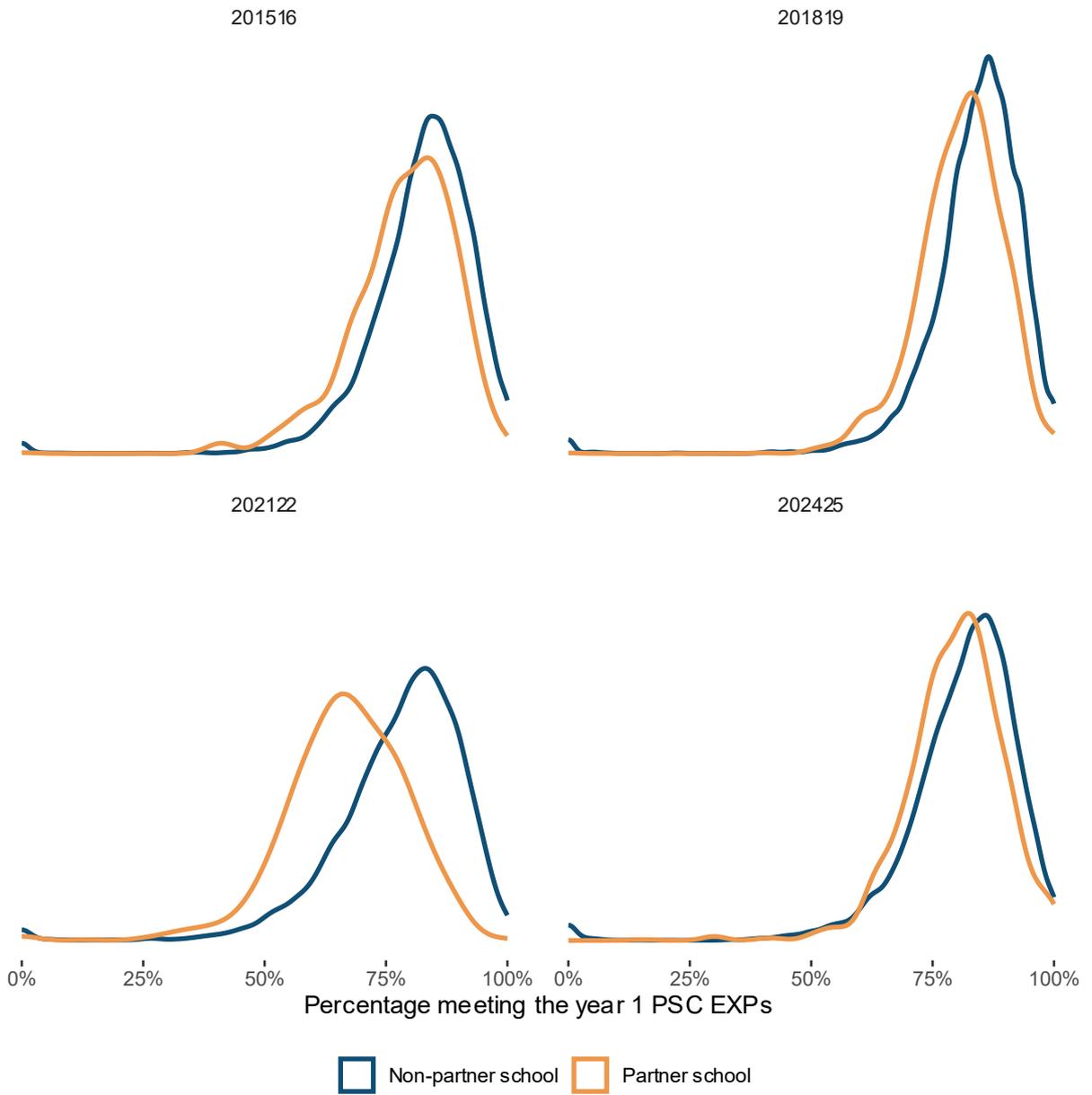


Figure 10: School distribution of phonics screening check results – AY 23/24 cohort



Annex C: Parallel trends tests

First cohort (wave 1)

Figure 11: School-level difference-in-differences impact estimates by year – wave 1 cohort

