

EVALUATION PROTOCOL AND  
STATISTICAL ANALYSIS PLAN

# Alternative Provision Specialist Taskforces

RAND Europe, University of Westminster,  
FFT Education Datalab

Principal investigator: Dr Emma Disley  
Impact leads: Professor Richard Dorsett and  
Dave Thomson

# Evaluation of the Alternative Provision Specialist Taskforce (APST)

## Evaluation Protocol and Statistical Analysis Plan

Evaluating institution: RAND Europe, FFT Datalab, University of Westminster

Principal investigator(s): Dr Emma Disley



Project title	<i>Evaluation of Alternative Provision Specialist Taskforces (APST)</i>
Developer (Institution)	<i>Department for Education (DfE)</i>
Evaluator (Institution)	<i>RAND Europe, University of Westminster, FFT Education Datalab</i>
Principal investigator(s)	<i>Dr Emma Disley</i> <i>Impact leads: Prof Richard Dorsett and Dave Thomson</i>
Protocol/SAP author(s)	<i>Natalie Picken, Dr Emma Disley, Prof Richard Dorsett, Dave Thomson, Dr Elle Wadsworth, Dr Ana FitzSimons, Elena Rosa Brown, Judith Ajebon</i>
Trial design	<i>Quasi-experimental design (QED) using difference-in-differences method.</i>
Trial type	<i>Efficacy</i>
Evaluation setting	<i>State-funded alternative provision (AP) schools</i>
Target group	<i>Pupils in Year 7 to Year 11 attending state-funded AP schools</i>
Number of participants	<i>2,000 to 2,500 pupils per year in 22 AP schools</i>
Primary outcome and data source	<b><i>For pupils in Year 7-10: Re-integration into a mainstream school (NPD)</i></b> <b><i>For pupils in Year 11: Initial post-16 participation (NPD + ILR)</i></b>

**Secondary outcome and data source**

***For pupils in Year 7-11:*** Social and emotional outcomes (SDQ), Conduct and hyperactivity symptoms (SDQ), Participation in education (NPD + ILR)

***For pupils in Year 7-10:*** Attendance (NPD)

***For pupils in Year 7-9:*** Reintegration into a mainstream school (NPD)

***For pupils in Year 10:*** Reintegration into a mainstream school (NPD)

***For pupils in Year 11:*** Key Stage 4 attainment in English and maths (NPD), Sustained post-16 participation (NPD + ILR)

## Study Plan version history

**Table 1: Study plan version history**

VERSION	DATE	REASON FOR REVISION
<b>V4.0</b>	17 March 2023	Revision following inputs from Department for Education
V3.0	27 February 2023	Revision following peer review comments
v.2.0	18 <sup>th</sup> November 2022	Revision following feedback on original version
1.0 [original]	13 October 2022	Original

## Table of contents

Study Plan version history.....	3
Table of contents .....	4
Table of Tables .....	6
Table of Figures.....	6
Study rationale and background.....	8
Policy context.....	10
Theory of Change.....	12
Impact evaluation .....	16
Implementation and process evaluation (IPE).....	46
Research questions.....	46
Research methods .....	48
Cost evaluation .....	60
Research questions.....	60
Research methods .....	60
Categorising cost .....	61
Data analysis.....	62
Ethics.....	64
Ethical considerations related to the impact evaluation.....	64
Ethical considerations related to the process and cost evaluations. ....	64
Data protection .....	65
Data roles, legal bases and approaches related to the impact evaluation .....	65
Data roles, legal bases and approaches related to the process evaluation .....	68
Data security.....	68
Stakeholders and interests .....	69
Developer and delivery team .....	69
The evaluation team.....	70
Evaluation funders.....	71
Risks .....	72
Timeline.....	76
Appendix A- Alternative Provision Specialist Taskforces: Preliminary Analysis .....	80
Background.....	80
Descriptive Statistics.....	81

Matching.....	85
Regression Discontinuity .....	87
Conclusion .....	108
Addendum 1 to Appendix A: Balance achieved for each outcome.....	109
Addendum 2 to Appendix A: Data tables .....	113
Appendix B - Review Process .....	114
Appendix C - Extra information regarding the impact evaluation.....	115
C.1 Identifying Comparison Schools .....	115
C.2 Issues with absence data in NPD.....	117
C.3 Points scores used in attainment measures.....	118
Appendix D: Theory of Change .....	124

## Table of Tables

Table 1: Study plan version history.....	3
Table 2: Research questions for the impact evaluation (EQ15-21).....	16
Table 3: Design of the impact evaluation .....	20
Table 4: Outcomes available by year group .....	25
Table 5: Empirically observed MDES for administrative data outcomes .....	26
Table 6: Sample size calculations for SDQ outcomes .....	28
Table 7: Primary outcomes by cohort.....	29
Table 8: Secondary outcomes by cohort .....	29
Table 9: Covariates used in models to estimate treatment effects .....	35
Table 10: Administrative datasets for the impact evaluation .....	36
Table 11: Timeline for establishing pre- and post-treatment trends.....	39
Table 12: Evaluation questions for IPE (EQ1-14) .....	46
Table 13: Overview of methods used in the IPE.....	50
Table 14: Evaluation questions for the cost evaluation (EQ22) .....	60
Table 15: What personal data is gathered in different strands of the evaluation .....	65
Table 16: Anticipated risks, mitigation strategies and risk likelihood following mitigation (18/11/22).....	72
Table 17: Timeline of the evaluation .....	76
Table 18: Further information on calculation of outcomes .....	83
Table 19: Mean post-16 participation and re-integration outcomes by year .....	89
Table 20: Measures used to select local authorities to participate in the programme. ....	116
Table 21: Selected percentiles of the distribution of student population. ....	118
Table 22: Points scores used in attainment measures .....	118

## Table of Figures

Figure 1: Timeline for availability of cohort datasets .....	37
Figure 2: Overview of data collection approaches within the formative process evaluation.....	49
Figure 3: Distribution of the serious violence and hospital admission scores .....	89
Figure 4: School outcomes and mean, linear and quadratic fits above and below cutoff, by bandwidth and cohort .....	91
Figure 5: Characteristics of schools' pupils and mean, linear and quadratic fits above and below cutoff, by bandwidth and cohort .....	98
Figure 6: Two-step method (preferred).....	110

Figure 7: One-step method.....	110
Figure 8: Two-step method (preferred).....	111
Figure 9: One-step method.....	111
Figure 10: Two-step method (preferred).....	112
Figure 11: One-step method.....	112
Figure 12: Two-step method (preferred).....	113
Figure 13: One-step method.....	113
Figure 14: Love plot showing balance in observed characteristics between APST and comparison group schools.....	117



## Study rationale and background

The evaluation of the Alternative Provision Specialist Taskforces (APST) programme is being delivered by a consortium from three organisations: RAND Europe, the University of Westminster (UW) and FFT Education Datalab (FFT).

The Consortium will carry out an independent evaluation that will include:

- A mixed methods process evaluation that aims to understand delivery of APST and the experiences of those involved and that provides ongoing feedback and lessons learnt across the 22 alternative provision (AP) schools implementing APST.
- An impact evaluation that aims to estimate the causal effect of APST on a range of pupil outcomes (see 2.2.1.2). As randomising AP schools or AP pupils was not feasible (since the intervention schools had already been selected by DfE), the impact analysis uses a quasi-experimental approach. Reflecting the likelihood of there being unobserved differences between participating and non-participating schools, the impact evaluation will take the form of a difference-in-differences study.
- A cost evaluation that describes the costs associated with delivery of APST at both the Department for Education (DfE) level and the school level.

Within the Consortium:

- RAND Europe holds overall responsibility for the delivery of the evaluation, including project management and ensuring that all elements of the research are integrated.
- RAND Europe leads on the formative aspect of the evaluation, process evaluation, the cost evaluation, and the primary data collection to inform the impact evaluation.
- FFT and UW lead on the quasi-experimental impact evaluation, including establishing the counterfactual, linking datasets, and all outcome analyses.

For most outcomes evaluated in the impact evaluation, data will be taken from administrative sources. By definition these sources cover all participating and non-participating schools in England.

However, primary data collection to inform the impact evaluation is required to collect data on social and emotional outcomes. This is being undertaken via the Strengths and Difficulties Questionnaire (SDQ)<sup>1</sup>. It was necessary to recruit comparison schools for the purposes of evaluating this outcome measured by the SDQ. To do this, three matched comparison AP

---

<sup>1</sup> Further information regarding the SDQ may be found in YEF, Core Measurement Guidance: Strengths and Difficulties Questionnaire. Accessed online on 04 April 2023: <https://youthendowmentfund.org.uk/wp-content/uploads/2022/04/18.-YEF-SDQ-guidance-April-2022.pdf>.

schools were identified for each of the 22 participating AP schools.<sup>2</sup> We attempted to recruit one of the three potential comparison APs schools for each participating AP school. This was achieved in 15 cases, with reserve matches being recruited in 6 cases. This was considered the optimal balance between a sufficiently powered analysis (see Table 7) and fieldwork costs. Comparison schools have been asked to administer the SDQ to their students. Further details about the selection of comparison AP Schools are included in Appendix C.

## Intervention

The APST programme aims to embed teams of specialists in 22 alternative provision (AP) schools in serious violence (SV) hotspots across England.

Working with each other, other colleagues at the AP school, and local stakeholders, the specialists work to support children and young people (CYP) at the school and their families.

The aim of APST is to improve the outcomes of CYP at the AP school and to improve knowledge and confidence about supporting these students and their families and working in partnership amongst the AP school and relevant local stakeholders. This approach has the potential to meet pupil needs more quickly and holistically, reduce the need for referrals to outside services, which are already stretched, and provide opportunities for practitioners to build person-centred and trusted relationships with AP pupils due to their consistent on-site presence. Before APST, some AP settings had existing or emerging relationships with specialist practitioners, but these were often limited and had not been properly evaluated.

Drawing on funding from the Shared Outcomes Fund<sup>3</sup>, the DfE is piloting this approach over two academic years, between November 2021 and August 2023. At the outset of the programme, DfE estimated that it would reach around 2,500 KS3 and KS4 children across 22 AP schools each year.<sup>4</sup>

The DfE requires all schools implementing APST to have in place:

- Support of the Senior Leadership Team (SLT) and a designated SLT lead
- A project coordinator in post
- A team that includes at least four out of the seven types of specialists:
  - Speech and language therapist
  - Mental health professional

---

<sup>2</sup> An additional participating school was recruited in January 2022, after comparison schools had been recruited, taking the total number of participating schools to 22 and comparison schools to 21.

<sup>3</sup> Department for Education (2021) Shared Outcomes Fund Round 2. Accessed on 04 April 2023: <https://www.gov.uk/government/publications/shared-outcomes-fund-round-two>

<sup>4</sup> YEF (2021) An Invitation To Tender (ITT) for the evaluation of Alternative Provision Specialist Taskforces.

- Post-16 Transition coach
- Youth workers
- Educational Psychologists
- Family Support Workers
- Youth Offending Worker
- Adequate co-location of specialists (working at least 3 days a week<sup>5</sup> onsite and with at least 2 days of overlap between the whole team) in order to allow collaborative working and information sharing between specialists.

Beyond this, APST is a highly differentiated programme. With the support and challenge of the DfE and strategic partners and within the guiding principles above, schools design a programme that is appropriate for their school and their CYP and families' needs. This means that materials, procedures, the format and frequency of support, dosage, and the content of the support vary depending on the school.

The intended recipients of APST are CYP attending the 22 participating AP schools during the two academic years of the pilot's intervention. Schools can decide whether the provision is universal or targeted and whether any particular children or groups of children require more or less support. It is expected that having APST in the school will improve the outcomes of all students attending the school, regardless of whether they have received direct support from specialists. This is due to the work that APST specialists will do to support wider staff training and awareness about how to engage with pupils and families (captured in Activity 12 of the theory of change), to ensure better partnership working with external agencies and a more integrated whole-school approach (captured in Output 2 of the theory of change).

## Policy context

AP schools are attended by children of compulsory school age who do not attend mainstream or special schools and who would not otherwise receive suitable education, for any reason.<sup>6</sup> AP therefore provides education to pupils who are unable to attend mainstream schools<sup>7</sup> for various reasons including: school exclusion, behaviour issues, school refusal, or short- or long-term illness.

---

<sup>5</sup> Or 0.6FTE if part-time staff.

<sup>6</sup> Department for Education (2018) Creating opportunity for all: our vision for alternative provision. Accessed on 04 April 2023:  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/713665/Creating\\_opportunity\\_for\\_all\\_-\\_AP\\_roadmap.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/713665/Creating_opportunity_for_all_-_AP_roadmap.pdf)

<sup>7</sup> House of Commons Library (2019) Alternative Provision Education in England. Accessed on 04 April 2023:  
<https://commonslibrary.parliament.uk/research-briefings/cbp-8522/>

AP schools liaise with a variety of local stakeholders in terms of receiving and support pupils including:

- Local authorities: responsible for arranging suitable education for permanently excluded pupils and for other pupils who would not receive suitable education without such arrangements being made.
- Other local statutory services: who may be providing support to pupils and their families. For example, health and social care.
- Governing bodies of schools: arrange full-time education from the sixth day of a fixed period exclusion for pupils.
- Schools: arrange off-site education for some students.<sup>8</sup>

AP state place-funded provision includes schools with different governance structures: local-authority maintained pupil referral units<sup>9</sup> and AP free schools or academies. As well as state placed-funded AP, other AP may be provided by independent schools, further education colleges or tuition centres, unregistered provision, or hospital schools. Only provision in state-funded AP schools is included in this evaluation: other types of AP provision are not included in this evaluation.

Pupils in state place-funded AP have different characteristics than those in mainstream schools, with higher levels of need and disadvantage. In January 2021<sup>10</sup> of the 12,785 pupils single or dual-main registered<sup>11</sup> at state place-funded AP schools:

- 78% were in Years 9, 10 or 11 (compared to 50% in state-funded secondary schools)
- 73% were boys (compared to 50% in state-funded secondary schools)
- 71% were from White British ethnic groups (compared to 65% in state-funded secondary schools)
- Some minority ethnic groups were also over-represented in AP (including White and Black Caribbean, Black Caribbean, and Gypsy/Roma pupils)
- 53% were known to be eligible for free school meals (compared to 19% in state-funded secondary schools)

---

<sup>8</sup> DfE (2016) Statutory guidance: Alternative Provision. Accessed on 04 April 2023: <https://www.gov.uk/government/publications/alternative-provision>

<sup>9</sup> DfE (2022) Academic year 2021/22 Schools, pupils and their characteristics. Accessed on 04 April 2023: <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics>

<sup>10</sup> Further information regarding the SDQ may be found in YEF, Core Measurement Guidance: Strengths and Difficulties Questionnaire. <https://youthendowmentfund.org.uk/wp-content/uploads/2022/04/18.-YEF-SDQ-guidance-April-2022.pdf>.

<sup>11</sup> Single-registered pupils are those on roll solely at an AP school. Dual-main registered pupils are those who attend more than one school but whose main registration is at the AP school.

- 83% were identified with special educational needs (compared to 13% in state-funded secondary schools)
- 24% had an education, health, and care plan (compared to 2% in state-funded secondary schools)<sup>12</sup>

Students in AP are therefore more likely to be boys, have particular ethnic groups overrepresented, have special educational needs, and be eligible for free school meals.

We note that the published statistics for numbers of AP schools are most likely underestimates as subsidiary registered pupils (those who attend AP for part of the week but mainly attend another school) are not included and pupils join AP schools throughout the year.<sup>13</sup>

## Theory of Change

A theory of change (ToC) has been developed by the evaluation team, building on a draft prepared by the DfE and is set out in Appendix D. The outcomes & impacts, inputs, activities, and outputs are described in this section.

### Outcomes & impacts

The programme aims to achieve various outcomes (OC) including:

- Local stakeholders having better knowledge and confidence about how to address the CYP and families' needs, more understanding about local APs, and a better partnership with APs (this relates to OC 2-3 in the ToC)
- AP school staff having better knowledge and confidence about how to support CYP and families (OC4)
- AP schools having an integrated approach to supporting CYP and families with local stakeholders and generally improved partnerships with local stakeholders (OC5-6)
- CYP attending the AP school to have improved outcomes in terms of engagement with the school and education, socio-emotional wellbeing, attendance of school, likelihood of reintegration into mainstream school where appropriate, attainment, transition into post16 destinations where appropriate (OC8-13)
- CYP attending the AP school to be less likely to be involved in youth violence (OC14)

---

<sup>12</sup> Taken from DfE (2021) Alternative provision analysis. Unpublished PowerPoint by Ellis Stephenson and Hester Clarke from EYTSSAR, Department for Education.

<sup>13</sup> Dave Thomson (2019) Timpson Review Reflections. Accessed on 04 April 2023: <https://ffteducationdatalab.org.uk/2019/05/timpson-review-reflections-part-one-not-all-pupils-who-end-up-in-alternative-provision-have-been-permanently-excluded/>

We understand that the intention is that having the APST taskforce in the school will bring these benefits to every pupil in every APST school: not just those who work directly with the specialists to receive interventions. This is because APST specialists will work to upskill other members of AP schools' staff, aim to build and improve relationships between the AP school and local agencies and because the benefits that are experienced by students receiving direct support (e.g., better wellbeing, better attendance) will directly improve the school environment for all remaining children.

Envisaged longer term impacts include reductions in the number of pupils becoming NEET (not in education, employment or training) later in life, a reduction in youth offending and serious violence, and general wider societal and economic benefits. These are beyond the scope of this evaluation.

## **Inputs**

The DfE identified the 22 participating AP schools in 'serious violence' hotspots by combining two indicators of serious violence:

- hospital admissions for assault with a sharp object (April to September 2020) aggregated at lower-tier local authority level; and
- recorded crime data about the volume of serious violence offences aggregated at community safety partnership level.

DfE had converted both indicators into percentile scores with a minimum of 0 and a maximum of 1. The sum of the two scores was used to select areas to participate. The 22 participating local authorities had scores ranging from 1.83 to 2. The DfE approached the largest AP in each of the 22 top serious violence hotspots and invited them to complete an application form to be part of the programme. DfE reported that "local intelligence" was used to assure those choices, but, as no red flags about the capacity of schools to take part were raised, proceeded with the largest providers. The largest AP school in each of the 22 local authority areas with the highest scores subsequently agreed to participate.

After recruitment, the inputs<sup>14</sup> for APST include funding from the Shared Outcomes Fund (administered by the DfE), and support from the DfE to set up a taskforce (including by issuing guidance, minimum requirements). In addition, the DfE has gathered cross-government support (including from the Home Office, Department for Work and Pensions, Department of Health and Social Care, Department of Culture, Media and Sport, Youth Justice Board, and Ministry of Justice) and strategic partners. Strategic partners work with each of these department's frontline services to embed specialists in AP schools full-time to work in

---

<sup>14</sup> Please see the ToC in Appendix D for a full list of inputs.

partnership with AP leaders and their teams<sup>15</sup> and provide support to the six types of specialists by convening Communities of Practice (CoP)<sup>16</sup> and sharing learning.

The knowledge and skills of AP specialists, AP staff and local stakeholders also input into the design of the model in each school.

## Activities

The activities (A)<sup>17</sup> that make up the APST programme and are listed in Appendix D include:

- The **DfE, cross-government Programme Board, and strategic partners** provide ongoing support and challenge to APST schools and facilitate learning between taskforces by organising CoP, maintaining an online Hub, and other ad hoc events (A1, A2).<sup>18</sup>
- The **AP school SLT** support the recruitment and retention of APST specialists and then support and facilitate the taskforce's work in the school, including by ensuring they receive training and induction. The SLT also ensure ongoing monitoring and reporting of metrics to the DfE about the progress of APST (A5-7).
- The **AP school staff** work with APST specialists to support students and families (A8).
- **Local stakeholders** work with APST specialists to support students and families (A9).
- The **APST specialist teams** work with CYP and families to provide support (including by engaging them, assessing need, making referrals, designing support and delivering tailored support). They work together as one team, with other staff in the AP school and with those from local stakeholders to support CYP and families and to share learning and knowledge about supporting these students. They also may communicate with their counterparts in other APST schools to provide and receive support and training (A10-14).

These activities aim to achieve the following outputs (OP): the development of trusted relationships between CYP and their families and the APST specialists, and CYP and families receiving integrated support and having better access to timely support that meets multiple needs (OP2-4).

---

<sup>15</sup> Information provided by the Department for Education to the evaluation team.

<sup>16</sup> Online events bringing together all specialist types for shared learning. Events can take different forms depending on the strategic partner and needs of the groups but may include discussion of progress, challenges and adaptations; informal catching up; formal training and learning sessions.

<sup>17</sup> Please see the full list of Activities in Appendix D.

<sup>18</sup> Beyond these stated activities, we understand that the DfE also uses broader learning from the SEN and AP teams within the DfE to feed into the APST programme and takes learning from APST to inform other programmes: these remain outside the scope of the programme ToC.

## **Sustainability**

An emerging aspect of the APST programme is whether elements of APST can continue after the current funding period ends (November 2021 to August 2023) in a way that the AP schools feels meets CYPs' and families' needs (OC7).

At the time of writing (November 2022), the SLT in participating AP schools are investigating if and how elements of APST can continue, which involves liaising with local stakeholders, the DfE, APST specialists, and the wider AP schools (A3). Support and challenge are being provided by the DfE (A9). The Consortium understand that not all participating AP schools will necessarily continue to deliver APST beyond the end of the current funding period.

## **Evaluation activities**

In addition, the APST programme aims to generate evaluation evidence about the impact of the programme through the independent evaluation that is funded by the Youth Endowment Fund (YEF) (OC1). The ongoing data collection activities of the evaluation inform both regular formative feedback and final publishable reports. The independent evaluation of the programme consists of three components: the impact evaluation, implementation and process evaluation (IPE) and cost evaluation. The details of each component, including the statistical analysis plans, are discussed below.



## Impact evaluation

The impact evaluation is the first component of the YEF funded independent evaluation of the APST programme. It is co-led by FFT and UW. Primary data collection for the impact evaluation is coordinated by RAND Europe.

### Research questions or study objectives

The impact evaluation seeks to estimate the impact of a pupil's school participating in APST on a range of outcomes.

**Table 2: Research questions for the impact evaluation (EQ15-21) <sup>19</sup>**

<b>For pupils in Years 7-10 attending state-funded AP schools</b>	<i>EQ15 – Primary Outcome</i> What is the difference in reintegration of pupils in Years 7 to 10 in the 22 participating AP schools in comparison to those pupils in comparison schools receiving business as usual?
	<i>EQ15 – Secondary Outcome</i> What is the difference in reintegration of Key Stage 3 (Year 7 to Year 9) pupils in the 22 participating AP schools in comparison to those pupils in comparison schools receiving business as usual?
	<i>EQ15 – Secondary Outcome</i> What is the difference in reintegration of Year 10 pupils in the 22 participating AP schools in comparison to those pupils in comparison schools receiving business as usual?
	<i>EQ16 – Secondary Outcome</i> What is the difference in attendance of pupils in Years 7 to 10 measured by attendance at AP schools and state-funded schools of pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?
<b>For pupils in Year 11 attending state-funded AP schools</b>	<i>EQ17 – Primary Outcome</i> What is the difference in post-16 outcomes measured by initial post-16 destinations of pupils in schools receiving

---

<sup>19</sup> EQ1-14 relate to the IPE, see section below.

	APST in comparison to those pupils in comparison schools receiving business as usual?
	<p><i>EQ17 – Secondary Outcome</i></p> <p>What is the difference in post-16 outcomes measured by sustained post-16 destinations of pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?</p>
	<p><i>EQ18 – Secondary Outcome</i></p> <p>What is the difference in attainment, measured separately by Key Stage 4 Attainment, in English and Maths of pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?</p>
For all pupils attending state-funded AP schools in Years 7 to 11	<p><i>EQ19 – Secondary Outcome</i></p> <p>What is the difference in social and emotional outcomes measured by the total difficulties score on the Strengths and Difficulties Questionnaire (SDQ)) of pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual</p>
	<p><i>EQ20 – Secondary Outcome</i></p> <p>What is the difference in conduct and hyperactivity symptoms, as measured by the externalising score (the sum of the conduct and hyperactivity sub-scales) of the SDQ, in pupils receiving APST in comparison to those pupils in comparison schools receiving business as usual?</p>
	<p><i>EQ21 – Secondary Outcome</i></p> <p>What is the difference in participation<sup>20</sup> in state-funded education of pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?</p>

---

<sup>20</sup> Pupils in Year 11 will be classified as participating if they are in a sustained post-16 destination; pupils in Years 7 to 10 will be classified as participating if they are absent for <22% of sessions).

## Design overview

We are proposing a quasi-experimental impact evaluation of the programme based on a two-group pre/post design with impact. Estimation will use a difference-in-differences methodology. This will use data from administrative sources (National Pupil Database) where available, supplemented with primary data from the SDQ collected directly from schools. The evaluation will compare the difference in outcomes between participating and non-participating AP schools before the intervention starts to the difference in outcomes following the interventions, controlling for differences in the characteristics and school histories of pupil populations.

This choice is motivated by the need to control as fully as possible for differences between participating and other schools. If such differences influence outcomes, a simple comparison of the outcomes of children at participating and non-participating schools will not capture solely the impact of APST but will also partly reflect differences in the characteristics of schools. One option is to control for these differences directly using matching or inverse probability weighting estimation. However, this relies on having data on all potentially important school characteristics. The appeal of difference-in-differences is that differences that are consistent over time can be netted out, thereby controlling for stable unobserved influences. Furthermore, we can conduct pre-programme tests to assess the stability of such influences in the past.

Prior to preparing the statistical analysis plan (SAP), we undertook preliminary analysis of pre-treatment differences between participating and non-participating schools to explore the feasibility of different analytical choices. This is presented as Appendix A..

The aims of the preliminary analysis were to:

- Develop several outcome measures (as described in the sections on primary and secondary outcomes) that were relevant for the study population, in particular defining a measure of re-integration.
- Examine pre-existing differences between participating and non-participating areas in outcomes. We examined differences with and without pupil-level controls.
- Test three different methodological options for estimating treatment effects: difference-in-differences, matching and regression discontinuity.
- Write code to estimate treatment effects and used this to calculate them for a “placebo” year to ensure that non-significant pre-treatment differences were recovered.

Analysis of pre-existing differences for each of the proposed outcomes revealed no apparent trends, except in the case of the Key Stage 4 maths outcome. Outcomes for participating schools tended to be lower than those of comparison schools.

Based on the preliminary analysis, our proposed approach is to compare participating AP schools to all other AP schools within a difference-in-differences framework. This is the option

that maximises the data available and yields the lowest minimum detectable effect sizes (MDES) for each outcome. Additionally, trends in pre-treatment outcomes between participating and non-participating schools are largely parallel (further detail is presented in Appendix A) and placebo tests of treatment effects prior to the treatment starting were not significantly different from zero.

By definition, levels of serious violence are lower in the areas in which non-participating schools are located. The appeal of difference-in-differences is that comparison areas don't have to be similar to participating areas with regard to outcome levels, only that the trends in these outcomes run parallel.

Nonetheless, some of the comparison schools will be in areas with low levels of serious violence and this may be associated with outcomes. Therefore, for our primary outcomes, we also plan as a robustness check to apply difference-in-differences to the subset of schools that are within areas for which the serious violence score is reasonably close to the cut-off used to confer APST eligibility. This provides another means of reducing differences between participating and non-participating schools and is in the spirit of a pre-processing step.<sup>21</sup> Assuming schools close to the cut-off are relatively similar on average, this approach can be interpreted as a difference-in-regression-discontinuity (DRD) estimator, offering a stronger basis for the identification of causal effects. In practice, the DRD approach will restrict comparison schools to those in areas whose serious violence score lies within 0.2 of the threshold used to identify hotspots (this tolerance was chosen as being sufficiently wide to cover the highest serious violence score). This translates into schools in 18 local authorities outside APST areas.

We plan to evaluate a number of outcomes in line with the ToC (Appendix D). The outcomes are sourced from either administrative sources (the National Pupil Database (NPD)) or Individualised Learner Record (ILR) or from primary data collection (SDQ). By definition we have to take a different approach to our analysis of SDQ outcomes as data for past cohorts in all AP schools are not available from the NPD. Instead, we compare outcomes in participating AP schools to a set of specifically recruited matched comparison schools (Appendix C1). Consequently, the comparison AP schools and population coverage both vary with respect to outcomes as indicated in Table 3 below.

---

<sup>21</sup> Ho, D., Imai, K., King, G. & Stuart, E. (2007) Matching as nonparametric pre-processing for reducing model dependence in parametric causal inference. *Political analysis*, 15, 199– 236.

**Table 3: Design of the impact evaluation**

<b>Design</b>	QED (Two group pre/ post-test using difference in difference methodology)			
<b>Unit of analysis</b> (school, social worker, young person)	Pupil Treatment at level 2 (AP schools) with outcomes at level 1 (pupils)			
<b>Type of outcome</b>	<b>SDQ: pupils in Years 7 to 11</b>	<b>NPD: pupils in Year 11</b>	<b>NPD: pupils in Years 7 to 9/ pupils in Year 10</b>	<b>NPD: pupils in Years 7 to 10</b>
<b>Comparison group</b>	21 recruited comparison schools	All non-participating AP schools	All non-participating AP schools	All non-participating AP schools
<b>Number of schools to be included in the analysis<sup>22</sup></b>	43 AP schools (Intervention: 22, Comparison: 21)	c300 AP schools (Intervention: 22, comparison c270)	c310 AP schools (Intervention: 22, comparison c290)	c320 AP schools (Intervention: 22, comparison c300)
<b>Number of pupils to be included in analysis</b> (Intervention, Comparison)	Based on enrolment numbers for the 2020/21 academic year, we anticipate 100 pupils per year in participating AP schools and 80 pupils per year in comparison schools	Before the intervention (across years 2014 to 2021), there were on average 58 pupils per year aged 15 in participating AP schools and 23 in comparison schools (approx. 1275 pupils per	Before the intervention (across years 2014 to 2021), there were on average 48 pupils per year in Years 7 to 9 and 38 pupils in Year 10 in participating AP schools	Before the intervention (across years 2014 to 2021), there were on average 86 pupils per year in Years 7 to 10 in participating AP schools and 33 in comparison schools (approx.

<sup>22</sup> Note that not all AP schools have pupils in all year groups (Year 7 to Year 11). Consequently, the number of schools varies in each column.

		<p>(approx.2200 pupils per cohort in participating AP schools and 1680 in comparison schools)</p> <p>Numbers in participating AP schools tend to be larger than the numbers in comparison schools (this is likely because the largest school in each selected LA was recruited to be part of APST). Numbers in both participating and non-participating schools are likely to be lower during the intervention due to falling numbers in AP schools since the start of the pandemic.</p>	cohort in participating schools and approx. 6200 in comparison schools)	and 19 pupils in Years 7 to 9 and 14 pupils in Year 10 in comparison schools (approx. 1050 Year 7 to 9 and 850 Year 10 pupils per cohort in participating AP schools and approx. 5500 Year 7 to 9 pupils and 4400 Year 10 pupils in comparison schools)	1900 pupils per cohort in participating AP schools and approx. 9900 in comparison schools)
<b>Primary outcome</b>	<b>Variable</b>		<b>Post-16 participation</b>		<b>Re-integration of pupils at a state-funded mainstream school</b>

	<b>Measure (instrument, scale, source)</b>		Enrolment at a school or FE provider on 31 <sup>st</sup> October in the year following Year 11 (NPD linked to ILR)		Enrolment of 11-14 year-olds at a state-funded mainstream school in the following year (NPD)
<b>Secondary outcomes</b>	<b>Variable(s)</b>	Social and emotional outcomes  Conduct and hyperactivity symptoms	a) Key Stage 4 attainment in English and maths  b) Sustained post-16 participation  c) Participation in education	a) Re-integration of Year 7 to 9 pupils at a state-funded mainstream school in the following year  b) Re-integration of Year 10 pupils at a state-funded mainstream school in the following year	a) Attendance  b) Participation in education
	<b>Measure(s)</b>	Strength and Difficulties Total Difficulties score (primary data collection)	a) Points score in KS4 English and mathematics qualifications, (NPD)	a) Enrolment of Year 7 to 9 pupils at a state-funded	a) Attendance rate at state-funded schools in the following year (NPD)

	<b>(instrument, scale, source)</b>	Strength and Difficulties Questionnaire, Externalising score (primary data collection)	b) Enrolled at a school, college or work-based learning provider for at least 180 days continuously in the following year, (NPD linked to ILR)  (c) As b)	mainstream school in the following year (NPD)  b) Enrolment of Year 10 pupils at a state-funded mainstream school in the following year (NPD)	b) An attendance rate of 78% or higher at state-funded schools in the following year (NPD)
<b>Baseline for primary outcome</b>	<b>Measure (instrument, scale, source)</b>		None, although a set of covariates derived from NPD will be used- see the section headed other data		None, although a set of covariates derived from NPD will be used- see the section headed other data
<b>Baseline for secondary outcome</b>	<b>Variable</b>	Baseline social and emotional outcomes total difficulties score  Baseline Conduct and hyperactivity symptoms subscales score	a) Prior attainment b) None, although a set of covariates derived from NPD will be used- see the section headed other data  c) As above	None, although a set of covariates derived from NPD will be used- see the section headed other data	a) Previous attendance rate  b) As above



	<b>Measure (instrument, scale, source)</b>	<p>Strength and Difficulties Questionnaire, total difficulties score collected at baseline (primary data collection)</p> <p>Strength and Difficulties Questionnaire, conduct and hyperactivity score collected at baseline, (primary data collection)</p>	(a) Standardised KS2 attainment (NPD)		a) Attendance rate in previous year (NPD)

Table 4 summarises the outcomes available for each year group.

**Table 4: Outcomes available by year group**

	Year group				
	7	8	9	10	11
<b>Primary outcomes</b>					
Initial post-16 destinations					Y
Re-integration	Y	Y	Y	Y	
<b>Secondary outcomes</b>					
Attendance	Y	Y	Y	Y	
Key Stage 4 attainment					Y
SDQ total difficulties and conduct and hyperactivity	Y	Y	Y	Y	Y
Re-integration (Y7-9 and Y10)	Y	Y	Y	Y	
Participation	Y	Y	Y	Y	Y
Sustained post-16 destinations					Y

### **Participants**

Our proposed intervention sample for the impact evaluation includes:

- All pupils in Years 7 to 11 in the 22 APs in the academic year Sep 21-Aug 22 (“cohort 1”)<sup>23</sup>
- All pupils in Years 7 to 11 in the 22 APs in the academic year Sep 22-Aug 23 (“cohort 2”)

All pupils in Years 7 to 11 who enrol (or who are already enrolled) at each of the 22 participating AP schools are assumed to be intended recipients of the intervention. In other words, we use an intention-to-treat design. The cohorts will include all pupils who are enrolled at AP schools in each year regardless of whether they are single or dual registered and regardless of the length of time they are enrolled.

Some pupils may appear in both cohorts if their enrolment in the AP sector spans two or more academic years. For both cohorts 1 and 2, we observe outcomes in the following academic year to the year in which we observe them on the roll of an AP school. In other words, we observe outcomes in 2022/23 for pupils enrolled in the 2021/22 academic year. The exception to this is outcomes related to attainment and SDQ, which are observed in the same year as enrolment at an AP school. This is covered in greater detail in the section on outcomes below.

### **Sample size calculations**

#### **NPD Outcomes**

For NPD outcomes we use a difference-in-differences specification over multiple years. We undertook preliminary analysis and estimated MDES empirically using the observed standard

---

<sup>23</sup> For cohort 1 we stop observing outcomes after 2022/23 unless they continue in the AP sector. In which case, they will also feature in Cohort 2.

errors from placebo tests on pre-treatment data (Table 5). We outline the methodology used in Appendix A. These are presented below for both a) our main specification, which uses all comparison AP schools and b) our robustness specification (see section on robustness checks). Further information and results for the main specification are presented in Appendix A., Tables 7 and 9.

We have assumed an alpha level of 0.05. However, as there will be two primary outcomes, we will also correct for multiple comparisons using a Benjamini-Hochberg correction to control the false discovery rate (see Inference below).

As seen in Table 5, the MDES for the two primary outcomes (re-integration and post-16 participation) are both below 0.2, indicating that the analysis is expected to achieve the power threshold typically required by the EEF for school trials. However, the MDES for Key Stage 4 English (a secondary outcome) is higher and exceeds the benchmark of 0.2 for our main specification.

**Table 5: Empirically observed MDES for administrative data outcomes**

Outcome	Year groups (cohorts)	Main specification (all comparison schools)	Robustness specification (comparison AP schools in the 18 non-participating local authorities with the highest levels of serious violence)
<b>Primary outcomes</b>			
Re-integration at a state-funded mainstream school	Years 7 to 10	0.12	0.16
Initial post-16 participation	Year 11	0.08	0.12
<b>Secondary outcomes</b>			
Sustained post-16 participation	Year 11	0.09	0.12
Attendance (whole year)	Years 7 to 10	0.14	0.20
Key Stage 4 attainment in English and maths	Year 11	0.25 (English); 0.13 (Maths)	0.32 (English); 0.21 (Maths)
Re-integration at a state-funded mainstream school	Years 7 to 9	0.18	0.28

Re-integration at a state-funded mainstream school	Year 10	0.13	0.22
Participation in state-funded education	Years 7 to 11	0.07	0.06

### **SDQ outcomes**

For SDQ outcomes, we use a more conventional baseline/ endline design. A baseline SDQ is administered to all pupils who enrolled for the first time at the 22 participating and 21 comparison AP schools during the 2021/22 and 2022/23 academic years. An endline SDQ is administered to all pupils who leave a participating or comparison AP school during the 2021/22 and 2022/23 academic years or at the end of 2022/23 if they are still enrolled.<sup>24</sup> Consequently, the length of time between baseline and endline is likely to vary by pupil but this in itself may be endogenous and therefore not suitable for inclusion within the regression framework we propose. We do not plan to adjust for this, but instead note it as a feature of the data.

For SDQ power calculations, we have calculated MDES using the PowerUpR<sup>25</sup> package in R on the basis of known and assumed data in Table 6. Data on number of settings is known. We have assumed an average cohort (cluster) size of 180 per setting (200 in participating AP schools and 160 in comparison schools) over the course of two years based on historic cohort sizes and adjusted for lower enrolment numbers following the start of the pandemic. The correlation at pupil level between baseline and follow-up SDQ conduct and hyperactivity has been assumed based on the known correlation in pre-test/ post-test in the SDQ total difficulties score<sup>26</sup> and lowered to reflect the lower degree of reliability in individual subscales.<sup>27</sup>

We have assumed an intraclass correlation coefficient of 0.1 based on the values observed for NPD outcomes from our preliminary analysis (see the column headed “rho” in Appendix A, Table 7 and Table 9). However, if it transpires that it is higher (e.g., 0.2) then the MDES for SDQ outcomes would also increase to 0.28 and 0.29 respectively.

At the time of preparing this SAP, preliminary analysis of the numbers of SDQ surveys and unique UPNs returned from participating and AP comparison schools suggests that there may

---

<sup>24</sup> Due to delays in delivery and implementation, we note that endline SDQs have only been administered to students leaving the AP from February 2021 onwards.