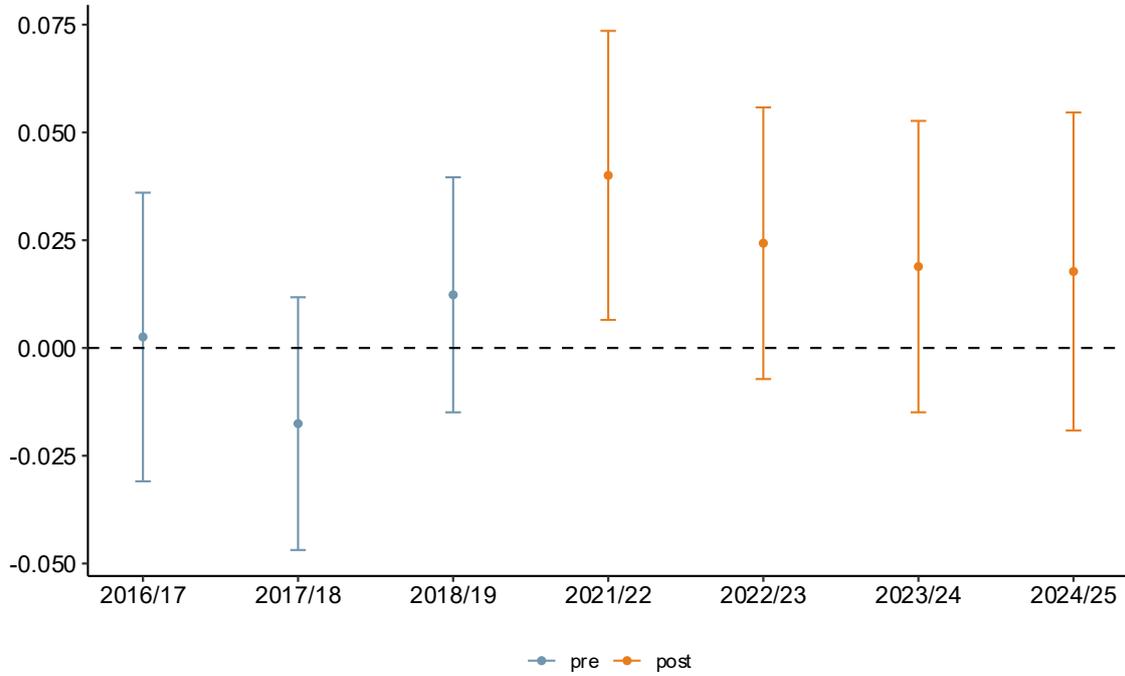
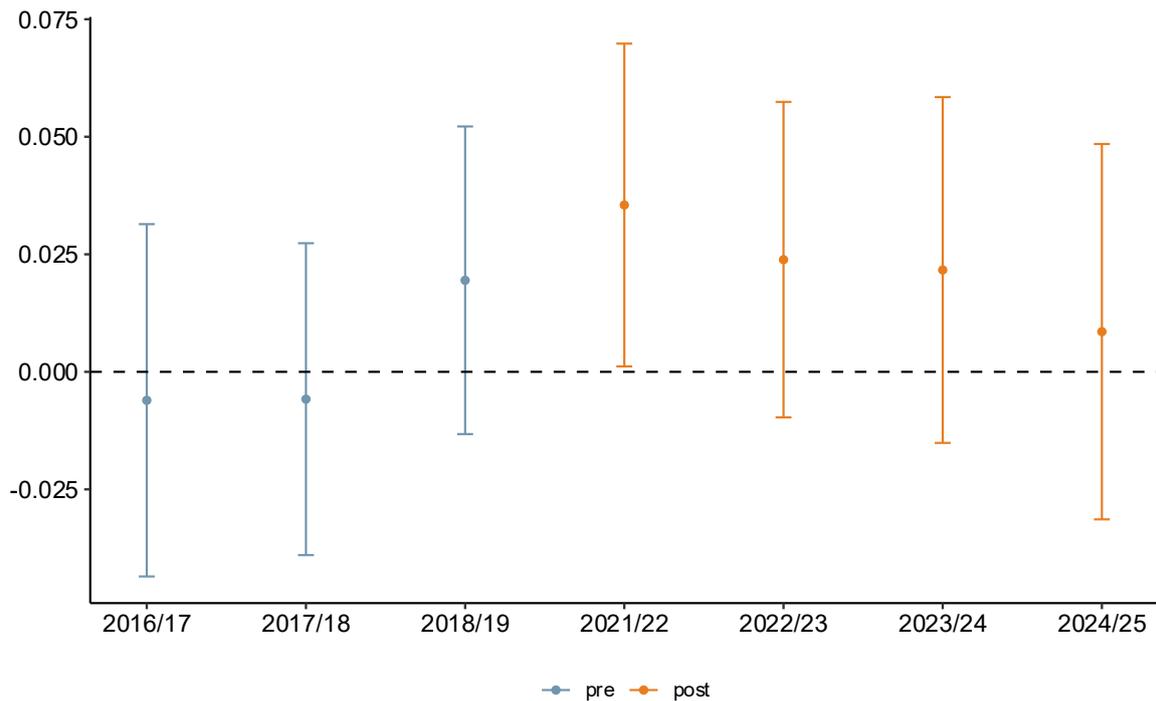
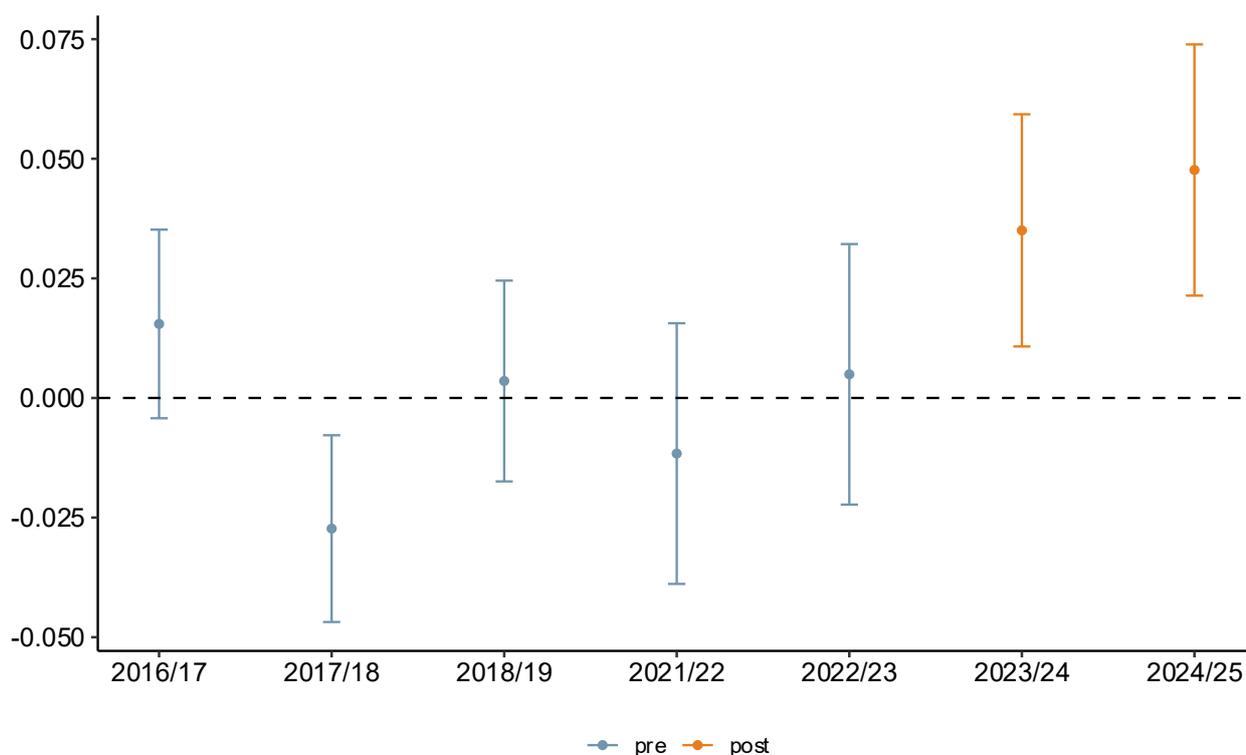