<sup>25</sup> <https://cran.r-project.org/web/packages/PowerUpR/PowerUpR.pdf>

<sup>26</sup> <https://www.eif.org.uk/files/resources/measure-report-child-sdq.pdf>

<sup>27</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4854391/#:~:text=SDQ%20measurements&text=It%20consists%20of%20five%20subscales,relationship%20problems%2C%20and%20prosocial%20behaviours.>

be non-response in both participating and comparison AP schools, with a larger amount of non-response in the latter. Non-response may result in the average cluster being smaller than 180 (as used in Table 6). Assuming a lower average cluster size of 120 would increase the MDES only negligibly.

**Table 6: Sample size calculations for SDQ outcomes**

		SDQ Total Difficulties Score	SDQ Conduct and Hyperactivity Score
<b>Minimum Detectable Effect Size (MDES)</b>		0.20	0.21
<b>Pre-test/ post-test correlations</b>	<b>level 1</b> (participant)	0.79	0.65
	<b>level 2</b> (school)	0.5	0.5
<b>Intracluster correlations (ICCs)</b>	<b>level 2</b> (school)	0.1	0.1
<b>Alpha</b>		0.05	0.05
<b>Power</b>		0.8	0.8
<b>One-sided or two-sided?</b>		Two	Two
<b>Average cluster size (if clustered)</b>		180	180
<b>Number of schools</b>	Intervention	22	22
	Comparison	21	21
	Total	43	43
<b>Number of pupils</b>	Intervention	4400	4400
	Comparison	3360	3360
	Total	3870	3870

## Outcome measures and other data

We will use a combination of administrative data from the National Pupil Database (NPD) and primary SDQ data collected from schools.

As very little information about the outcomes of young people who access AP schools is published, we undertook preliminary analysis to define and construct a range of measures that fitted with the evaluation plan published in December 2021 (see Appendix A).

We use data on enrolments in AP schools in both 2021/22 and 2022/23 to define the cohorts whom the treatment is intended to treat. We then largely focus on their outcomes in the following year(s):

- Cohort 1 (enrolled in AP schools in 2021/22): We will observe their outcomes in 2022/23, apart from attainment for Year 11, which we observe in 2021/22.
- Cohort 2 (enrolled in AP schools in 2022/23): We will observe their outcomes in 2023/24, apart from attainment for Year 11, which we observe in 2022/23.

Table 7 and Table 8 show the primary and secondary outcomes that will be included for each cohort in the impact evaluation delivered in December 2024.

**Table 7: Primary outcomes by cohort**

Outcome	Measure	Year Groups included	Cohort 1 (2021/22)	Cohort 2 (2022/23)
Post-16 participation	Initial destinations in year 1 of post-16	Year 11	Y	Y
Re-integration	After 1 year	Years 7-10	Y	Y

**Table 8: Secondary outcomes by cohort**

Outcome	Measure	Year Groups included	Cohort 1 (2021/22)	Cohort 2 (2022/23)
Post-16 participation	Sustained destinations in year 1 of post-16	Year 11	Y	
Re-integration	After 1 year	Years 7 to 9	Y	Y
Re-integration	After 1 year	Year 10	Y	Y
Attendance	After 1 year (whole year)	Years 7-10	Y	
	After 1 year, (Autumn and Spring terms)	Years 7-10	Y	Y
KS4 Attainment	After 1 year	Year 11	Y	Y
SDQ	Total difficulties score, externalising problems subscale	Years 7-11	Y	Y
Participation	Sustained post-16 destinations for pupils in year 11 or on attendance rate of 78% or higher for pupils in Years 7 to 10	Years 7-11	Y	

Each measure for each outcome is described below. A timeline showing the release of administrative datasets necessary to calculate each outcome then follows.

### Primary Outcomes

We propose two primary outcomes:

- Initial post 16 participation – primary outcome for students in Year 11
- Reintegration into mainstream education – primary outcome for students in Years 7 to 10.

We propose two outcomes to ensure that all age groups receiving APST (Y7-11) are covered. Other outcomes which relate to all year groups were not appropriate for primary outcomes. Both primary outcomes feature in the ToC (Appendix D).

Year 10 pupils are less likely to be re-integrated than younger pupils (Appendix A, table 13). Some providers told us that their efforts on re-integration were concentrated on pupils in Key Stage 3 (Years 7 to 9). Consequently, we include re-integration of Key Stage 3 pupils as a secondary outcome.

Our preferred option would have been to use a measure of sustained post-16 destinations for Year 11 pupils. However, the necessary administrative data covering the whole of the first year of post-16 for the second APST cohort (those on roll in 2022/2023) would only become available in April 2025, outside the project timescales.

For that reason, we include an initial post-16 destination measure to act as the primary outcome in the evaluation.

Potential primary outcomes considered but not selected are as follows. These have been included as secondary outcomes:

- **Attendance:** Attendance data has also been affected by COVID-19, with some regions disproportionately more affected than others, which may in turn have an effect on the estimate of treatment effects (e.g., if participating areas are disproportionately affected). In addition, attendance is affected by missing data in NPD (see Appendix C2).
- **Attainment:** the attainment measures had some unattractive properties. The preliminary analysis for English revealed lower statistical power than for other outcomes, whereas for maths there did not appear to be common pre-treatment trends between participating and non-participating schools. We also note that data for 2020 and 2021 is unavailable due to the cancellation of national exams. However, this does not affect our analysis.
- **Conduct and hyperactivity as measured by SDQ:** attrition and missing data in collecting SDQs meant this was not selected as a primary outcome.

### ***Initial Post-16 participation***

**Measure:** Enrolment at a school or further education (FE) provider on 31<sup>st</sup> October in the year following Key Stage 4

**Population:** All pupils observed in School Census on roll at AP schools in England for 1 day or more in Year 11. Pupils observed at multiple AP schools are allocated to the first participating school at which they are observed else the first non-participating school at which they are observed.

**Years available:** 2014 to 2023

**Definition of the measure:** Using School Census for pupils in schools and Individualised Learner Record (ILR) for pupils in further education (including work-based learning) we identify pupils who on the 31 October were a) enrolled at a school or b) enrolled on one or more learning aims in ILR.

**Rationale:** A reduction in the number of young people classified as NEET (and therefore an increase in young people in education, employment and training) is one of the longer-term impacts of APST identified in the ToC (Appendix D). From our previous research, rates of initial participation among pupils who experience alternative provision are low (fewer than 60% of pupils were observed to be participating among the 2018 cohort, for example).<sup>28</sup>

### ***Re-Integration***

**Measure:** Enrolled at a state-funded mainstream school continuously for at least 180 days in the following year and spent less than 180 days in alternative provision the following year

**Population:** All pupils observed in School Census on roll at AP schools in England for 1 day or more in Years 7 to 10. For each year pupils observed at multiple AP schools are allocated to the first participating AP school at which they are observed else the first non-participating school at which they are observed. No adjustment is made for pupils observed in multiple years. Re-integration for all pupils in Years 7 to 10 (aged 11-14) will be a primary outcome. We will also analyse re-integration of a) KS3 pupils (age 11-13) and b) Year 10 (age 14) pupils separately as secondary outcomes.

**Years available:** 2014 to 2023

**Definition of the measure:** For each pupil observed attending an AP school in year  $y$ , we scan School Census and the local authority alternative provision census in year  $y + 1$  and all subsequent years. Using the leaving date at each school, and adjusting for changes in school identifiers, we calculate the number of days between the start of  $y + 1$  and the leaving date for each enrolment spell at each school (in days). We also count the total number of days spent in alternative provision (both AP schools and local authority AP). Here we combine all schools attended. Pupils observed as spending at least 180 days continuously enrolled at

---

<sup>28</sup> [https://ffteducationdatalab.org.uk/wp-content/uploads/2021/09/working\\_paper4.pdf](https://ffteducationdatalab.org.uk/wp-content/uploads/2021/09/working_paper4.pdf)



mainstream schools and less than 180 days in total in alternative provision are considered to have been re-integrated.<sup>29</sup>

**Rationale:** An increase in re-integration of pupils into mainstream schools is an outcome identified in the ToC (OC12). While reintegration is an outcome included in the ToC, our understanding is that it may not be appropriate as an outcome for Y10 students.

## **Secondary Outcomes**

### ***Social and emotional wellbeing & conduct and hyperactivity***

**Measure:** The impact analysis will measure the effect of attending a school with APST on pupils' social and emotional wellbeing (as measured by total difficulties score across all SDQ subscales) and the externalising problems sub-scale score.

**Population for baseline:** Baseline. All pupils in Years 7 to 11 in the 22 schools participating in APST and in 21 comparison schools during the 2021/22 and 2022/23 academic year were eligible to take the baseline SDQ when joining the school for the first time. This includes all pupils on roll at the start of the academic year 2021/22 and subsequent joiners.

**Population for endline:** Endline. All pupils in Years 7 to 11 in the 22 schools participating in APST and in 21 comparison AP schools during the 2021/22 and 2022/23 academic year were eligible to take the endline SDQ when leaving the school at any point from February 2021 onwards. This will include all pupils on roll in Years 7 to 11 at the end of the academic year 2022/23.

**Years available:** 2021/22 and 2022/23

**Definition of the measure:** For both baseline and endline SDQ we will calculate a) total difficulties score across all SDQ subscales (except the prosocial subscale) and b) the externalising problems score (as measured by the conduct and hyperactivity subscales).

**Rationale:** These outcomes are included as, as outlined in the ToC, the support provided by APST aims to (a) improve pupils' social and emotional wellbeing and (b) reduce youth violence, in which conduct, and hyperactivity is correlated with offending<sup>30</sup>. The SDQ is a core measure recommended by the YEF, as it supports the measurement of broader cognitive and behavioural outcomes as well as offending outcomes. The SDQ has been validated for a UK population and have shown good validity and reliability.<sup>31</sup>

---

<sup>29</sup> There are no official measures of re-integration used within the education system at present, therefore we created one for the purposes of this evaluation. We chose 180 days (in effect 6 months) as a suitable benchmark for two reasons. Firstly, it was felt from discussions with the Department for Education that this gave a reasonable indication of placement stability. Secondly, we could observe such a duration in administrative datasets that would be available within the project timescales.

<sup>30</sup> van Domburgh, L., Doreleijers, T. A., Geluk, C., & Vermeiren, R. (2011). Correlates of self-reported offending in children with a first police contact from distinct socio-demographic and ethnic groups. *Child and adolescent psychiatry and mental health*, 5(1), 22.

<sup>31</sup> <https://res.cloudinary.com/yef/images/v1623145467/cdn/18.-YEF-SDQ-guidance/18.-YEF-SDQ-guidance.pdf>

### ***Sustained Post-16 participation***

**Measure:** Participation in education or training for at least 180 days at academic age 16 (Year 12).

**Population:** All pupils observed in School Census on roll at AP schools in England for 1 day or more at academic age 15 (Year 11). Pupils observed at multiple AP schools are allocated to the first participating school at which they are observed else the first non-participating school at which they are observed.

**Years available:** 2014 to 2023.

**Definition of the measure:** Using School Census for pupils in schools and ILR for pupils in further education (including work-based learning) we calculate the maximum number of days pupils were continuously a) attending schools or b) enrolled on ILR learning aims up to the end of July in the year in which they were aged 16. We adjust for attrition by excluding any pupils who are not observed in any of the data sources at age 16.

**Rationale:** A reduction in the number of young people classified as NEET is one of the longer-term impacts of APST identified in the ToC (Appendix D). This measure covers being in sustained education, employment and training (EET) the first year of post-16 study. Rates of participation among pupils who experience alternative provision are low.<sup>32</sup>

### ***Attainment***

**Measures:** Key Stage 4 score English score, Key Stage 4 maths score.

**Population:** All pupils observed in School Census on roll at AP schools in England for 1 day or more at academic age 15 (Year 11). Pupils observed at multiple AP schools are allocated to the first participating school at which they are observed else the first non-participating school at which they are observed.

**Years available:** 2014 to 2019, 2022, 2023.

**Definition of the measure:** Grades are converted into points as follows as per Annex G.

**English:** Based on grades in GCSE English language and other general qualifications in literacy or communication.

**Maths:** Based on grades in GCSE maths and other general qualifications in numeracy.

**Rationale:** Increased attainment is an outcome identified in the ToC (OC13). English and maths are the two subjects that all pupils must study. Consequently, outcomes can be observed for all pupils. Exams are marked externally and quality assured by awarding bodies. Furthermore, grades in English and maths are associated with successful transition to post-16 study, consistent with the longer-term aim of APST to reduce propensity to be not in education, employment or training post-16.

---

<sup>32</sup> [https://ffteducationdatalab.org.uk/wp-content/uploads/2021/09/working\\_paper4.pdf](https://ffteducationdatalab.org.uk/wp-content/uploads/2021/09/working_paper4.pdf)

## **Attendance**

**Measure:** Rate of attendance in the following year.

**Years available:** 2014 to 2018, 2020-2023. Due to the COVID-19 pandemic resulting in schools being closed to the majority of pupils in state schools in the 2019/20 academic year, there is no endline absence data for the 2019 cohort. Data for the 2020 cohort is also partially affected by school closures due to COVID-19.

**Population:** All pupils observed in School Census on roll at AP schools in England for 1 day or more in Years 7 to 10. For each year pupils observed at multiple AP schools are allocated to the first participating school at which they are observed else the first non-participating school at which they are observed. No adjustment is made for pupils observed in multiple years.

**Definition of the measure:** For each pupil observed attending an AP school in year  $y$ , we scan absence data for year  $y + 1$ . We sum the following:

- Sessions absent due to authorised absence ( $a$ )
- Sessions absent due to unauthorised absence ( $b$ )
- Total possible sessions of attendance ( $c$ )

*The absence rate for a pupil is  $(a + b)/c$ .*

The attendance rate for a pupil is  $1 - ((a + b)/c)$ .

Due to lags in absence data for year  $y + 1$  being made available, we also propose to use a measure of attendance based on the Autumn and Spring terms only in order to deliver the impact evaluation within the timescales of the project.

**Rationale:** Improvement in attendance at schools is an outcome identified in the ToC (OC10). However, we have a number of concerns about the quality of absence data in NPD. Firstly, there is the uneven impact of COVID-19 on absence, with London appearing to suffer less impact compared to other regions. In addition, we noted two additional problems with absence data during our preliminary analysis (not all pupils appear in the absence data for the following year and there is variation among the group of pupils for whom absence is recorded). These are summarised in Appendix C.

## **Participation in state-funded education**

**Measure:** A composite indicator based on a) absence for pupils in Years 7 to 10 and b) post-16 participation pupils in Years 11.

**Years available:** 2014 to 2018, 2020-2023. There is no endline absence data for the 2019 cohort. Data for the 2020 cohort is affected by school closures due to COVID-19.

**Population:** All pupils observed in School Census on roll at AP schools in England for 1 day or more at academic age 11 to 15 (Years 7 to 11). For each year pupils observed at multiple AP schools are allocated to the first participating school at which they are observed else the first non-participating school at which they are observed. No adjustment is made for pupils observed in multiple years.

**Definition of the measure:** A binary measure of participation in education in the following year. For Year 11 pupils we use the sustained post-16 participation measure described above. For Year 7 to Year 10 pupils, we define participation has having an absence rate lower than 22%. This threshold was selected as the proportion of Year 7 to Year 10 pupils with this level of absence equalled the proportion of Year 11 pupils in sustained post-16 participation in the most recent year for which data was available during our preliminary analysis.

**Rationale:** This is an attempt at producing a measure from administrative data that is applicable to all pupils within the scope of the APST programme. Both attendance and post-16 participation are identified within the ToC. However, this is a speculative measure and has lower precedence than the other proposed outcomes.

### Other data

A set of pupil-level measures available in NPD and known to be associated with outcomes<sup>33</sup> were included as covariates in models where the outcome is sourced from NPD (see Appendix A). These were used to adjust for differences in the pupil populations served by AP schools. We plan to use these covariates in the evaluation for both NPD and SDQ outcomes. They are listed in Table 9 below.

**Table 9: Covariates used in models to estimate treatment effects**

Outcome					Control	Description
Attainment	Post-16	Attendance	Reintegration	SDQ		
Y	Y	Y	Y	Y	female	pupil is female
Y	Y	Y	Y	Y	social_care	pupil has ever been in need or looked after
Y	Y	Y	Y	Y	also_mainstream	pupil also attended a mainstream school in the year in question
Y	Y	Y	Y	Y	first_ap_age_gp	academic age pupil first enrolled at an AP school
Y	Y	Y	Y	Y	prior_sen	highest SEN category before joining current AP school
Y	Y	Y	Y	Y	prior_perm_ex	pupil had been permanently excluded prior to joining current school
Y	Y	Y	Y	Y	prior_suspensions	number of suspensions prior to joining current AP school

<sup>33</sup> <https://ffteducationdatalab.org.uk/2021/09/investigating-alternative-provision-part-one/>

Y	Y	Y	Y	Y	prior_fsm_percent	% of terms aged 5-15 observed as eligible for free school meals in School Census
Y	Y	Y	Y	Y	prior_sen_percent	% of terms aged 5-15 observed recorded as SEN in School Census
Y	Y	Y	Y	Y	prior_idaci	IDACI score of home postcode
Y	Y	Y	Y	Y	ethcode	Ethnic background
Y	Y	N	N	Y	ks2_avz	Mean standardised KS2 score in English and maths
N	N	Y	N	N	abs_pre1	Absence rate in previous year
N	N	N	Y	N	months_ap_pre1	Months enrolled at an AP school in previous year

After primary data collection of SDQ scores is completed, pupil-level baseline and endline scores together with identifiers (name, UPN, date of birth) will be securely transferred from RAND Europe to DfE. DfE will match pupils to the anonymous pupil identifier in NPD. They will then deposit a file of SDQ scores linked to anonymous NPD identifiers in SRS for FFT Datalab and University of Westminster to use within the evaluation alongside the administrative datasets listed in the following section.

### ***Administrative datasets***

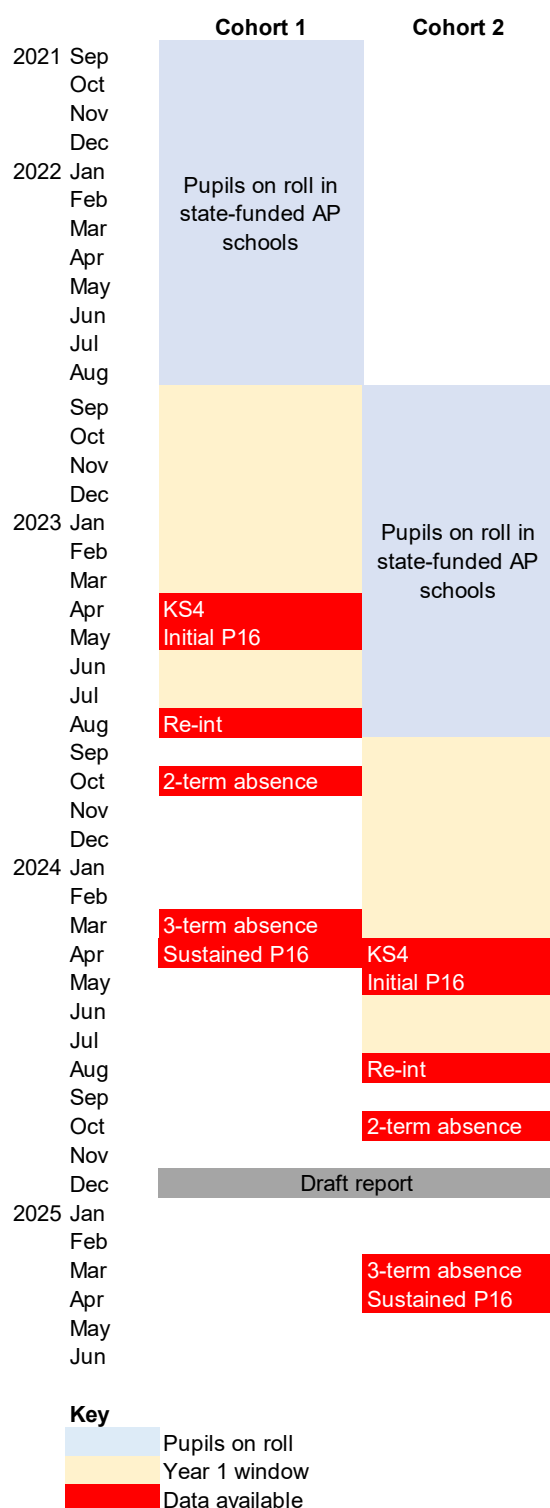
Table 10 shows which administrative datasets we plan to use to define each of the outcomes.

**Table 10: Administrative datasets for the impact evaluation**

<b>Initial post-16 destinations</b>	Autumn Term School Census ILR SN04
<b>Re-integration</b>	Autumn Term School Census Spring Term School Census Summer Term School Census
<b>Sustained post-16 destinations</b>	Autumn Term School Census Spring Term School Census Summer Term School Census ILR SN14
<b>Key Stage 4 Attainment</b>	Final Key Stage 4 Pupil and Exam data
<b>Attendance (2-term)</b>	2-term Absence
<b>Attendance (3-term)</b>	3-term Absence

The timeline below shows when the relevant datasets for each cohort become available to researchers from the Department for Education.

**Figure 1: Timeline for availability of cohort datasets**



### ***Selection of the comparison group and identification assumptions***

The schools participating in the programme were selected by the DfE using two measures of serious violence:

- Number of hospital admissions for assault with a sharp object (all ages) April to September 2020, measured at lower-tier local authority level.

- Number of police-recorded serious violence offences in calendar year 2020, measured at community safety partnership level (CSP).

CSPs are largely (but not exactly) coterminous with lower-tier local authorities. For those that were coterminous, percentile scores were calculated for each measure and summed by DfE. The areas with the 22 highest scores were selected to participate. In each area, the largest AP school was selected to participate in the programme.

### ***NPD outcomes***

All non-participating AP schools (both in selected and non-selected areas) form the basis of the comparison group for the purposes of the difference-in-differences analysis of outcomes available in administrative data.

However, as a robustness check, we will also restrict the analysis to the schools in the 18 non-APST areas with the highest levels of serious violence (the DRD estimator mentioned above). This is motivated by the intuition that schools in areas that are more similar on this definition are likely to have outcomes that are more similar to those that would be expected in non-APST areas absent the treatment.

We also gave some consideration to using non-participating AP schools in the local authorities selected to participate in APST but ruled this out for a number of reasons. Firstly, by definition, they have fewer pupils. Secondly, 8 of the participating areas did not have any other AP schools covering the full secondary range (i.e., pupils up to age 16). Thirdly, 16 of the 30 non-participating AP schools in participating areas are free schools. These tend to serve different pupil intakes than pupil referral units and alternative provision academies. Finally, we were concerned about the possibility of endogenous sorting, in other words, the availability of APST interventions may influence placement decisions in participating local authorities.

### ***SDQ outcomes***

As SDQ outcomes are not present in the administrative sources used, we will use a different comparator group when analysing them. This involved recruiting a set of 21 comparison schools in which to administer baseline and endline SDQs during 2021/22 and 2022/23. The motivation here is that the comparison schools serve pupil intakes with similar needs and potential outcomes as the schools participating in APST.

The Study Team undertook the following steps to identify comparison schools.

- Initially, all 320 open state-funded AP schools in England were included as possible comparison schools (minus the 21 already participating in APST at that point<sup>34</sup>). This includes other AP schools in the participating areas.

---

<sup>34</sup> A further school was subsequently recruited to participate in APST.

- Any schools that did not cover the full 11-15 age range were removed.
- The remaining AP schools were linked to the area-level measures of serious violence above and measures of local area population size. Other school-level measures relating to attainment, post-16 destinations, absence and pupil characteristics were available up to the end of 2019 from a separate research project undertaken by one of the study team.<sup>35</sup>
- Based on this, three matched schools were selected for each of the 21 participating AP Schools.
- The study team attempted to recruit one matched school for each of the 21 participating AP schools. In six cases, reserve matches had to be found due to none of the three initial matches agreeing to participate.

An incentive of £1,000 in total will be provided to the 21 comparison AP Schools for their participation in the evaluation: £500 was paid to comparison AP schools upon completion of the first round of baseline SDQ by April 2022, and £500 will be paid upon completion of endline SDQ surveys in September 2023.

Further details of the matching routine can be found in Appendix C.

## Analysis

### Difference-in-Differences (DiD)

In the preliminary analysis (see Appendix A), we establish a seven-year pre-treatment trend in outcomes between participating schools and non-participating schools (and separately for the matched comparison set) as per Table 11 below. We do this for all outcomes except SDQ for which we will only have a baseline score for the year in which the treatment started.

**Table 11: Timeline for establishing pre- and post-treatment trends**

Academic Year									
2014/ 15	2015/ 16	2016/ 17	2017/ 18	2018/ 19	2019/ 20	2020/ 21	2021/ 22	2022/ 23	2023/ 24
Pre-treatment							Post-treatment		

The treatment is assumed to begin at the same time in all 22 participating AP schools. We assume that differences in outcomes among the treatment group compared to the non-treatment group are due to the intervention. The idea is that the non-treatment group can capture the expected growth in mean outcomes absent treatment so netting this from growth seen in the treatment group allows the impact of treatment itself to be identified.

<sup>35</sup> <https://ffteducationdatalab.org.uk/2021/09/investigating-alternative-provision-part-one/>



For this to be credible requires that trends in the non-treatment group would have applied to the treatment group had the treatment not occurred. Pre-treatment trends for outcomes measured using administrative data are shown in Appendix A, Tables 6 and 8. Briefly, with the exception of Key Stage 4 maths, we do not observe any important pre-treatment trends for the NPD outcomes using a difference-in-differences specification with all comparison schools and the controls listed in Table 9. For Key Stage 4 maths, the placebo test of a significant effect in the placebo year of 2019 is not significant( Appendix A , Table 7). However, we propose a robustness test controlling for the prior trend. This is by way of robustness check; the headline estimates will remain those of the primary model specified above.

For each NPD outcome, we will use a pupil-level estimation dataset containing up to seven pre-treatment cohorts and two post-treatment cohorts for each school. Imbalance between schools in pupil populations will be explicitly controlled for using a set of covariates known to be associated with outcomes (Appendix A, Table 10). Some pupils may attend more than one AP school. In such cases, we will regard the first AP school they attend each year as fixed for that year. To assess sensitivity of the results to this, we will conduct a separate analysis whereby such pupils are excluded.

## Primary analysis

### *Primary & secondary outcome analysis*

Assuming intention-to-treat all pupils in the 22 participating AP schools, we will estimate the school-level average treatment effect on the treated (ATET) using Equation 1.

### Equation 1

$$y_{ist} = \beta \cdot X_{i0} + \mu_s + \phi_t + \gamma \cdot POST_t \cdot T_{si} \cdot POST \cdot T_i \quad (1)$$

In the equation above,  $y_{ist}$  represents an outcome for pupil  $i$  in setting  $s$  at time  $t$ . Here,  $\mu_s$  is the setting fixed effect,  $\phi_t$  is a time effect,  $POST_t$  indicates that  $t$  falls after the introduction of the treatment and  $T_{si}$  indicates that pupil  $i$  is in a treatment setting.  $X_{i0}$  is a set of covariates (see section headed other data above) summarising pupils' educational and social care histories prior to treatment (the 0 in the subscript is to emphasise that all such controls pre-date the treatment).

This is the standard two-way fixed effects regression. Since treatment is at the setting level, we will use standard errors that take account of clustering at the AP level, following the guidance of Abadie et al (2023).<sup>36</sup> The coefficient of interest is  $\gamma$ , which provides the estimate of ATET. As written, it can be estimated over multiple years. With outcomes taken from administrative data, we have up to seven years of data, for SDQ outcomes we have two periods (baseline and endline).

---

<sup>36</sup> Athey, Abadie, Imbens and Wooldridge (2023) "When Should You Adjust Standard Errors for Clustering?" The Quarterly Journal of Economics, Volume 138, Issue 1, Pages 1–35.

The advantage of DiD relative to before-after estimators is that it controls for changes that would have happened anyway. This is particularly important when ambient conditions are unstable, such as during or after a pandemic.

The primary analysis will include all AP schools with Year 7 to 11 pupils in the dataset. As we set out in the conclusion to the preliminary analysis set out in Appendix A, we prefer this approach to other DiD approaches we tested (e.g., the regression discontinuity DiD) as treatment effects were estimated most precisely. However, we will also run the analysis of primary outcomes on the subset of schools in non-participating areas with the highest levels of serious violence (see the section on robustness checks below) using the regression discontinuity DiD.

In preparing the study plan, placebo tests of a treatment effect in the pre-treatment period have been performed. Results are shown in Appendix A Tables 7 and 9. As we should expect, all of these return a non-significant treatment effect of attending an AP school prior to treatment.

For outcomes where the preliminary analysis suggests pre-treatment trends in outcomes (e.g., Key Stage 4 maths) we will run an alternative difference-in-differences specification in which the linear trend,  $L$ , is explicitly modelled and allowed to vary with treatment status.

#### **Equation 2**

$$y_{ist} = \beta \cdot X_{i0} + \mu_s + \phi_t + \delta \cdot L + \delta \cdot t \cdot T_i + \gamma \cdot POST \cdot T_i. \quad (2)$$

Models will be fitted using STATA.

#### *Inference*

Treatment effects will be reported with standard errors, clustered by school, along with p-values of a t-test that the treatment effect is not equal to zero.

As we have two primary outcomes, albeit for different populations, we will use the Benjamini-Hochberg procedure to control the false discovery rate. This will be set at 5 %.

#### *Robustness checks*

For primary outcomes, we will evaluate the sensitivity of our results to different analytical choices.

Firstly, the difference-in-differences estimates use all non-participating AP schools as the comparison group. However, the 22 participating AP schools are located in areas with the highest rates of serious violence. By definition, comparison schools are located in areas with lower rates.

In order to test the robustness of our findings to differences in the definition of the comparison set of schools, we will run a version of the difference-in-difference specifications in which we restrict the set of schools included in the comparison set to all those in the 18 non-participating local authorities with the highest rates of serious violence based on the index used by DfE to select areas for APST. Formally, we include local authorities within a cut-

off of 0.2 index points below the lowest-ranked participating area (in this case, Ealing). An exploration into the feasibility of this approach is provided in Appendix A.

There are two other decisions arising from our preliminary analysis that we would also wish to test the robustness of findings to:

- Comparison schools tend to be smaller than participating schools. Consequently, we will test the effect of school size by including it as an additional control in our main specification.
- Some pupils attend more than one AP school each year. For our main specification, we allocate pupils to the first school they attend. However, we will undertake a robustness specification in which pupils who attend more than one AP school are excluded.

Clearly, a 'failed' robustness check sounds a note of caution for our results. Depending on the nature of the finding, it may be a challenge to fully reconcile results. However, our reporting will have to interpret the results in the context of the robustness checks. Of the three checks mentioned:

- Should impact estimates using only those local authorities with the highest rates of serious violence based as comparators differ from those using all local authorities as comparators, this implies something about the nature of area/school selection
- Should including the school size variable in the regression alter the estimated impact, we would need to reflect on why differencing does not account for this
- Should results when excluding pupils who attend more than one AP differ from the main results, this may tell us something about the nature of that subgroup of pupils

As emphasised earlier, the headline estimate will remain that from the primary mode, consistent with YEF guidance.

### **Subgroup analyses**

We will produce sub-group analyses for the primary outcomes using our main specification. We will interact the treatment effect with the following variables:

- Cohort (1 - 2021/22 or 2- 2022/23)
- Gender (male/ female)
- Ethnic group (White vs non-white)
- Ever been looked after or in need (yes/ no)

The statistical power of these sub-group analyses will likely be lower than when considering full sample impact estimates and will thus be considered exploratory. Consequently, we only

propose to use the broad groupings listed above. Indicative numbers of pupils in participating and non-participating schools can be found in Appendix A, Tables 11 and 12.

### **Treatment effects in the presence of non-compliance**

We assume that all AP schools selected to participate in APST did so. During the course of the APST programme, DfE will collect information on the deployment of specialists in 22 participating AP schools: for example, the types of specialists in place, the types of interventions they offer, and the number of hours offered.

Schools meeting the following criteria will be considered to be meeting the minimum level of service:

1. **Having at least 4 specialists in place from the core list of specialists** (youth worker; family worker; speech and language therapy (SALT) worker; mental health worker; post-16 worker; youth offending worker);
2. **These specialists being co-located:** specialists will need to work at least 3 days a week onsite in the AP school; if there is more than one person working part time or out of hours there should be at least 2 days of overlap between the whole team to allow collaborative working and information sharing between specialists;
3. **Having a SLT who oversees APST;** and
4. **Having a project coordinator** <sup>37</sup>

As the programme will take place over two years, it is possible that some AP schools may deliver the minimum level of service for only part of the two-year period. Towards the end of the project, we will use any information provided to us by the DfE over the two years of the project to define a period over which the minimum level of service will be measured. Options might include:

- Any school where they had this in any place at any point in 2021/22 or 2022/23
- Any school where this was in place for at least 50% of the time in 2021/22 and 2022/23

If some participating schools do not meet this minimum level of service, we will run an additional primary outcome analysis on the subset that do meet the minimum level.

Alternatively, should the DfE be able to identify the earliest date on which each AP school had the minimum level of service in place, we could divide the schools into two groups of 11 schools based on this date to compare the outcomes of the early adopters vs the later adopters.

---

<sup>37</sup> These criteria were described by DfE at the ToC workshop on 15 August 2022 and form the basis of EQ1.

These analyses can only be undertaken should the DfE be able to provide clear data about the 4 criteria above on each of the 22 schools for the entire programme period as part of the metrics provided in DS5. At present, it is not clear to the study team that information on criteria 2, 3 or 4 is available via these metrics. No extra data collection by the evaluation team is planned. Consequently, there is a risk that assessing treatment effects in the presence of non-compliance becomes infeasible due to missing or partial data.

### **Missing data**

By definition, no outcome data for attainment, re-integration and post-16 destinations will be considered missing. Instead, we will assume missing outcomes reflect pupils who did not achieve qualifications, join a state-funded mainstream school, or engage in sustained post-16 education respectively. This risks some degree of bias due to emigration (including to other parts of the UK). However, this is indistinguishable in the data from remaining in England and disengaging from education.

Some SDQ data is likely to be missing. Based on data collected for the 2021/22 academic year and dividing the number of UPNs collected by the estimated number of pupils eligible for inclusion in the SDQ sample implies fuller coverage in participating than non-participating schools; UPNs were collected for 93% of those estimated to be eligible in participating schools compared to 66% for non-participating schools. Among pupils for whom a UPN is collected, the proportion providing baseline SDQ data is similar among participating and non-participating schools (95% and 98% respectively). This is also true for endline SDQ data albeit (so far) at a much lower rate (25% and 23%). We estimate that this would translate into matched baseline and endline data being available for 88% and 64% of eligible pupils.

The two main concerns are therefore the imbalance in coverage and the low proportion of children providing endline SDQ. With regard to the first of these, the worry is that children are selecting non-randomly into the SDQ group in non-APST schools. We can attempt to control for this but cannot take account of unobserved characteristics that affect the probability of providing an SDQ response. Consequently, it is difficult to assert that impacts based on the achieved SDQ sample are capturing causal effects rather than compositional differences.

Missing SDQ endline data may arise in the following circumstances:

- A treated or comparison school drops out of the project
- A pupil is long-term absent
- A school only administers the tests for a subset of year groups
- A pupil leaves a school before the test can be arranged

Post-treatment absence data may also be missing in cases where pupils leave the state-funded school system (e.g., move into home education, independent AP or emigrate).

Where necessary, we will summarise the extent of missing data using cross-tabulations. This will include breakdowns by pupil-level and school-level characteristics, and an assessment of the extent to which missing data is concentrated within particular schools. We will use multiple imputation (MI) to reduce sample loss arising from missing covariates.

We do not propose to impute missing outcome data for our main estimates. This reflects the fact that MI relies on missingness being explained by observed characteristics and so cannot capture the effect of unobserved factors associated with being in the treatment group. While this applies also to imputation of covariates, it is likely to influence results more substantially in the case of outcome variables. If more than 10% of outcomes are missing, we will use logistic regression to test how well missingness can be predicted by the covariates. As a sensitivity analysis, we will present results where outcomes are imputed.

### Effect size calculation

For NPD outcomes, the effect size for each outcome is the difference-in-differences estimate  $\gamma$  from equation (2) divided by the AP population standard deviation in the outcome for all pre-treatment years (i.e., years unaffected by the treatment) in the dataset combined.

#### Equation 3

$$ES = \frac{\gamma}{\sigma_{pre}}$$

The lower and upper confidence intervals for each treatment effect will also be divided by the population standard deviation to calculate confidence intervals for the effect size. For binary outcomes, risk ratios will also be presented.

For SDQ outcomes, population standard deviations will not be available. Consequently, we will use Hedges'  $g$ , in which estimated effects are divided by the pooled and weighted standard deviation in baseline SDQ scores pupils in participating and comparison schools.

#### Equation 4

$$ES = \left(1 - \frac{3}{4(n_T + n_C) - 9}\right) \cdot \frac{\gamma}{\sqrt{\frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T + n_C - 2}}}$$

Where  $s_T^2$  and  $s_C^2$  are the variances in SDQ outcomes in participating and comparison schools respectively and  $n_T$  and  $n_C$  represent the number of pupils in participating and comparison schools respectively.

The lower and upper confidence intervals for each treatment effect will also be divided by the denominator shown above and then multiplied by the correction factor above.

## Implementation and process evaluation (IPE)

The implementation and process evaluation (IPE) is the second component of the independent evaluation funded by the YEF. It is led by RAND Europe.

We intend to deliver a mixed-methods process evaluation that aims to understand delivery of APST and perceptions of outcomes. We intend to provide ongoing formative feedback through three reflection points (September 2022, February 2023, November 2023) and a final summative report incorporating overall findings (to be delivered November 2023).

### Research questions

The evaluation questions (EQs) set out in the table below explore the implementation of APST and the processes that underpin APST's functioning.

Through exploring these EQs, we intend to consider the extent to which OC2-6 and OC 14 in the ToC (Appendix D) are achieved.

**Table 12: Evaluation questions for IPE (EQ1-14)<sup>38</sup>**

	Implementation and operation of the APST
EQ1	To what extent were the APSTs implemented <i>as planned</i> in relation to (i) co-location of teams (ii) the inclusion of at least four different specialists (including transition coach, speech & language, mental health (including counsellors), youth offending, youth worker, family support, educational psychologists, social worker) (iii) having a project coordinator and (iv) having a designated SLT lead?
EQ2	How do APST specialists work with children and young people and families as part of APST on a day-to-day basis?
EQ3	What were the barriers and facilitators that affected implementation and operation of APST as planned? What, if any, adaptations were made to the implementation and operation of APST as planned?
EQ4	To what extent does implementation and operation of the APSTs differ between AP schools and why?
EQ5	To what extent do AP schools have plans to continue delivering elements of APST in the AP school after the end of the DfE-funded pilot? What are the barriers and facilitators to sustainability?

---

<sup>38</sup> These research questions were refined following the first round of process evaluation data collection with the aim of consolidating and clarifying the questions posed, to ensure that they are clear, comprehensive, and gather the information that is important to the operation of APST.