Figure 12: Pupil-level difference-in-differences impact estimates by year – wave 1 cohort



Fifth cohort (AY 23/24)

Figure 13: School-level difference-in-differences impact estimates by year – AY23/24 cohort

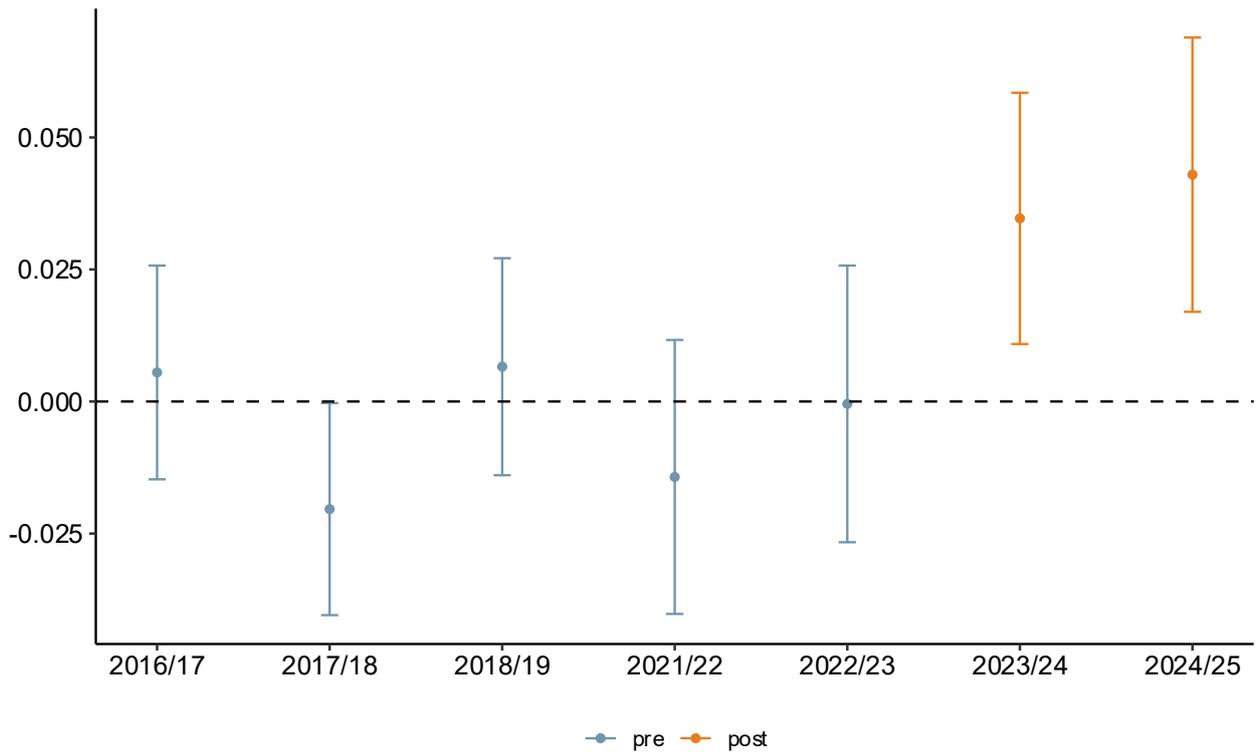


There is one pre-programme year (2017) in which the school-level difference-in-differences model does not fully pass the parallel trends test, as the test finds a small but statistically significant negative apparent impact of the programme⁴⁷. On inspection, this is due to substantial volatility in the average PSC result of matched non-partner schools used in this model. The same model controlling for the PSC trajectory only for the two most recent PSC results before support starts (and not the three most recent), which does pass the parallel trends test due to lower volatility, actually gives a higher estimate of impact (5.1 percentage points (3.6 to 6.7)⁴⁸). This robustness check together with the use of the synthetic control method (in which the synthetic control of non-partner schools matches the pre-programme PSC time series of partner schools very closely), provides assurance that the AY 23/24 school-level difference-in-differences model is not unduly influenced by a measurable violation of the parallel trends assumption.