<b>EQ6</b>	To what extent do APSTs account for and respond to diversity in students' ethnicities and genders?
	<b>Partnership working within the APST team</b>
<b>EQ7</b>	<p>To what extent did APST specialists within the same schoolwork in partnership as a team?</p> <p>To what extent and how did APST specialists working in partnership within the same school make a difference?</p> <p>What were the barriers to partnership working between APST specialists within the same school? What adaptations were made?</p> <p>What were the facilitators to partnership working between APST specialists within the same school?</p>
	<b>Communication between APST taskforces in different schools</b>
<b>EQ8</b>	<p>To what extent did APST specialist teams and AP SLT leads from different AP schools communicate with each other?</p> <p>To what extent and how did this communication make a difference?</p> <p>What were the barriers to communication between different APST teams? What adaptations were made?</p> <p>What were the facilitators to communication between different APST teams?</p>
	<b>Partnership working between APST team and AP school</b>
<b>EQ9</b>	<p>To what extent did APST specialist teams and AP staff (including the APST SLT lead) coordinate and work together?</p> <p>To what extent and how did this coordination and working together make a difference?</p> <p>What were the barriers to coordination and working together? What adaptations were made?</p> <p>What were the facilitators to coordination and working together?</p>
	<b>Partnership working with local stakeholders</b>
<b>EQ10</b>	<p>To what extent did APST specialist teams and local stakeholders effectively coordinate and work in partnership with each other?</p> <p>To what extent and how did coordinating and working in partnership make a difference?</p> <p>What were the barriers to coordinating and working in partnership? What adaptations were made?</p> <p>What were the facilitators to coordinating and working in partnership?</p>



	Outcomes <sup>39</sup>
EQ11	To what extent and how was APST perceived by relevant stakeholders to contribute to the stated outcomes [that are being measured by the impact evaluation]?
EQ12	To what extent and how was the APST perceived by relevant stakeholders to contribute to parental and pupil engagement with the AP school and education?
EQ13	To what extent and how was APST perceived by relevant stakeholders to have the potential to reduce youth violence amongst CYP attending the AP during the length of the APST pilot?
EQ14	To what extent did the APSTs result in unintended consequences for all stakeholders during the length of the APST pilot?

## Research methods

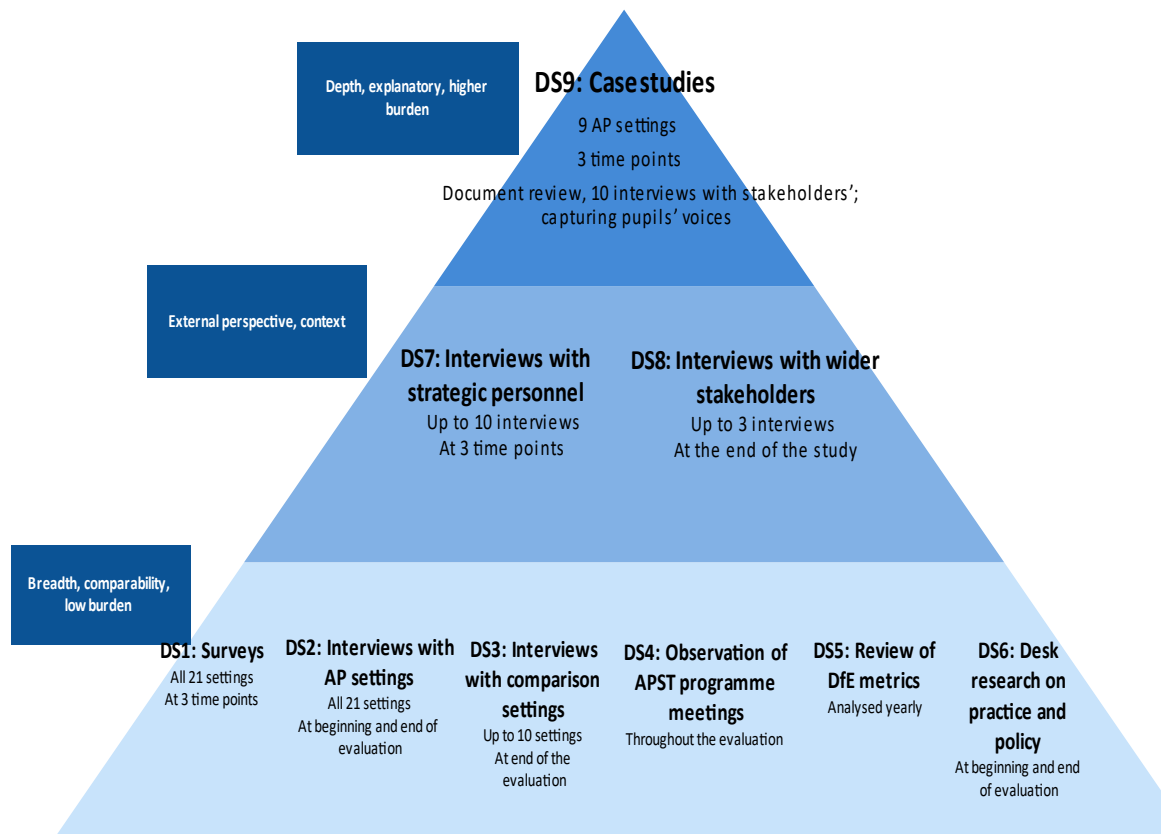
The Consortium propose a tiered approach to data collection as part of the process evaluation. We intend to use nine different data sources (DS) in order to answer our evaluation questions.

Figure 2 shows how the proposed combination of methods aims to achieve depth, breath and understanding of the wider context. Table 14 sets out the research methods and associated data collection methods, participants and data analysis methods. Below the table, the methods are described in further detail, providing information on the sampling strategy, timing, and rationale for each method.

---

<sup>39</sup> While exploring the extent to which outcomes were achieved is key to the impact evaluation, the parts of the questions below referring to 'how' APST was perceived to contribute to outcomes is key to the implementation and process evaluation.

**Figure 2: Overview of data collection approaches within the formative process evaluation**



**Table 13: Overview of methods used in the IPE**

Research methods	Data collection methods	Timing	Participants/ data sources (type, number)	Rationale for interview
<b>Interviews</b>	Interviews with SLT at the 22 participating AP schools x 2 (DS2)	Round 1: in early implementation (November 2021-February 2022) Round 2: towards the end of implementation (June 2023-July 2023)	We will invite all 22 members of SLT who oversee the task force in each APST school to take part in the interview <sup>40</sup>	To understand more about implementation and operation of APST, perception of outcomes, and perception of partnership/collaborative working
<b>Interviews</b>	Interviews with SLT from the comparison schools x 1 (DS3)	Towards the end of implementation (June 2021-July 2022)	We will aim to interview up to 10 members of SLT from the comparison schools. We will use purposive sampling to select interviewees from a range of schools (considering geographical location, type of school, size of school)	To understand more about the AP practice context by understanding the “business as usual” offer in other AP schools
<b>Interviews</b>	Interviews with strategic personnel x 3 (DS7)	Round 1: in early implementation (April 2022)	We aim to interview 6 individuals in each round (18 interviews in total). Interviewees will include those working in the DfE (policy,	To understand more about implementation and operation of APST, perception of outcomes,

<sup>40</sup> We carried out interviews with 16/22 SLT in Round 1.

		<p>Round 2: towards the beginning of the second year of implementation (October 2022)</p> <p>Round 3: towards the end of implementation (March 2023)</p>	<p>delivery, and programme management teams) and those working for strategic partners</p>	<p>and perception of partnership/collaborative working</p>
<b>Interviews</b>	Interviews with stakeholders in broader AP context x 1 (DS8)	Towards the end of implementation (June 2023-July 2023)	<p>We will aim to interview 3 stakeholders who have insights into AP policy and practice. We envisage approaching senior practitioners (e.g., those involved in the AP Stakeholder group and PRUsAP), other funders working on AP and school exclusions and in the policy space. We will be guided by advice from YEF, DfE, and emerging findings and remain flexible to ensure these are as useful as possible</p>	<p>To understand more about the AP policy and practice context</p>
<b>Case studies</b>	Three rounds of 3 case studies, each focusing on one APST taskforce	<p>Round 1: in early implementation (June 2022)</p> <p>Round 2: towards the beginning of the second year of implementation (November 2022)</p>	<p>In each case study, we will interview personnel in case study settings (APST specialists, other members of staff at the school, people from local agencies who work with APST specialists), review documents and data</p>	<p>To understand more about implementation and operation of APST, perception of outcomes, and perception of partnership/collaborative working</p>

	(DS9). <sup>41</sup> Case studies will comprise interviews, observations, and documentation review	Round 3: towards the end of implementation (April 2023)	provided by the school, and consider pupil voice  We will use purposive sampling to select the schools involved in case studies (considering range of experiences with APST, geographical location, size, type of school)	
<b>Surveys</b>	Online surveys of SLT and APST specialists at the 22 participating AP schools (DS1)	Round 1: in early implementation (April -May 2022)  Round 2: towards the beginning of the second year of implementation (October 2022)  Round 3: towards the end of implementation (-April-May 2023)	We will invite all SLT leads from 22 schools to take part in the SLT survey  We will invite all APST project coordinators and specialists from all 22 participating AP schools to take part in the specialist survey	To understand more about implementation and operation of APST, perception of outcomes, and perception of partnership/collaborative working
<b>Observations</b>	We will observe relevant programme-level meetings about APST programme delivery (DS4)	Throughout the evaluation	This includes the quarterly Programme Board meetings, in person meetings of APST specialists/SLT, and other meetings to which the evaluators	To understand more about implementation and operation of APST, perception of outcomes, and perception of partnership/collaborative working

---

<sup>41</sup> In Round 1, we altered our plan slightly given the early stage of implementation. We carried out case studies of 3 APST schools drawing on virtual interviews with APST specialists and SLT.

			are invited and which appear to be relevant	
<b>Documentation and data review</b>	We will review data provided by the DfE delivery team on the metrics collected from intervention AP schools (DS5). We will also review documentation and data about each intervention AP school and about the wider policy context (DS6) at the start and end of the evaluation	Throughout the evaluation	DfE metrics & other documentation shared by DfE Policy documents relevant to APST Other documents relevant to schools	To understand more about implementation and operation of APST, perception of outcomes, and perception of partnership/collaborative working

## Interviews

### *Interviews with SLT in participating AP schools*

We propose to conduct **interviews with the SLT lead (DS2)** at each of the 22 participating AP schools at two points in time. The first round of interviews was conducted between November 2021 and February 2022). The second round of interviews is planned for June and July 2023).

The first round of interviews focused on understanding the context of the school, emerging barriers and facilitators and progress with roll out. They were also an opportunity for the Consortium to address schools' questions about the administration of the SDQ, and a chance to get to know the school better (in order to facilitate future surveys and case studies). Interviews were conducted with 16 SLT across 16 schools.

The second round of interview will be a chance to reflect on the two years of implementation and identify learning points.

### *Interviews with SLT in comparison AP schools*

We propose to conduct **interviews with an SLT representative in up to 10 of the 21 comparison schools (DS3)** at the close of the evaluation (in June-July 2023). Understanding that AP schools are diverse and may have a range of ongoing interventions that may be similar in scope and focus to APST, we intend to use these interviews to explore any other interventions as well as general key challenges and facilitators faced by comparison schools during the evaluation period.

We will use purposive sampling to ensure that the comparison AP schools selected for interview represent a diversity of AP schools in their geographic region and other key characteristics (including geographical location, type of school, size of school). We will identify the most relevant person for interview within the school in liaison with the DfE delivery lead.

We propose to conduct interviews at just one time point, at the end of the evaluation, for the following reasons: we think this will be sufficient to gather the information needed about business as usual in these AP schools over the evaluation period; interviewing representatives from comparison AP schools earlier in the evaluation carries a small risk of contamination (i.e. these schools might change their approach after engaging in conversations about multi-agency support for pupils); we are keen to minimise burden on the comparison AP schools and use evaluation resources carefully, and we do not feel that an additional round of interviews, earlier in the evaluation, would return enough benefit to the evaluation to justify the cost and burden.

### *Interviews with strategic personnel*

We propose **three rounds of interviews with strategic personnel (DS7)** (from the DfE delivery team, other DfE teams, the YEF, other departments that are involved in the APST pilot, and other relevant stakeholders). The first round of interviews was conducted in April 2022; the second round were conducted in October 2022; a third round is planned for March 2023. In total this will result in around 18 interviews across the evaluation. These will coincide with

data collection from the intervention AP schools (via the surveys and interviews) and aim to capture cross-programme issues and lessons. These interviews also provide opportunities for the research team to interact with key personnel involved in APST delivery and are therefore part of the formative nature of the evaluation.

#### *Interviews with stakeholders in broader AP context*

At the end of the evaluation, we propose **interviews with stakeholders involved in the wider AP context (DS8)** who are not directly involved in the design and delivery of the APST. These aim to gather outside perspectives, place the APST in the wider policy and practice context, and inform assessments about the potential for wider roll-out and generalisability of the evaluation findings.

### **Case studies**

We propose to conduct a total of 9 case studies (in 3 rounds of 3 case studies) each focusing on a participating AP school. The first round of case studies was conducted in June 2022; the second round is commencing at the time of writing in November 2022. The third round is planned for April 2023.

The first round of case studies involved up to five virtual interviews with APST specialists and one interview with a SLT lead at each of three case study AP schools (this was a change compared to the approach included in the initial evaluation plan, which had included in-person visits: see Box 1 below).

The second and third round of case studies will involve a two-day researcher visit during which we will: conduct interviews with APST specialists, other AP staff and relevant local partners and agencies; review relevant documentation and data, and observe meetings of the APST specialist taskforce where practicable. In addition, we will work with staff at the school to consider how, if at all, pupil voice might be captured in the school and will be led by their expertise and experience. Examples of how this might work, gathered from individual schools consulted during co-design, have included RAND observing existing pupil forums or RAND asking staff to hold focus groups with pupils in which RAND observes and gathers overall impressions and feedback.

The sampling for the Round 2 case studies takes a different approach compared to that included in the initial evaluation plan. The initial plan was to approach different AP schools in each round. In discussion with YEF and DfE, it was instead decided to approach the same three schools as Round 1, rather than a new set of schools, because the APST programme is complex and unique in each school and useful, detailed data could be gathered by following up with the same schools. Two of the three schools from Round 1 agreed to take part in Round 2, meaning one new school was chosen.

The consortium will decide whether Round 3 continues to focus on the same schools or will focus on a new set of schools in consultation with DfE and YEF and reflecting on learning from round 2.



We have and will continue to use purposeful sampling to select schools, ensuring a balance across different regions, schools with different experiences of implementation, and of different sizes.

### **Online surveys**

We propose an **online survey (DS1)** in all 22 intervention AP schools at three time points. The first survey was administered in April-May 2022; the second in October 2022 and a third is planned for April-May 2023). In each of the 22 participating AP schools the following will be invited to participate:

- The headteacher or SLT lead.
- The APST project coordinator.
- APST specialists.

Two tailored survey instruments are used: one for the headteacher or SLT lead and one for APST project coordinators and specialists.

Contact details for respondents were obtained through initial conversations with the 22 participating AP schools and liaison with the DfE delivery leads. Surveys include a mix of 'perennial' questions to consider change over time while also allowing for variation in questions, in order to explore different issues and emerging themes.

The first survey focused on experiences of roll-out and implementation, while the second survey followed up on these issues and explored perceptions on how APST may contribute to outcomes.

We will also use the surveys with the headteacher/SLT lead to collect information to inform the cost evaluation (please see Cost evaluation).

### **Observations**

We will **observe relevant meetings (DS4) about APST programme delivery** in order to gather more information about emerging themes, challenges and learning, and to understand the governance of the programme. These include the quarterly cross-governance group meetings (with wider departmental attendance) and ad hoc meetings with the 22 participating AP schools.

We also observed nine Communities of Practice (CoP) events carried out between January and June 2022. This was not foreseen as an evaluation activity in the initial evaluation plan, as CoP emerged during early implementation. It was decided to trial observations, in order to build our understanding of what was happening in the APST specialist teams. At the first formative feedback point, the Consortium, DfE and YEF reflected on the learning as part of this activity and determined that observing *online* CoP meetings was of limited usefulness.

We determined, going forwards, to attend and observe any *in-person* meetings that took place.

### **Data and documentation review**

We will **review data and metrics provided by the DfE (DS5)**. We have liaised closely with the DfE delivery team to lend our expertise to decisions regarding which metrics to collect and the format of collection. We intend to review and analyse relevant metrics throughout the evaluation, including (where possible) to consider treatment in the case of non-compliance (see Impact). We also intend to monitor and observe activity on the online Collaboration Teams hub (which is maintained by the DfE as a forum for the 22 participating AP schools to collaborate and communicate) on a monthly basis.

We propose to conduct **desk research on practice and policy (DS6)**. At the beginning of the intervention, we reviewed the applications and delivery plans submitted by the 22 participating AP schools to the DfE to ensure a good understanding of the context and other activities. We propose to review the policy context at the beginning and end of the evaluation in order to understand national priorities, strategies, programmes, etc. that are relevant to AP as well as to the outcomes the APSTs aim to achieve.

### **Analysis**

We use an evaluation grid, which maps out evaluation questions, the ToC components and cross-cutting themes, in order to develop data collection tools with questions and prompts closely linked to the questions. This means that data collection is aligned with the ToC and the research questions.

Collecting data in this way facilitates the analysis of data in a way which is also structured according to the evaluation questions and informed by the theory of change.

Data from interviews, case studies, open-text survey responses observations and document reviews are coded deductively (to identify data relevant to each question and to each theory of change construct, such as activities and outcomes), and inductively (to enable identification of unanticipated themes and findings). All coded data, from each data source, are entered into an analysis matrix, which categorises data according to our evaluation questions, aligned with the Theory of Change. This enables us to triangulate and synthesise all data, from across data sources, that are relevant to an evaluation question. Our evaluation questions are designed to enable us to test and refine the intervention logic, by assessing the evidence relating to each theory of change construct (inputs, activities, outputs and outcomes). Several of our evaluation questions ask ‘how’ activities make a difference or contribute to outcomes, which will enable us to elucidate mechanisms of change underpinning the intervention.

Data from closed survey questions are analysed through descriptive statistical analysis, which can also be fed into the analysis matrix.

## Formative feedback and reporting

We report emerging findings from each data collection round in formative feedback presentations to the DfE and the 22 participating AP schools delivering APST at three time points (September 2022, February 2023, summer 2023).

We will synthesise and report findings from all data collected in the process and cost evaluations in a report to be delivered in November 2023 (unpublished).

### Box 1: Summary of changes in the process evaluation since the evaluation plan on 6 December 2021

As reflected in this section of the SAP, a few changes have been made to the process evaluation compared to the initial evaluation plan dated 3 December 2021 (unpublished). These changes were made in response to adaptations and changes in delivery of APST by the DfE in school and ongoing learning from the evaluation.

The changes are:

- **We have refined our process evaluation research questions** in order to consolidate and clarify the themes explored and to ensure clarity, comprehensiveness and relevance. The refined research questions are included above.
- **We adapted the methodology for Round 1 of the case studies (DS9): to focus on the APST specialists within 3 case study schools rather than the wider context and to include observation of CoP events attended by specialists.** We made this decision based on the learning so far that many APST teams remained in the early stages of establishment in June 2022 and speaking to the wider school was less valuable than initially anticipated. CoP events only started after our initial evaluation plan, and we wanted to ensure that we attended these to capture learning.
- **We may focus our 9 case studies (DS9) on a smaller number of schools:** we initially intended to carry out case studies on 9 separate schools but have since realised that in order to achieve the depth of learning these were intended to bring, we may need to focus more closely on a smaller number of schools and following them more closely over time as they deliver APST.
- **We have learnt more about the meetings that are ongoing in the APST programme management and updated our plan for observations (DS4) accordingly:** we attend and observe the quarterly Programme Board meeting, monthly YEF-DfE-Evaluation catch-ups (included in our initial evaluation plan). We have also ensured that we attend and observe ad hoc meetings of APST leads and project coordinators when scheduled (not included in the initial evaluation plan)
- **We have learnt more about the data and documentation that is collected and made available by the DfE and adapted our plan accordingly (DS5 and 6):** we

intend to review the metrics and any further documentation that the DfE makes available to us.

## Cost evaluation

The cost evaluation is the third component of the independent evaluation funded by the YEF. It is led by RAND Europe. We propose a cost evaluation that will describe the costs associated with delivery of APST at both the level of delivery and at the school level.

### Research questions

In line with guidance from YEF around cost evaluation, we propose one research question for the cost evaluation.

**Table 14: Evaluation questions for the cost evaluation (EQ22)**

	Cost evaluation
EQ22	What are the costs of delivering APST in schools?

### Research methods

We propose to collect data on costs from two sources: the DfE and the 22 APST schools.

#### *Data from DfE on costs incurred by schools*

We understand that the DfE gathers and records the costs incurred by the 22 participating AP schools when delivering APST, recording these as “Project leadership/management” costs, “Specialist” costs, “Equipment” costs, “Capital building” costs, and “Other” costs. Using these data gathered by DfE will be the main way in which we collect cost data in this evaluation.

#### *Data from DfE on costs incurred by DfE*

The YEF cost guidance requests that costs incurred by the intervention provider should be included in a cost evaluation. The DfE have confirmed that they are not able to provide data on the costs incurred by DfE in administering and supporting the APST programme.<sup>42</sup>

We will explore the costs involved in delivering APST for the DfE qualitatively through our final round of strategic personnel interviews (DS7).

#### *Data from APST schools*

We understand that there may be some costs incurred by the 22 participating AP schools that are not directly reimbursed by the fund from the DfE. We aim to understand if these exist,

---

<sup>42</sup> While the DfE maintains a record of administrative costs of the DfE team working across APST, SAFE and other initiatives, their record does not break down how much time is dedicated to APST specifically.

what they cover, and estimate the costs. We intend to use process evaluation data collection methods to explore this: including round 2 and 3 of the survey to SLT (DS1), the final round of interviews with AP schools (DS2) and case studies (DS9).

## Categorising cost

We propose the below cost categories for costs (see Table 16):

- Staff costs
- Buildings and facilities costs
- Materials and equipment
- Miscellaneous

In proposing the categories outlined below in Table 16, we considered carefully the categories outlined in the YEF guidance<sup>43</sup> and consulted with the DfE regarding the ways in which they categorise the costs incurred by schools. The categories that the DfE uses to currently record costs are displayed in Table 16 and align with how we will categorise costs in our evaluation.

There are two categories suggested in the YEF guidance that we do not intend to use:

- ‘Programme procurement costs’<sup>44</sup>: these are not collected by DfE, and we understand that these are likely to be less relevant for APST (which is not a designed programme being implemented).
- ‘Incentives for taking part’: we understand that these are not used in APST. Incentives used as part of the evaluation (for comparison schools) are not included.

**Table 16: Cost evaluation categories proposed and how these relate to data provided by DfE**

The cost categories we will use in the evaluation (taken from YEF cost guidance)	Relevant DfE data cost categories
Staff costs	Project leadership/management Specialists <sup>45</sup>

<sup>43</sup> These are: staff costs, programme procurement costs, buildings and facilities, materials and equipment, and incentives for taking part. See YEF (2022) Cost reporting guidance. <https://youthendowmentfund.org.uk/wp-content/uploads/2022/01/21.-YEF-Cost-reporting-guidance.pdf>

<sup>44</sup> This is a category requested in the YEF Cost reporting guidance.

<sup>45</sup> The “Specialist” category used by the DfE mostly includes staff costs but does sometimes include service-based costs, i.e., administration fees. The DfE will provide us with the schools who have invoiced for a service so that we can separate staff costs from administration fees.

Buildings and facilities costs	Capital building
Materials and equipment	Equipment
Miscellaneous	Other

Each of the four cost categories outlined above will be broken down further into types of cost:

- **Pre-requisite costs:** any costs incurred and invoiced by schools in cost categories “Project Oversight” and “Overheads” before November 2021<sup>46</sup>
- **Set-up costs:** any costs incurred or invoiced by schools in all cost categories between November 2021 and January 2022.
- **Recurring costs:** any costs incurred or invoiced by schools in all cost categories after January 2022.

We have operationalised these cost types using YEF guidance and our knowledge of the APST delivery so far.

## Data analysis

By collecting these costs, we aim to construct a comprehensive picture of the cost incurred by the 22 participating AP schools of administering the APST programme. These will be combined to give an overall cost, which can further be broken down per school. We will endeavour to break these down per pupil using official on-roll data but note that this may be less accurate and useful given the fluid nature of the AP pupil population.

We propose to provide three final costs:

- The total cost of the intervention (all 22 schools, all pupils KS3 and KS4) over the two academic years of delivery.
- The average cost of the intervention per school over the two academic years of delivery.
- The average cost per pupil for the intervention over the two academic years of delivery.

---

<sup>46</sup> The DfE have advised to omit any “Specialist” costs invoiced before November 2021 in this pre-requisite category. Many began work in advance of November as they were assured they would receive grant funding.

As per YEF guidance, within these categories, we will indicate how the total costs break down to pre-requisite, set-up and recurring costs.

We will follow YEF's cost guidance when calculating the full cost of delivery, including adjusting costs to constant prices using GDP deflators. YEF recommends using the year that the delivery begins as the base year, in this instance, 2021.

We will account for uncertainty in the costings provided and document all assumptions made in the final calculations. We do not intend to complete sensitivity analyses.

Where estimates of staff wages or resources cannot be retrieved from schools, market estimates will be used. We will consider and document uncertainties and assumptions and present these alongside our calculations.

We understand that the cost evaluation should consider benefits outside the programme: for example, if equipment or materials are purchased that will be used outside the programme, only a percentage of that equipment's cost should be attributed to the intervention. However, the DfE confirmed that schools would only be using the equipment specifically for APST programme activities and so there would be no benefits outside the programme to consider.

Since delivery of APST relies heavily on the use of practitioners, and since schools may be using practitioners for other purposes, our cost calculations will take into account these overlapping uses so as not to inflate incorrectly the costs of running the APST programme. However, the DfE confirmed that schools already account for practitioners used for other purposes and the invoicing will only include those hours worked directly for the APST programme.

#### **Box 2: Changes in the cost evaluation since the evaluation plan on 6 December 2021**

We suggest a few revisions to our strategy for the cost evaluation since the evaluation plan on 6 December 2021. These are in response to reviewing the updated YEF cost guidance and following consultation with DfE about data available.

The changes are:

- We no longer propose to evaluate the cost of business-as-usual in the 21 comparison AP schools, as this is not recommended by the YEF cost evaluation guidance.
- We no longer aim to identify the value of at least one outcome using unit costs and compare our outcomes to business-as-usual in the 21 comparison AP schools, as this is not recommended by the YEF cost evaluation guidance.
- We no longer intend to complete formal sensitivity analyses to account for heterogeneity in costs, as this is not recommended by the YEF cost evaluation guidance. We will provide sufficient detail around any assumptions and variations of costs (including reasons for variations of costs) across schools.



## Ethics

The evaluation team have expertise working with vulnerable groups and a good understanding of the ethical and practical considerations involved. Our approach to ethics includes considering the value of including CYP as active participants and working to ensure that their participation is meaningful and to design appropriate approaches to data collection and consent for them and all stakeholders.

We recognise a number of ethical issues to be considered and addressed in our evaluation, including the data that we intend to collect from pupils as part of the impact evaluation, incorporating pupil voice into the evaluation, and ensuring researcher wellbeing during fieldwork visits.

We have received full ethical approval from the RAND Europe Human Subjects Protection Committee and from University of Westminster Business School College's Research and Knowledge Exchange Ethics Committee for all aspects of the study.

### **Ethical considerations related to the impact evaluation**

We have considered carefully how we intend to gather data on social and emotional wellbeing (through the Strengths and Difficulties Questionnaire) from pupils attending AP schools.

We have carefully considered how to ensure that parents and pupils are informed of the use and gathering of the data and know their rights around withdrawing from the evaluation, taking into account UK GDPR (as considered below) and ethical conduct.

We have developed accurate yet accessible documentation for pupils and parents that include information on the APST, the evaluation, and their role within it. We have asked schools to disseminate this to pupils and parents before administering the baseline and endline SDQs. This includes clear information about pupils' and parents' rights around participating and providing data. We have ensured that this documentation makes it clear who is the point of contact for any questions and concerns about data being used.

We have provided schools with guidelines around collecting SDQ data, including guidance about how to introduce this to pupils and parents and the common concerns or questions that may arise.

### **Ethical considerations related to the process and cost evaluations.**

We have also considered ethical conduct across our process and cost evaluations as part of our ethical review process. To draw on pupil voice in the case studies as part of the process evaluation, we will use mechanisms in place at AP schools to do so. We will work closely with AP schools to ascertain the most appropriate method of data collection (if any).

We propose for researchers to visit AP schools in order to carry out case studies. This means that there is the need to closely consider the safety of both researchers and research subjects within this context. We will draw upon RAND Europe's relevant policies for researchers working outside of the premises that are clearly explained and publicised. We intend to use researchers who are experienced in conducting research in schools or other settings and to ensure that there are a pair of researchers working in each school. We will ensure researchers are aware of the safeguarding policies in each school and follow these, in line with YEF Safeguarding guidance. All members of RE staff hold a basic DBS clearance and we will arrange enhanced DBS clearance for those undergoing fieldwork. We will establish a point of contact at the relevant school and ensure that researchers follow all processes required by the school before entry. Finally, we will ensure that decisions about in-person visits are made taking into account the local and national COVID-19 situations, and will ensure that data collection can take place remotely if required by the situation.

## Data protection

In this project, we identify and use personal and special category data in the process and impact evaluations.

**Table 15: What personal data is gathered in different strands of the evaluation**

	Personal data	Special category data
Process evaluation	Interviewee's name, email address	NA
Cost evaluation	NA	NA
Impact evaluation	Pupil name Pupil Date of Birth (DoB) Pupil UPN	Pupils' responses to SDQ (indicating social and emotional wellbeing)  Pupils' data held on NPD regarding gender, race/ethnicity, looked after status, SEN, whether or not they receive free school meals

## Data roles, legal bases and approaches related to the impact evaluation

RAND Europe and the DfE are joint controllers of the personal data (pupil name, DoB, UPN, SDQ scores) and special category data (data on health, gender, ethnicity, SEN, LAC, FSM) gathered for the impact evaluation. The University of Westminster and FFT Education Datalab are data processors.

The stipulated roles and responsibilities of RAND Europe and the DfE as joint controllers will include:

- RAND Europe will be the point of contact for all data subjects and administer all requests for withdrawal.
- **RAND Europe will be responsible for collecting pupil administrative data (name, DoB, UPN) and the SDQ tests from the AP schools**, using a secure courier service. RE will provide the AP schools with guidance on how to collate this information and administer SDQs, how to share the relevant ethical and data protection material, how to share the information with RE, and how to ensure withdrawals are included. RE has set up DSAs between the AP schools and RE to allow this data to be shared.
- RE will securely store pupil administrative data (name, DoB, UPN) and the SDQ tests.
- **RE will securely transfer pupil administrative data to DfE for linkage.** RE will liaise with DfE to establish a DSA to cover this sharing of data.
- DfE will be responsible for linking the dataset from RAND to other datasets in the NPD/ILR and liaising with both processors (FFT, University of Westminster) around accessing this in the SRS
- DfE will be responsible for transferring the required datasets to YEF at the end of the evaluation (for the YEF archive).
- After the evaluation ends, the YEF will become the data controller and deidentified data will be stored in the YEF archive for use by future researchers.<sup>47</sup>

Further information on the data flows involved in the personal and special category data gathered through the impact evaluation may be found in the Impact Evaluation Privacy Notice and Data Protection Impact Assessment.

The legal basis for RAND Europe to process students' personal data is legitimate interests, detailed in Article 6(1)(f) of the UK GDPR. The legal bases for processing students' special category data are for reasons of substantial public interest and because it is necessary for archiving purposes in the public interest, scientific or historical research purposes as detailed in Article 9(2)(g) & (j) respectively of the UK GDPR. To ensure that all processing is fair and

---

<sup>47</sup> You can find more information about the YEF archive on the YEF's website:  
<https://res.cloudinary.com/yef/images/v1625734531/cdn/YEF-Data-Guidance-Participants/YEF-Data-Guidance-Participants.pdf>.

lawful, RAND Europe have also completed a Legitimate Interest Assessment and a Data Protection Impact Assessment and completed an application to the RAND internal review board for ethical approval. RAND Europe will process only what is required to meet these legal bases and will ensure security and safeguards are in place to protect the information.

The legal basis for the DfE to process students' personal data is where it is necessary for the performance of a task carried out in the public interest as set out in Article 6(1)(e) of the UK GDPR. The legal basis for processing special category data is for reasons of substantial public interest as detailed in Article 9(2)(g) of the UK GDPR and will process any criminal offence data using research as its condition for proceeding in Schedule 1 of the DPA 2018. The DfE also conducts a Data Protection Impact Assessment, and its Data Sharing Approval Panel ensures any sharing of data through the ONS's Secure Research Service meets ONS's 'Five Safes data protection framework' to make sure that the people, projects, settings, data and outputs are safe.

RE has provided students and their parents with fair processing privacy notices that explain the use, storage and secure handling of the data. This makes it clear that during or after the intervention, participants may contact RAND Europe and/or the DfE to request deletion of their personal data. Moreover, students/parents are told of their rights to access, correct or erase their data, to object to the processing of their data, or to request that the data is restricted or withdraw their consent (they are also provided with a withdrawal form for this purpose). Parents/students are provided with the RAND DPO email address in order to contact the evaluation team about exercising these rights.

The RAND Europe evaluation team have appropriate security measures in place to keep personal data secure and to prevent any unauthorised access to or use of it, in line with UK GDPR. Data transferred between parties will be encrypted or use secure file transfer protocols. The Evaluation Team will collect and store all evaluation data in accordance with the Data Protection Act (2018) and GDPR (General Data Protection and Regulation) requirements. Evaluation data will be stored on secure servers. Analysis will be conducted in the ONS' SRS – a secure environment that can only be accessed by accredited researchers.

The RAND team have measures in place to mitigate the risks involved in collecting and accessing special data. We will pseudonymise all data to prevent identification through linking several datasets. Data identifiers will be stored securely in a separate folder to the linked data spreadsheets and will only be accessed if the data subject (or in this case their parent) requests to modify or erase the data. No individual school or child will be identified in any report arising from the research.

With the support of their dedicated DPO team, RE has established and follows processes, procedures and logs around the transfer of data, the withdrawal of pupils from the evaluation, the deletion of data, and any data breaches.

At the end of the project, the YEF will become 'controllers' of the data collected throughout the project. Data will be securely transferred into their archive.

Before the data goes into the YEF archive, the Department for Education will take out names and other personal details like addresses. This means that no one who looks at the information in the YEF archive will know who it relates to. In the future, people can ask to use the YEF archive to do more studies to find out whether this project has helped young people. Only researchers who are approved by the YEF will be able to look at the archive. The police can't use the information in the YEF archive.

### **Data roles, legal bases and approaches related to the process evaluation**

In our process evaluation, we intend to gather personal data of interviewees: including those at the DfE, strategic partners, AP schools, and local areas who are invited for interview. We do not intend to gather special category data. We do not intend to gather personal data from students.

RAND Europe are the sole data controllers of the process evaluation data. There are no data processors. The data will not be transferred to processors or other bodies. All personal data will be stored in secure, on-shore servers in restricted-access folders. The personal data will be kept separately from the interview data and will be pseudo-anonymised. This data will not be shared with, or accessed by, anyone not directly involved with the study.

All personal data collected by the project will be stored in secure, on-shore servers in restricted-access folders. The personal data will be kept separately from the interview data and will be pseudo-anonymised. This data will not be shared with, or accessed by, anyone not directly involved with the study. These measures will prevent any unauthorised access to or use of interviewees' personal data in accordance with Data Protection Act (2018) and UK GDPR requirements. No data will be saved on servers or shared with processors outside the UK. RAND Europe will securely delete all data held on its secure server six months after the end of the project.

### **Data security**

In terms of data storage, RAND Europe has implemented a company wide Information Security Management System (ISMS). RAND Europe is accredited for ISO 27001 certification and Cyber Essentials Plus. We have a senior management team that supports the continuous review and improvement of the company ISMS. Key controls RAND Europe has implemented include:

- An Information Security Risk Assessment Process that assesses the business harm likely to result from a security failure and the realistic likelihood of such a failure

occurring in the light of prevailing threats and vulnerabilities, and controls currently implemented.

- An Information Classification and Handling Policy including compliance with regulations under the Data Protection Act to protect client, partner, supplier, our own and personal employee information which is not in the public domain.
- A Business Continuity Plan to counteract interruptions to business activities and to protect critical business processes from the effects of major failures or disasters.
- Defined security-controlled perimeters and access to controlled offices and facilities to prevent unauthorised access, damage and interference to business premises and information and data that might be held there.
- Mandatory Information Security awareness guidance for all company employees.
- Background screening of all company employees.

SDQ data will be held on a server located in RAND Europe's Cambridge, UK office, only.

NPD data supplied by the Department for Education will be held in the Secure Research Service provided by the Office for National Statistics. Access to data is governed in line with the principles of the five safes framework.<sup>48</sup>

## Stakeholders and interests

In this section, we define the roles and responsibilities held by colleagues within the developer and delivery team and the evaluation team. We consider any other stakeholder involvement in the design, conduct and analysis of the evaluation. We declare the sources of funding and support and consider any other potential interests.

### Developer and delivery team

The developer and delivery team is the DfE. Within this team, we identify the following stakeholders:

- The Preventing Serious Youth Violence Unit.
- The Serious violence Strategy (HO crime plan) team, the SAFE taskforces, Serious violence duty, Serious violence governance, Spending Review, Business of State

---

<sup>48</sup>

<https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/aboutthesecureresearchservice>

- The delivery of the AP taskforces team: engaging with local partners and AP schools, solving challenging, promoting AP networking and reporting into project management team
- The DfE policy team, delivery team, PMO team.

We also identify the strategic partners who support the Department for Education in delivering APST. These include:

- Royal College of Speech and Language Therapists (provide support to the speech and language therapists)
- NHS (provide support to the speech and language therapists and mental health professionals)
- Department for Work and Pensions (provide support to post-16 transition coaches)
- National Youth Agency (provide support to youth workers)
- Association of Educational Psychologists (provide support to educational psychologists)
- What Works Children and Social Care (provide support to social workers)
- Youth Justice Board (provide support to youth justice workers).<sup>49</sup>

## The evaluation team

**Evaluation team leadership and oversight:** The evaluation is led by **Dr Emma Disley**, Director of the Home Affairs and Social Policy team at RAND Europe and Principal Investigator of this study. The evaluation is managed by **Natalie Picken**, Senior Analyst at RAND Europe.

**Impact evaluation:** The impact evaluation strand is co-led by **Dave Thomson** of the FFT and **Professor Richard Dorsett** of the University of Westminster. RE contributes primary data collection and coordination, with a team of researchers overseen by **Natalie Picken** with the support of **Dr Emma Disley** (PI).

**Process evaluation:** The process evaluation is led by **Natalie Picken**, working alongside a team of experienced researchers. These include **Dr Ana FitzSimons** (who has extensive experience conducting evaluations, including theory of change evaluations of interventions to support

---

<sup>49</sup> Information provided by the Department for Education to the evaluation team.

vulnerable children, young people and families), **Judith Ajebon, Sarah Angell, Iris Leussink, Tamara Strabel and Dr Kankan Zhang.**

**Cost evaluation:** The cost evaluation is led by **Elle Wadsworth.**

### **Evaluation funders**

The evaluation is funded by the **Youth Endowment Fund (YEF)**. YEF is a fund established by the children's charity Impetus that holds an endowment from the Home Office. The YEF aim to prevent CYP becoming involved in violence, by finding out 'what works' and putting this knowledge into practice.



## Risks

Table 16 below sets out the main risks we have anticipated, our original planned mitigation strategies along with updates made since December 2021, and our assessment of the likelihood of each risk following the mitigation measures. The principal investigator and project manager will monitor potential risks to the project throughout the study and report on these regularly (along with mitigation strategies) to the YEF.