⁴⁷ This does not contribute to the estimated impact, rather is a test of its validity.

⁴⁸ 5.9 percentage point impact at the point of graduation (2025).

Figure 14: Pupil-level difference-in-differences impact estimates by year – AY 23/24 cohort



Sixth cohort (AY 24/25)

Figure 15: School-level difference-in-differences impact estimates by year – AY24/25 cohort

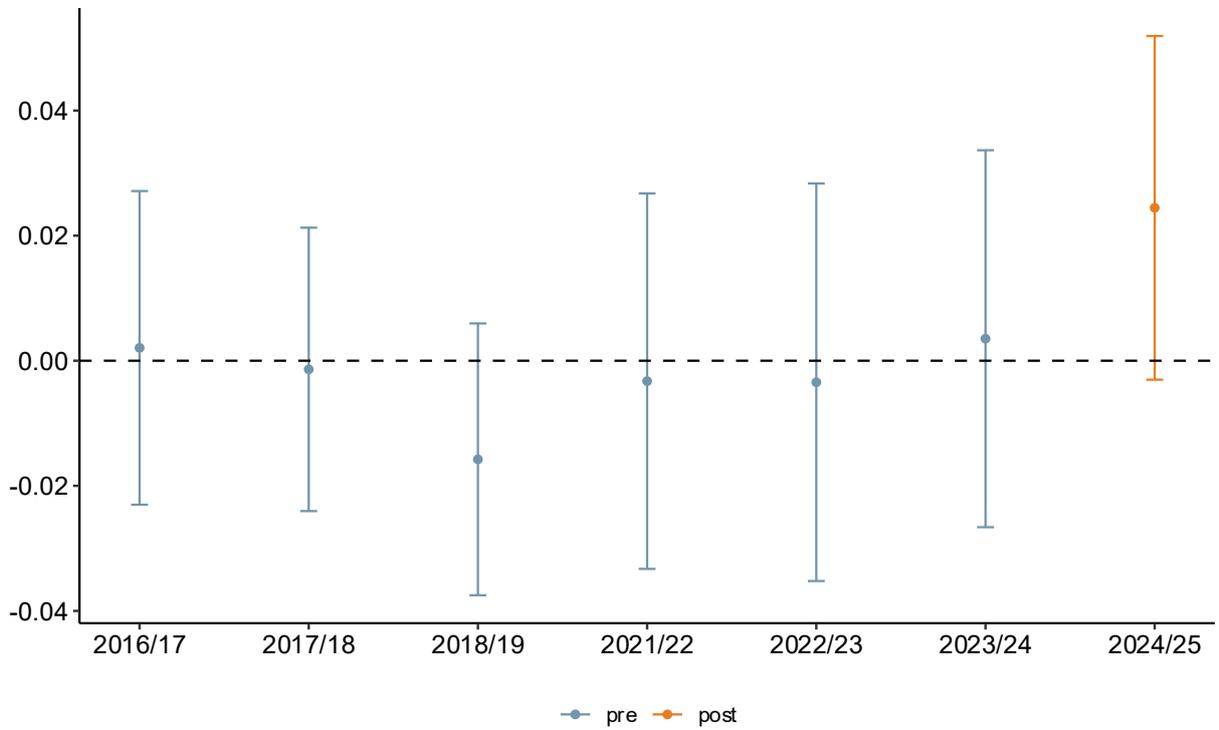
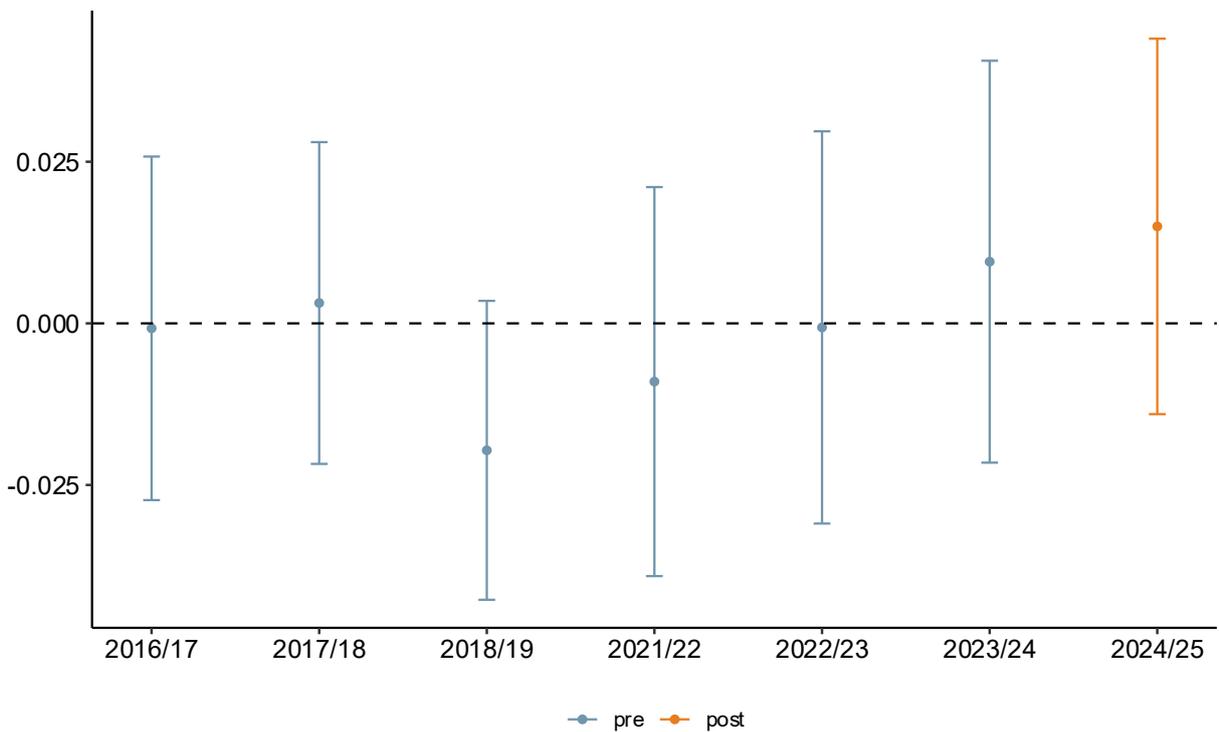