**Table 16: Anticipated risks, mitigation strategies and risk likelihood following mitigation (18/11/22)**

Risk	Mitigation	Likelihood following mitigation (H, M, L) – <i>updated on 12th October</i>
APST begins work before SDQ baselining is possible	This was a risk identified in September 2021, when conversations with DfE around data roles and school recruitment were at risk of delaying baselining.	All schools have baselined. <b>This risk has been resolved.</b>
Delay in ethical review hinders data collection	We were in the process of submitting an ethical application to the RAND internal board in September 2021, while data collection was planned, and faced delays in approval while updating this based on changes to the evaluation design. We now have ethical approval and data collection was not hindered.	We have ethical approval and data collection was not delayed. <b>This risk has been resolved.</b>
Stakeholders involved in the implementation are too busy to engage with the evaluation, leading to attrition	We will carefully consider data requests, ensuring they are not burdensome (e.g., short and targeted survey tools) and do not come at busy periods in the academic year, through consulting with schools and DfE and allowing plenty of time for planning.	M

<p>Risks associated with data protection (for example, accidental disclosures or breaches)</p>	<p>RE, UoW and FFT have extensive experience of complying with GDPR. RE's in-house data protection team provide guidance and advice. We have data sharing agreements in place with all 43 schools and the DfE and continue to work with DfE as joint controllers and to ensure the safe transfer of data. The research team has developed relevant procedures and processes around data transfer, withdrawal and breaches, which are followed and documented.</p> <p>We have updated this from L to M to reflect the increased risks of inadvertent data breach associated with the 43 schools sharing personal and special category data with us.</p>	<p>M</p>
<p>Risks associated with conducting research with vulnerable CYP to capture pupil voice (including challenges in organising and participation, risks associated with researchers visiting educational institutions)</p>	<p>We can draw on extensive experience of conducting research with vulnerable population groups. Our approach will be to liaise closely with the schools involved in the 9 case studies to tap into their existing structures for consulting pupils. We have budgeted sufficient time, based on our experience, to work with schools to plan these consultations. We have designed a clear informed consent process in liaison with our ethical review board and drawing on our experiences of conducting similar research. We will ensure that all researchers are aware of their safeguarding responsibilities and of procedures.</p>	<p>M</p>
<p>Ongoing COVID-19 pandemic</p>	<p>We have planned for meetings and data collection to take place virtually – but budgeted to allow these to shift to real-life depending on the COVID-19 situation. The flexible and agile approach we propose will allow us to adapt accordingly. We previously categorised this risk as High and now consider that this is Low, thanks to a lessening of the</p>	<p>L</p>

	pandemic and ongoing experience with conducting research during the pandemic.	
Impact analysis underpowered	We conducted this preliminary analysis to consider power and effect sizes and suggest outcomes and data analysis methods accordingly. We will incorporate control variables to increase power and will explore how power varies with the number of comparator schools.	L
Missing SDQ outcome data (as a result of schools withdrawing, surveys being incomplete etc.)	We have worked closely with schools and DfE to ensure administration of SDQ is as complete as possible and fed back issues to schools. We are aware of missing data from SDQs from Collections 1 and 2: however, we have seen an improvement in quality over time and we have considered this as part of the preliminary analysis and selected some outcomes from administrative data (which are unaffected by this issue). We closely monitor engagement of comparison schools and minimise burden involved in the SDQ collection by learning from feedback.	M
Delays or difficulties in availability of administrative data on SRS	We have experienced some delays in accessing data shared by DfE thanks to SRS capacity. We work closely with DfE and ONS to overcome issues. We budget in some additional time to allow for delays without affecting overall delivery. We consider the risk to be high but the impact to be low.	M
Unexpected results in the impact evaluation	The impact evaluation may produce unexpected results which require detailed investigation to explain. We consider the risk to be low as the methods we use are standard and the code we plan to use has been written and tested, however some risk will always remain in a project of this complexity.	L
Significant variation in implementation of APST in	We will use the compliance analysis to consider how far lack of implementation is an	H

the 22 participating AP schools means that an APST programme is difficult to describe and creates challenges for impact evaluation approach which looks at the effect of APST on pupils in all 22 schools	issue. We will use the data collection methods in IPE to understand and document this variation and remain flexible in how we use these to respond to the situation. For example, we have amended the approach to the case studies – revisiting the same schools in round 1 and 2 – in order to collect more detailed information from a smaller number of schools.	
Delays in implementation (for example, delays in recruiting specialists) means the evaluation period only captures a short period of delivery as intended	We will use the IPE to understand the process of implementation and contextualise the findings. Our compliance analysis will allow us to consider, dependent on data available from DfE, how far lack of implementation is an issue.	M
Data provided by DfE is not able to be delivered, is delivered late, or does not include the information expected. We anticipate this may affect the impact analysis (treatment in the case of non-compliance), process (metrics and other data and documentation), cost.	We will continue to liaise closely with DfE, be flexible in timelines as needed, and be adaptable where possible. We have clearly outlined in this SAP where analysis will be challenging without DfE data.	M
The primary and secondary outcomes we propose are complex: this makes it hard to communicate the proposed approach and risks different expectations of evaluation outputs.	We have had multiple discussions with YEF and DfE in the process of designing this SAP and are confident that we will be able to explain clearly the outcomes selected and the reasons for their choice.	L

## Timeline

**Table 17: Timeline of the evaluation**

Date	Status (as of 3 April 2023)	Activity	Staff responsible/leading
<b>Process Evaluation</b>			
November 2021- June 2022	Completed	Round 1 Data collection (including Interviews with SLT at APST schools Interviews with strategic personnel Survey of APST specialists and SLT at APST schools Case studies of 3 selected APST schools)	RAND Europe
August 2022-September 2022	Completed	Round 1 – Formative feedback	RAND Europe
September 2022-December 2022	Completed	Round 2 Data collection (including Interviews with strategic personnel Survey of APST specialists and SLT at APST schools Case studies of 3 selected APST schools)	RAND Europe
February-March 2023	Completed	Round 2 – Formative feedback	RAND Europe
March 2023-July 2023	In progress	Round 3 – Data collection (including Interviews with SLT at APST schools Interviews with strategic personnel Survey of APST specialists and SLT at APST schools Case studies of 3 selected APST schools Interviews with comparison schools Interviews with wider stakeholders)	RAND Europe
November 2023	Not yet started	Round 3 – Formative feedback	RAND Europe
<b>Impact Evaluation</b>			

November 2021-autumn 2023	In progress	RE support and guidance to schools around administering SDQs, collecting SDQs, etc.	RAND Europe
November 2021-July 2023	In progress	APST and comparison schools administer baseline SDQs to all students joining their school in KS3 and 4 for the first time	Schools
February 2022-September 2023	In progress	APST and comparison schools administer endline SDQs to all students leaving their school in KS3 and 4 & for all those remaining in the school after academic year 2023/24	Schools
March 2022-April 2022	Completed	Collection 1 of SDQs gathered so far from APST and comparison schools	RAND Europe
July 2022	Completed	Collection 2 of SDQs gathered so far from APST and comparison schools	RAND Europe
July 2022- September 2022	Completed	Preliminary analysis to inform the statistical analysis plan	FFT and University of Westminster
January 2023	Not yet started	Collection 3 of SDQs gathered so far from APST and comparison schools	RAND Europe
Summer-Autumn 2023	Not yet started	Collection 4 of SDQs gathered so far from APST and comparison schools	RAND Europe
September 2023	Not yet started	Application to use NPD data for the impact evaluation	FFT
December 2023	Not yet started	RE shares dataset with DfE for linkage with NPD outcomes	RAND Europe

Spring 2024 (TBC)	Not yet started	DfE shares full dataset with FFT/UoW via SRS for analysis	DfE
Summer 2024 (TBC)	Not yet started	Impact evaluation analysis	FFT and University of Westminster
<b>Cost Evaluation</b>			
Summer 2022-Summer 2023	In progress	Collecting cost data from DfE	RAND Europe & DfE
September 2022-July 2023	In progress	Collecting cost data from APST and comparison schools	RAND Europe
July 2023-October 2023	Not yet started	Analysing cost data from APST and comparison schools	RAND Europe
<b>Reporting and deliverables</b>			
December 2021	Completed	Delivery of evaluation plan	RAND Europe, FFT, UoW
October 2022	Completed	Delivery of preliminary analysis and statistical analysis plan (D2) with revised ToC	RAND Europe, FFT, and University of Westminster
November 2023	Not yet started	Process and cost evaluation synthesis report	RAND Europe
December 2024	Not yet started	First draft of final impact report provided to YEF and DfE	RAND Europe

June 2025	Not yet started	Publication of final impact analysis, process and cost analysis	RAND Europe
December 2025	Not yet started	FFT and University of Westminster to upload data to YEF archive	FFT and University of Westminster



## **Appendix A- Alternative Provision Specialist Taskforces: Preliminary Analysis**

### **Background**

The Alternative Provision Specialist Taskforces (APST) is a new programme introduced in the 2021/22 academic year by the Department of Education. It aims to improve outcomes for young people aged 11-15 who access state-funded alternative provision (AP) schools. 22 AP schools were selected to participate from the 22 local authority areas with the highest levels of serious violence.

FFT Education Datalab and University of Westminster have undertaken this preliminary analysis in order to inform the study plan of a quasi-experimental impact evaluation of the programme. Briefly, we examine pre-existing trends between participating and non-participating AP schools using administrative data from the National Pupil Database (NPD) for years 2014 to 2021, some data limitations (described below) notwithstanding.

The purpose of the analysis is to:

- Calculate suitable outcome measures for the evaluation
- Set up data processing steps to create analytical datasets
- Test different methodological options
- Assess pre-existing trends in outcomes
- Calculate approximate effect sizes
- Help make decisions about the methodology we will pursue when undertaking the impact evaluation of the programme following the 2022/23 academic year.

We test three different approaches

- Difference in differences
- School-level propensity score matching followed by pupil-level regression
- Regression Discontinuity

Additional information regarding the balance achieved for outcomes may be found in Addendum 1.

Data tables supporting this Appendix A may be found in Addendum 2 (attached Excel file). These tables comprise:

- Appendix A, Table 1 Number of pupils included in difference-in-difference models for attainment and post-16 participation

- Appendix A, Table 2 Number of pupils included in difference-in-difference models for absence and re-integration
- Appendix A, Table 3 Pre-treatment trends in outcomes for 15 year olds
- Appendix A, Table 4 Pre-treatment trends in outcomes for 14 year olds
- Appendix A, Table 5 Number of pupils with multiple records in the 11-14 dataset
- Appendix A, Table 6 Pre-treatment trends in outcomes: Outcomes for 15 year olds
- Appendix A, Table 7 Estimated Treatment Effects (Placebo tests) for outcomes for 15 year olds
- Appendix A, Table 8 Pre-treatment trends in outcomes: Outcomes for 11-14 year olds
- Appendix A, Table 9 Estimated Treatment Effects (Placebo tests) for outcomes for 11-14 year olds
- Appendix A, Table 10 Covariates and factors used as controls
- Appendix A, Table 11 Key Characteristics of 11-14 year-old pupils in participating and non-participating schools
- Appendix A, Table 12 Key Characteristics of 15 year-old pupils in participating and non-participating schools
- Appendix A, Table 13 Percentage of pupils re-integrated into mainstream schools the following year by age

## **Descriptive Statistics**

Very limited published statistics exist for pupils who attend AP schools. These typically report on subsets of the population of pupils in AP. For instance, end of Key Stage 4 statistics only include pupils attending AP schools with a current or main registration in January of Year 11.

We created suitable outcome measures for the AP population as described in the sections on primary and secondary outcomes in the analysis plan above. In some cases (e.g., absence), this was relatively straightforward as we took a commonly used definition and applied it to the AP population. By contrast, the measure of re-integration is a measure that we have created specifically for this project.

Summary data for each outcome for participating and non-participating AP schools can be found in Appendix A, Table 3 (outcomes for 15 year olds) and Table 4 (outcomes for 11-14 year olds). This also includes the standard deviation for each measure (to be used in

calculating measures of effect size) and numbers of pupils. These can be found in Appendix A, Tables 1 and 2 respectively.

### *Difference-in-differences*

For each outcome, we fit DiD models using pre-treatment data. This has a threefold purpose:

- To test whether pre-treatment trends are parallel
- To run placebo tests (we should not expect to find any significant difference for a fake treatment year)
- To get a handle on the likely size(s) of confidence intervals with which treatment effects will be estimated (“power calculations”).

For each outcome, we calculate three sets of models comparing treated schools with:

1. All other AP schools
2. The 3 most similar schools calculated using available data when the project commenced<sup>50</sup>
3. AP schools in local authorities which were within a cut-off of 0.2 based on the measure of serious violence used by the DfE to select local authorities to participate in APST (see the selection mechanism section in the study plan for further details). This cut-off was determined by the investigation into the regression discontinuity design, hence we refer to it as a RDD hybrid. 18 local authorities meet this criterion.

Option 3 replaced a previous version in which we selected the largest AP schools in the list of the next 22 local authorities ranked by the serious violence measures used to select areas for participation in the programme (labelled next 22 SV areas). This was abandoned following completion of the analysis into the regression discontinuity, which suggested a more principled approach to deriving the cut-off in the serious violence score used to select local authorities to participate in the intervention.

Both 2) and 3) act as a pre-processing stage to make the set of comparison schools more similar to the treatment schools.

For each of 1-3 we calculate models with and without controls. This means there are six specifications for each outcome.

The models are fitted in Stata. Standard errors are clustered by school.

---

<sup>50</sup> This was done to identify a comparison set of schools for which SDQ data would be collected. Only 21 of the 22 participating schools had been recruited when this was carried out. We include all 22 participating schools in the treatment group but comparison schools for just 21.

**Table 18: Further information on calculation of outcomes**

Outcome(s)	Parallel trends	Placebo tests
Attainment, Post-16, NEET	Appendix A, Table 6	Appendix A, Table 7
Absence, re-integration	Appendix A, Table 8	Appendix A, Table 9

Note that we plan to use linear probability models (LPM) for binary outcomes (NEET, post-16 participation, re-integration) although we also tested the effect of using probit and logistic regression. There is evidence to suggest that the LPM may even be more appropriate when estimating treatment effects both in experiments (RCTs) and observational (QED) studies<sup>51</sup>. Furthermore, a LPM is consistent with the assumptions underpinning DiD, unlike non-linear approaches (such as probit and logit). A potential drawback is that LPM does not constrain predicted probabilities to lie between 0 and 1. However, this is more likely to be a problem when considering individual probabilities; our interest is in mean probabilities which are more robust to this.

The number of pupils included in the estimation datasets are reported in Tables 1 and 2. Note that the number of cases used in the specifications with controls for attainment and post-16 are approximately 10% smaller due to missing data (missing Key Stage 2 data).

The controls are listed in Appendix A, Table 10.

We used information from the placebo tests to calculate the minimum detectable effect size (MDES) for each outcome we plan to evaluate on the assumption that the post-treatment data behaves in a similar fashion. We calculate the minimum detectable effect (MDE) as a multiple,  $M$ , of the robust standard error,  $s$ . Here,  $M$  is calculated as the 0.025 upper tail critical value of the  $t$  distribution (reflecting the 95% two-tail significance level) plus the 0.8 upper tail critical value of the  $t$  distribution (reflecting 80% power). Although the precise value of  $M$  depends on the degrees of freedom,  $M \approx 2.8$ . The MDES is obtained by dividing the MDE by the standard deviation :

$$\text{MDES} = \frac{s * M}{\sigma}$$

## Parallel trends

### *Attainment, post-16 participation, engagement*

The results contained in the trends tab of Appendix A, Table 6 show the parameter estimates of treated\_school \* year for each outcome each year (relative to the base year of 2014). Of the 150 estimates shown (5 outcomes x 6 specifications x 5 years), 7 reach statistical

---

<sup>51</sup> <https://opa.hhs.gov/sites/default/files/2020-07/lpm-tabrief.pdf>

significance at the 5 % level – roughly the proportion that would be expected based on chance alone.

Visual inspection of the trends across years shows that in maths and engagement, there is an upward trend between 2016 and 2019 for treated schools across all specifications. If necessary, this trend can be incorporated into the diff-in-diff specification (see equation 2 in the study plan).

#### ***Re-integration, absence***

Table 8 shows that there are no particular concerns about the prior trends for these measures. None of the values of `treated_school * year` are significantly different from zero for the specifications using all comparison schools and the RDD hybrid.

### **Placebo Tests**

#### ***Attainment, post-16 participation***

Results of placebo tests setting 2019 as a fake treatment year are shown in Appendix A, Table 7. Estimates for all four outcomes fitted under OLS are not significant. They all appear to be reasonably powered.

The estimated effect sizes for the placebo tests suggests that pre-existing differences between treatment and comparison groups (whether based on all comparison schools or the DRD subset) were generally small (all below 0.1 absolute standard deviations).

It should be noted that the controls do not offer much in the way of explanatory power. The  $R^2$  value for maths (0.21) implies a combined correlation of 0.45 between the outcome and the controls. Correlations are lower still for other outcomes.

However, there isn't a huge amount of between-school variance (as shown in the rho column).

The tab also shows the results of fitting the models using probit regression for the binary outcomes. These show that the effect sizes are similar to the linear probability models but that the MDES are slightly larger.

#### ***Re-integration, absence***

Results of placebo tests setting 2020 as a fake treatment year are shown in Appendix A, Table 9.

The estimated effect sizes for the placebo tests suggests that pre-existing differences between treatment and comparison groups (whether based on all comparison schools or the DRD subset) were generally small (all below 0.1 absolute standard deviations).

Estimates for both outcomes are not significantly different from zero. The analysis appears reasonably well powered. Absence in particular exhibits little in the way of between-school variation.

## Matching

We also attempt to calculate treatment effects using a two-stage process of propensity score matching of school-level data followed by regression using pupil-level data.

In this analysis, we also use the most recent year of data available for each outcome as a “dummy” treatment year. We first undertake school-level matching using data from previous years to find a set of comparison schools as similar as possible to the schools participating in the treatment. We then link the results of the matching exercise to pupil-level data to estimate treatment effects.

We use two approaches to matching:

- 1:1 nearest neighbour matching (NN)
- Covariate balancing propensity score matching (CBPS)

Both methods have their respective merits. NN is conceptually straightforward to understand. One non-participant is selected for each participant. However, balance in school-level matching variables is not necessarily achieved. CBPS achieves optimal balance in the school-level matching variables. However, this involves using weights in the regression models.

### *School-level matching*

We run matching exercises for each outcome separately. This means, for example, that the set of comparison schools for KS4 English may not be the same as the set for KS4 maths.

Before matching, we filter out potential comparison schools located in local authorities with a below average score on our serious violence measure, i.e., the sum of the percentile scores for serious violence offences and hospital admissions for assault with a sharp object. We do this because including it within the matching routine led to poor balance on the other variables (see Addendum 1 to Appendix A).

The variables we include in all specifications are:

- Variables related to the local authority in which the school is located
- 2020 population estimate for the local authority
- Variables related to the characteristics of pupils attending the school in the three years<sup>52</sup> pre-treatment
- Mean annual cohort size

---

<sup>52</sup> Or however many years are available if there are fewer than three

- % female pupils
- % of pupils observed to have ever been in need
- % of pupils observed to have ever been looked after
- % of pupils first admitted to an AP school before academic age 15
- % of pupils of a white ethnic background
- % of pupils ever permanently excluded prior to joining the school
- Mean number of suspensions per pupil
- % of pupils with an EHC plan
- % of pupils ever to have been eligible for free school meals
- % of pupils who also attended a mainstream school during Year 11
- Mean standardised KS2 score<sup>53</sup>
- Variables summarising mean outcomes for each of the previous three pre-treatment years

CBPS calculates a set of weights to balance schools on pre-treatment data. These sum to 1 for both treated schools and comparison schools. We adjust these weights in the following stage.