Figure 16: Pupil-level difference-in-differences impact estimates by year – AY 24/25 cohort



Annex D: Matching assessments

An illustration of the effectiveness of pre-matching in the difference-in-difference models is given by Table 12 and Table 13 in Annex B: Further supporting analysis, for school characteristics, and by Figure 5 and Figure 6 in the Impact by cohort section for PSC results.

Weighting assessment results for the synthetic control model for wave 1 are included as an example.

Table 14: Wave 1 synthetic control weight balance table

Variable	Partner Schools	Synthetic Control	All Schools
Intercept	308.00000	308.00001	308.00000
EAL	61.68220	61.68220	51.04341
SEN (any)	48.95092	48.95092	41.20298
Free School Meals	69.49754	69.49754	41.47646
Disadvantage	87.43465	87.43465	52.97868
Summer born	104.08872	104.08873	103.30332
Gender	158.30186	158.30186	158.35267
White ethnicity	209.63548	209.63549	219.12939
Special School	0.00000	0.00000	6.03268
Ofsted Rating	111.00000	111.00000	43.62861
Region: East Midlands	23.00000	23.00001	27.66367
Region: East of England	18.00000	18.00000	33.92966
Region: London	32.00000	32.00000	41.92880
Region: North East	14.00000	14.00000	13.73185
Region: North West	58.00000	58.00000	52.16102
Region: South East	27.00000	27.00000	52.86095
Region: South West	52.00000	52.00000	25.96386
Region: West Midlands	20.00000	20.00000	33.02976
Region: Yorkshire	64.00000	64.00000	26.69711
Urban flag	251.00000	251.00001	227.90867
Cohort size	13146.50000	13146.50061	13200.62969
PSC 2015 to 2016	229.96955	229.96956	253.20818
PSC 2016 to 2017	222.22334	222.22334	255.17904
PSC 2017 to 2018	231.22454	231.22455	251.06534
PSC 2018 to 2019	232.98696	232.98697	249.47369

Note that synthetic control weights are constrained to be in the range [0, 1]. Individual dots in the figure represent weights given to each non-partner school in wave 1. By design, weights are ‘sparse’ in that only a small number of units receive high weights (i.e. are very similar to partner school units).

Figure 17: Wave 1 synthetic control weight distribution

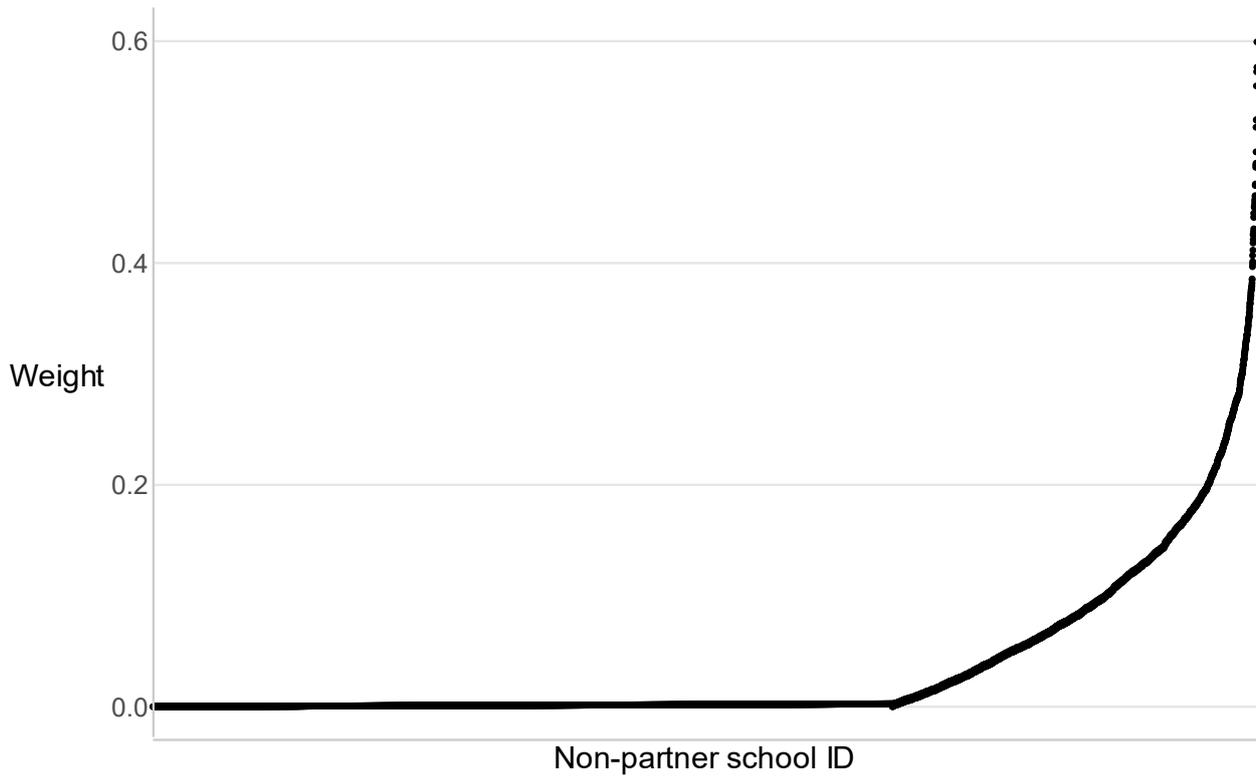


Table 15: Wave 1 synthetic control weight distribution

Min	1 st quartile	Median	Mean	3 rd quartile	Max
0.00000	0.00000	0.00000	0.03425	0.03381	0.59780

Annex E: Impact estimates using Phonics Screening Check score

Methods

We estimated the impact of the English Hubs programme using the PSC score as the impact measure (each pupil scores 0 to 40 on the PSC), in addition to the percentage meeting the expected standard in the PSC. This is to aid interpretation and to establish effect size results and equivalents in pupils' months' progress directly comparable to those of other programmes.

One additional impact measure we used is the mean PSC score. We used the PSC score for the pupil-level difference-in-differences model, and the school mean PSC score for the school-level models. We applied the same models using the PSC score impact measure as with the percentage meeting the expected standard.

School-level matching for both difference-in-differences models still used the percentage meeting the expected standard as a matching variable (rather than mean PSC score), but the synthetic control model weights schools using the mean PSC score. This is likely still effective in mitigating selection bias since any deliberate selection of partner schools by hubs based on phonics teaching standard may well be influenced by the percentage of year 1 pupils meeting the expected standard in each school more than the average PSC score. The results of parallel trends tests for the difference-in-differences models using PSC score are similar to those when using the percentage of pupils meeting the expected standard; there is one exception which is the pupil-level difference-in-differences model for wave 1, for which one pre-support year in which there is an apparent statistically significant impact. See the discussion of comparable parallel trends test results for models using the percentage meeting the expected standard in Annex C: Parallel trends tests.

To calculate effect size we used Hedges' *g* as described in Annex A: Detailed methods. This is consistent with the approach taken by the Education Endowment Foundation (Education Endowment Foundation, 2022)⁴⁹. Hedges' *g* is based on the assumption that the impact measure has a normal distribution, and this is not the case for pupil or school mean PSC scores. We use Hedges' *g* in spite of this, in the absence of established non-parametric alternatives suitable for results of difference-in-differences or synthetic control models (Fritz, et al., 2012). We use the Hedges' *g* effect size from pupil-level difference-in-differences impact estimates to convert to months' progress in the same way that the Education Endowment Foundation do (Higgins, et al., 2013)⁵⁰.

⁴⁹ See 'EEF statistical analysis guidance (2022)', [Evaluation design, EEF](#)

⁵⁰ [The Sutton Trust, EEF](#)

We also produced school-level impact estimates using median and various percentile PSC scores to give an indication of the extent to which the programme impact was different for pupils with low attainment.

Results

Figure 18 and Table 16 have PSC score impact estimates with confidence intervals for the first, fifth and sixth cohorts (wave 1, AY 23/24 and AY 24/25). The impact estimates for the AY 23/24 cohort are that pupils in partner schools scored approximately 1 point more in the PSC than they would have done in the absence of the programme. These are statistically significant, and we consider them valid given the parallel trends test. The impact estimates for the wave 1 and AY 24/25 cohorts are lower than for the AY 23/24 cohort, and not all statistically significant.

The effect size of programme impact for the pupil-level difference-in-differences model on the AY 23/24 cohort is 0.105 and on the wave 1 cohort is 0.035. This corresponds to up to two months' additional progress for pupils in the AY 23/24 cohort partner schools, and up to one month's additional progress for pupils in wave 1 partner schools. Effect sizes for school-level models range from 0.18 to 0.23 for the AY 23/24 cohort and are 0.12 for the wave 1 cohort.

Figure 18: Mean PSC score impact estimates by cohort

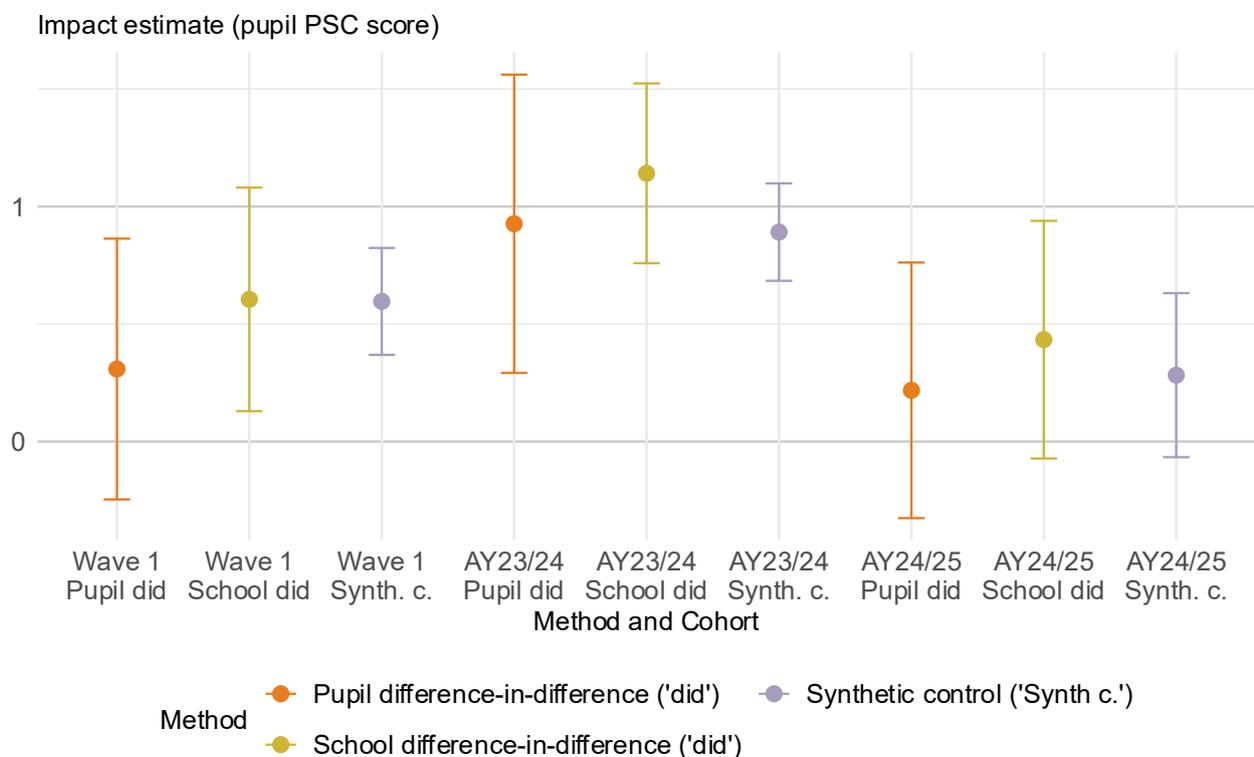


Table 16: Mean PSC score impact estimates by cohort

Cohort	Method	Central impact estimate (PSC score)	Lower confidence interval (PSC score)	Upper confidence interval (PSC score)	Effect size (Hedges' g)
Wave 1	Difference-in-differences (school-level)	0.60	0.13	1.08	0.122
Wave 1	Synthetic control	0.60	0.37	0.82	0.120
Wave 1	Difference-in-differences (pupil-level)	0.31	- 0.25	0.86	0.033
AY 23/24	Difference-in-differences (school-level)	1.14	0.76	1.52	0.230
AY 23/24	Synthetic control	0.89	0.68	1.10	0.179
AY 23/24	Difference-in-differences (pupil-level)	0.93	0.29	1.56	0.102
AY 24/25	Difference-in-differences (school-level)	0.43	- 0.07	0.94	0.087
AY 24/25	Synthetic control	0.28	- 0.07	0.63	0.056
AY 24/25	Difference-in-differences (pupil-level)	0.22	- 0.33	0.76	0.023

Impact estimates from school-level methods using school year 1 cohort percentile PSC score impact measures are larger when using percentiles corresponding to low-attaining pupils – between 1 and 3 PSC points for the wave 1 and AY 23/24 cohorts when using the PSC score of the 10th percentile pupil in each school (approximately corresponding to the fourth-lowest attaining pupil in each class), which was the percentile measure which had the highest results of those we calculated results for.

Discussion

The estimated effect size of 0.102 for the AY 23/24 cohort using the pupil-level model with pupils' PSC score is slightly lower than the effect size calculated using the percentage of year 1 pupils meeting the expected standard (0.134); the result using pupils' PSC score for the wave 1 cohort (0.035) is also slightly lower than the effect size estimated using the percentage of year 1 pupils meeting the expected standard (0.074). Effect sizes are different at school level for Hedges' g due to reduction in variation at

school level compared to pupil level; pupil level effect sizes allow for greater comparability with other studies.

The impact estimate results are a higher number of PSC points when using an impact measure specifically for lower-attaining pupils, which suggests the impact is greater for low-attaining pupils since pupils achieving low scores might typically find progress indicated by each additional PSC point more difficult; however, arguably they might alternatively find such measurable progress easier when they have a lower score as a starting point.

We have included these results using pupils' PSC score for comparison with other studies, but they are to be used with caution due to notable limitations. In particular, we are not confident in the use of months' progress as a measure of impact in itself. Using the mean PSC score as an impact measure potentially underrepresents the significance of the impact since many pupils will achieve the maximum or near the maximum score on the PSC even without the programme, and there would be no measured impact for those pupils using this method. The translation into months' progress is also dependent on an assumption about typical progress made by pupils each year, and it is unclear whether this assumption holds for progress with phonics by pupils in reception and year 1.

Further interpretation and conclusions are incorporated into the main body of this report.

Annex F: Robustness checks

We explored the sensitivity of impact estimates to a number of factors in the course of our analysis. Aspects of the methods we varied include the inclusion or exclusion of PSC results from academic year 2015 to 16 and the inclusion or exclusion of all control characteristics. Where estimates were sensitive to such choices about the method used, there is a justification for the approach taken for the main analysis.

We applied the same methods to the wave 2, 3 and 4 cohorts in addition to the wave 1, AY 23/24 and AY 24/25 cohorts. The results of these have not been included as it is not possible to control for selection bias for the waves 2, 3 and 4 cohorts as described in the 'Mitigating selection bias' section. Results of the analysis for the waves 2, 3 and 4 cohorts were similar to or higher than those for the wave 1, AY 23/24 and AY 24/25 cohorts; where results were higher, this is likely a result of selection bias.

Results were calculated using adapted versions of both the methods outlined in this report and those used in an earlier published impact analysis of the English Hubs programme (Shepherd & Fortescue, 2023) in order to establish the extent to which the difference between impact estimates in these two reports was a consequence of the change in methods. The change in impact estimate is largely due to the change in methods, for further detail see Annex G: Effect of methods change.

Effect sizes were also calculated using Cohen's *h* as well as the Cox index and Hedges' *g* (Cohen, 1988). Results using Cohen's *h* were slightly lower than those using the Cox index with the pupil-level difference-in-differences model with the percentage of year 1 pupils meeting the expected standard in the PSC, and similar to those using Hedges' *g* with the pupil-level difference-in-differences model with pupils' PSC score.

Annex G: Effect of methods change

We used methods to estimate the impact of the EHP which are a development of those used in an earlier publication (Shepherd & Fortescue, 2023). As a result, the differences in impact estimates between the earlier publication and this one are at least partly a result of the change in methods used, and do not necessarily reflect changes in the true impact of the programme since the publication of the earlier impact analysis.

The main aspects of methods used here which are different to methods used in the original analysis are controls for school mean pre-programme PSC results and three years of schools' pre-programme PSC trajectory; completing analysis of each wave separately; the use of PSC results from 2016 in addition to 2017 onwards; and the introduction of an additional method, synthetic control. The results in this report also reflect three further years of PSC results.

The change in methods accounts for most of the difference in impact estimates between the earlier publication and this one. There are no directly comparable results so we compare impact estimates from each approach for wave 1 partner schools in 2022. The range of central impact estimates for wave 1 partner schools in 2022 from the three methods in this report is 3.6 to 4.0 percentage points, and the range of central impact estimates from the methods used in the earlier publication is 6.0 to 7.3 percentage points. The difference between these ranges is exclusively a consequence of the change in methods.

A further reason the overall impact estimates for each cohort in this report are lower than in the original analysis is that they include impact over a period other than the point of graduation. The results in this report include the impact over a three-year period after intensive period ended for the wave 1 cohort, and the impact over the first year of intensive support for the AY 23/24 cohort. In contrast, the original analysis reflected the impact as in 2022, which was at the end of the full period of intensive support⁵¹.

The results in this report suggest that the recent impact of the programme is similar to that in the earlier period used in the original analysis. As noted in the Interpretation section, the range of estimates of impact at the point of graduation in this report are similar for the first and fifth cohorts (wave 1 and AY 23/24). However, there are other differences between these two cohorts including the characteristics of partner schools, the level of intensive support provided, and the contexts in which support was provided.

⁵¹ Although the impact estimates reported in the original analysis included the impact on wave 3 partner schools which had only had one year of support in the data used at that point, this would not have affected the results significantly since there were significantly fewer partner schools in wave 3 than in the combination of waves 1 and 2.

Annex H: Criteria for hubs' selection of partner schools

The following is the prioritisation criteria suggested by DfE for Hubs to review when selecting which schools to provide intensive support to:

Although not essential, Hubs should take into account the prioritisation criteria when considering a school's eligibility to become a Partner School:

- Lower than average percentage of pupils meeting the expected standard in the Phonics Screening Check (2019 PSC expected standard is 82%).
- Low attainment in bottom 20% of children
- A higher-than-average proportion of children eligible for pupil premium (2021 to 2022 the pupil premium national average is 27.3%).
- Ofsted judgement of Requires 'Improvement' or 'Inadequate';
- a high proportion of groups considered hard to reach, such as EAL, FSM, travellers, etc.
- schools referred by local partners (e.g., NLEs, RSCs, LAs);
- Schools that sit in an education investment area (EIA).



Department
for Education

© Crown copyright 2026

This publication is licensed under the terms of the Open Government Licence v3.0, except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Reference: RR1602

ISBN: 978-1-83870-760-6

For any enquiries regarding this publication, contact www.gov.uk/contact-dfe.

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