There are some issues with achieving strong matches because the treated schools are both unusual and homogenous with respect to several of the matching variables, most notably the local authority level serious violence measure. We attempt to correct for this by filtering out schools with a below average score for this measure before matching, as described above.

Addendum 1 to Appendix A discusses this in more detail and shows the balance achieved from the matching processes for each outcome, using both the methodology described above and an alternative method.

### **Outcome models**

We then apply the results from the school-level matching to pupil-level data and produce estimates of treatment effects for a dummy treatment year. These models are alternatives to the difference-in-difference falsification tests described above.

We again use regression with standard errors clustered by schools. The same pupil-level controls are used as in the difference-in-difference models.

---

<sup>53</sup> Currently not included in the matching for younger pupils as the data were not available

For CBPS, we divide the school-level weights calculated during the matching process by the number of pupils at each school. We then use these pupil-level weights in the regression.

#### ***Attainment, post-16 participation***

We report here on placebo tests obtained using the preferred two-step methodology. Results can be found in Appendix A, Table 7.

Estimates for three of the four outcomes fitted under OLS are not significant for either the NN or CBPS matching method. However, estimates for whether pupils are recorded as NEET for 1 month or more at academic age 16 are significant at the 5% level using CBPS weighting, but not using NN matching.

The tab also shows the results of fitting the models using probit and logit regression for the binary outcomes. These show effect sizes similar to the linear probability models and, again, none of the estimates are significant using NN matching, while the NEET outcome is significant at the 5% level using CBPS weighting.

There are also some differences in the estimated effect sizes depending on whether CBPS or NN is used, although the  $R^2$  is similar.

#### ***Re-integration, absence***

We report on results obtained using the preferred two-step methodology in Appendix A, Table 9.

None of the estimates presented are significant at the 5% level. OLS, probit and logit models are fitted to the binary re-integration outcome, and similar effect sizes are estimated under the different models, as well as under the estimates based on CBPS and NN matching.

### **Regression Discontinuity**

The sum of serious violence and hospital admission percentile scores is used to identify APST areas. Those CSPs/LAs with a combined score greater than 1.82 are APST areas and those with combined scores below 1.82 are non-APST areas. This discontinuity can be exploited for estimation purposes. The intuition is that areas just above the cutoff are likely to be similar to those just below the cutoff such that we can view treatment status among those close to the cutoff as being as good as randomly allocated. Treatment-control comparisons of outcomes among those close to the cutoff can therefore provide an estimate of treatment impact.

In the context of RDD, there is a distinction between sharp design and fuzzy design. With a sharp design, treatment status is determined entirely by the threshold whereas, with a fuzzy design, treatment status is influenced but not determined by the threshold (this can arise where there is non-compliance, for example). Our case is fuzzy in the sense that only a single school per LA is selected as an APST school in those areas that lie above the cutoff. Hence, while all schools below the cutoff are non-APST schools, not all schools above the threshold



are APST schools. Since the outcomes we consider are all pupil-level, we deal with this (and sharpen the design) by dropping non-participating schools in APST areas.

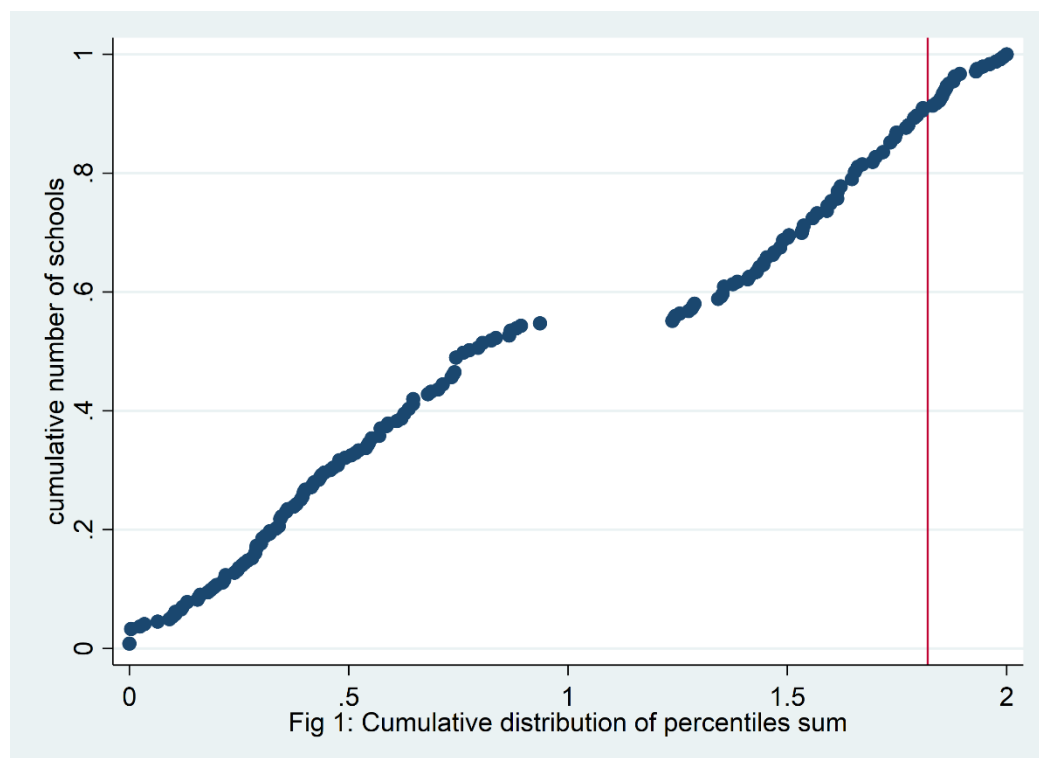
For an RD approach to be valid, we require the percentile score to be continuous in the region of the cutoff and also that the outcome that would prevail in the absence of the treatment be continuous around the cutoff.<sup>54</sup> The chart below (Figure 3) demonstrates this, showing the cumulative number of AP schools in 2021 as the sum of the percentile scores increases. There is a break in the series between 0.94 and 1.23 but this is sufficiently far from the cutoff (shown as a red vertical line) as to not be a concern (recall the estimator will only use observations close to the cutoff). More important is that there is no apparent break in the region of the cutoff; that it is smooth and continuous. This is to be expected since the use of prior scores for identifying APST areas preclude the sort of manipulation of the running variable that this type of graph is intended to show.

Subsequent graphs (Figure 3) concentrate on schools close to the cutoff and examine whether outcomes in previous years show a change at the cutoff. We define this closeness in two ways: whether the sum of the percentile scores is within 0.1 of the cutoff (in the jargon, a bandwidth of 0.1) and a bandwidth of 0.2 (since the maximum of the sum of percentile scores is 2, this latter case has the consequence of including all APST schools). Looking at earlier years shows whether and how outcomes differed between schools above and below the cutoff prior to APST. The rationale for examining this is that it provides a clue as to whether untreated potential outcomes (the unobserved outcomes that would have prevailed without APST) would be expected to be continuous around the cutoff after APST is introduced.

---

<sup>54</sup> We also require that the percentile score cannot be manipulated by schools (or others); this is guaranteed in our case.

**Figure 3: Distribution of the serious violence and hospital admission scores**



We perform this for the years 2014 to 2019. We avoid the years 2020 and 2021 so results are not affected by COVID-related distortions (including missing KS4 outcomes). The table below (Table 19) shows the number of observations for each year (final pair of columns). It also provides the means of the two outcomes considered here: whether a Year 11 pupil participated in education soon in Year 12 (top panel) and whether a KS3 (Year 7-9) pupil re-integrated in the following academic year (bottom panel):

**Table 19: Mean post-16 participation and re-integration outcomes by year**

	Initial post-16 participation %		Num. pupils	
	C	T	C	T
2014	52%	56%	5,633	1,191
2015	52%	57%	6,112	1,331
2016	49%	51%	6,538	1,226
2017	49%	49%	6,762	1,335
2018	49%	53%	6,590	1,375
2019	48%	49%	6,602	1,313
	Re-integration %		Num. pupils	
	C	T	C	T
2014	39%	40%	4,716	1,086
2015	37%	38%	4,997	1,075
2016	33%	33%	5,554	1,062
2017	32%	28%	5,862	1,195
2018	31%	33%	6,098	1,144

2019	30%	28%	5,871	1,044
------	-----	-----	-------	-------

The first pair of graphs below (within Figure 4). focus on the participation outcome in 2014 for Year 11s (top panel: variable name `initial_post`) and the re-integration outcome in 2014 for Years 7-9 (bottom panel: variable name `mains_ext`). In both cases, the results on the left hand side impose the bandwidth of 0.1 and the results on the right hand side impose a bandwidth of 0.2. In each chart the mean outcome for each outcome is plotted against that school's percentile sum, with the size of each circle reflecting the number of pupils at the school (schools with 10 or fewer pupils are not shown on the graph). The y-axis denotes the proportion of pupils at a school who participated or re-integrated, respectively.

Three lines are shown either side of the cutoff. The red line marks the mean outcome among schools on each side but within the bandwidth (the local mean). The green line shows an estimated linear relationship between the points on each side (the local linear regression). The blue line shows an estimated quadratic relationship between the points on each side (the local quadratic regression). The RD estimator is essentially the vertical difference between same-colour lines at the cutoff, where red, green and blue lines correspond to progressively more flexible ways of modelling the relationship between the outcome and the score variable. What we hope to see in these graphs is that there is smooth continuous progression around the cut point such that the RD estimate is close to zero.

Summarising the results for participation, we see that under the narrower bandwidth (0.1), local means (red lines) perform best (in the sense of providing an estimate closest to zero). Under the wider bandwidth (0.2), the local mean performs at least as well as the local linear regression. Both perform somewhat better than under the narrower bandwidth. The local quadratic regression again performs poorly.

For re-integration, the bandwidth of 0.2 again provides more consistently better estimates than the bandwidth of 0.1. The local mean estimator consistently out-performs the more flexible specifications.

**Figure 4: School outcomes and mean, linear and quadratic fits above and below cutoff, by bandwidth and cohort**

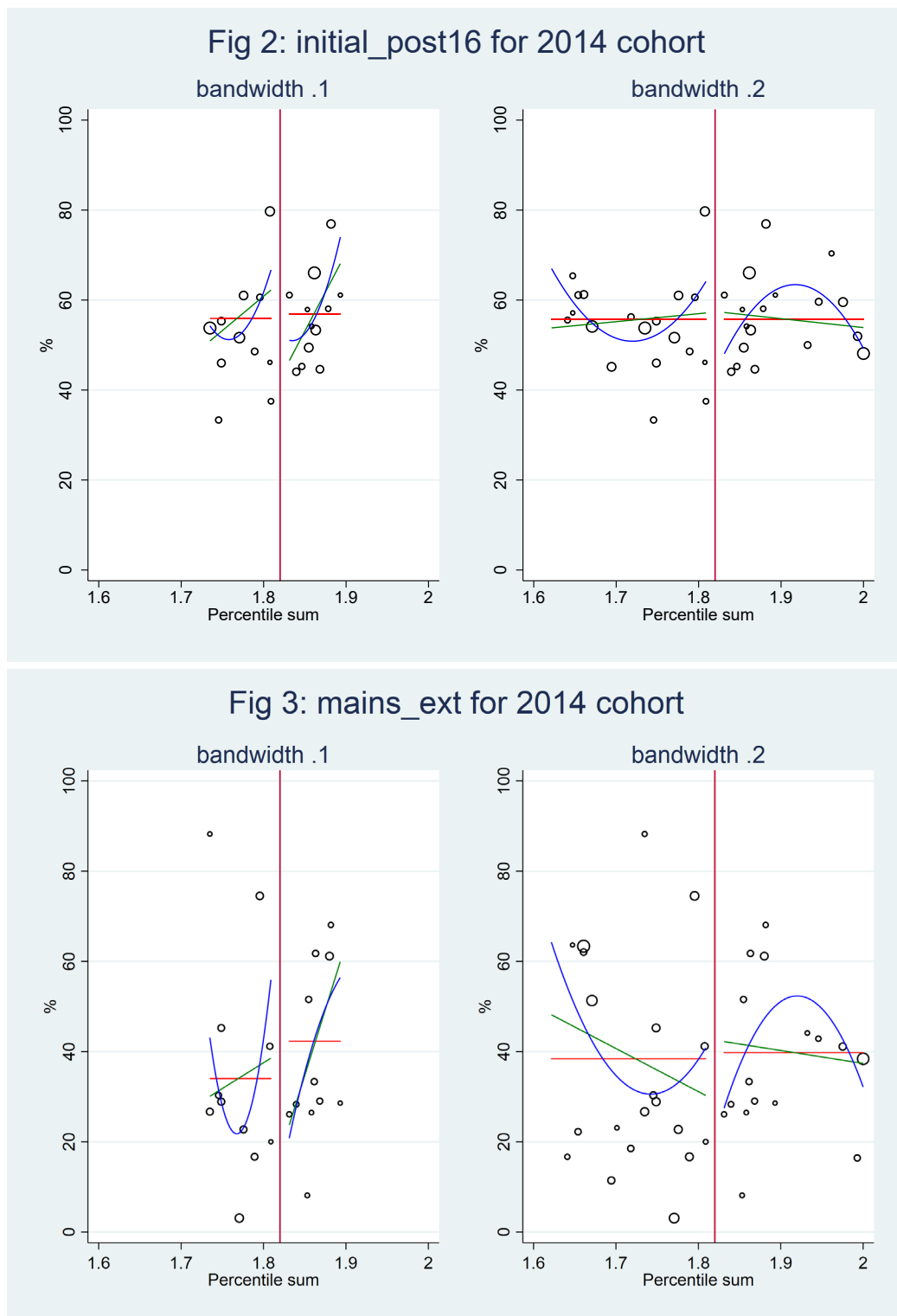


Fig 4: initial\_post16 for 2015 cohort

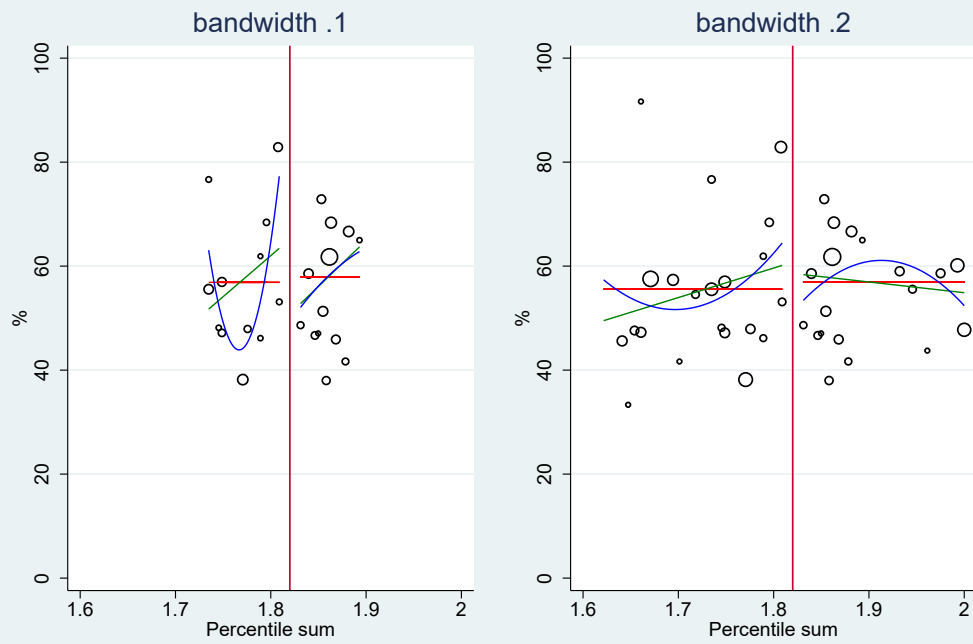


Fig 5: mains\_ext for 2015 cohort

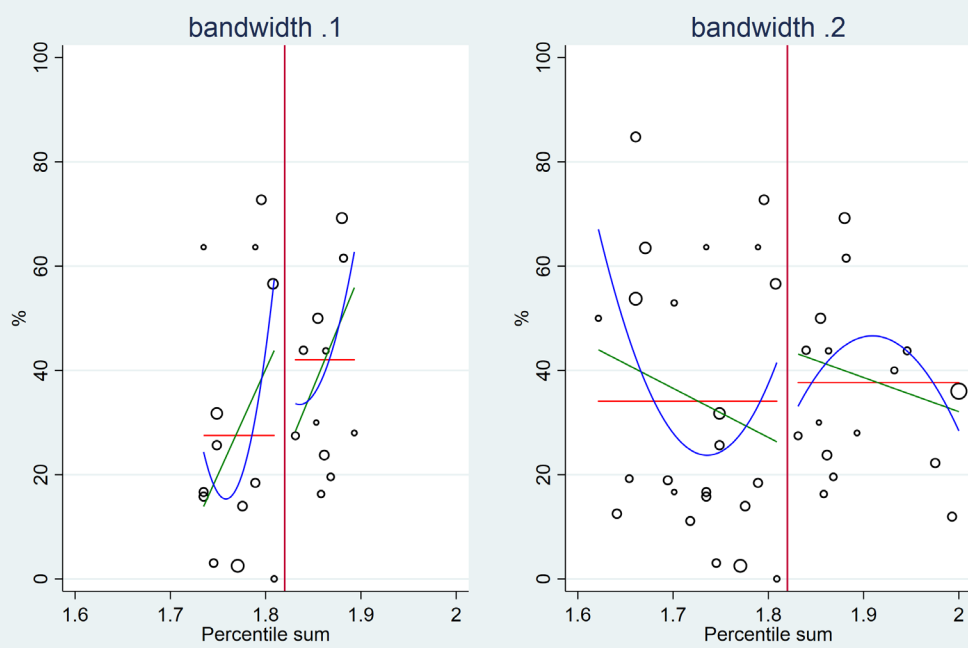


Fig 6: initial\_post16 for 2016 cohort

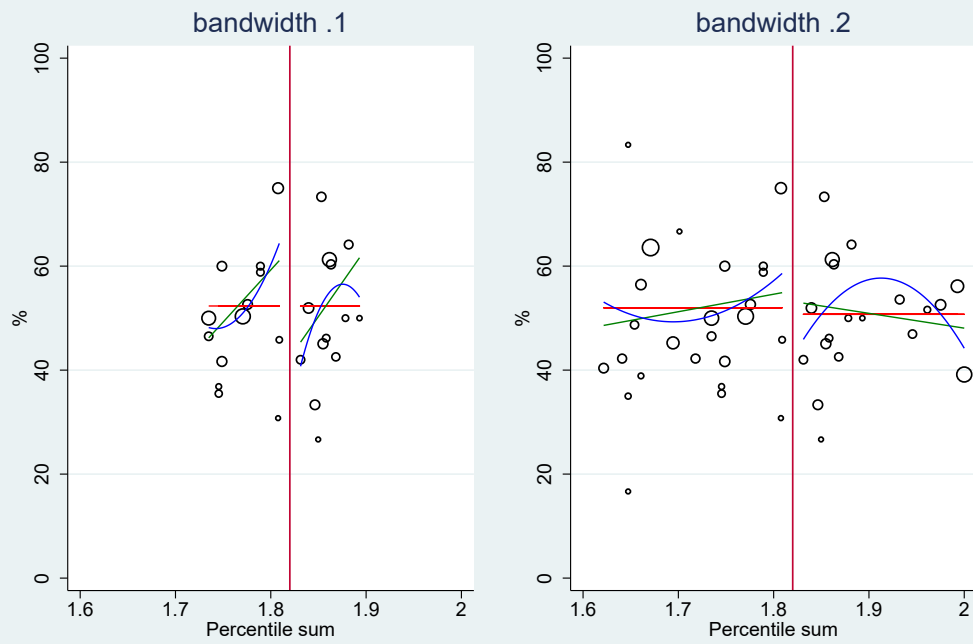


Fig 7: mains\_ext for 2016 cohort

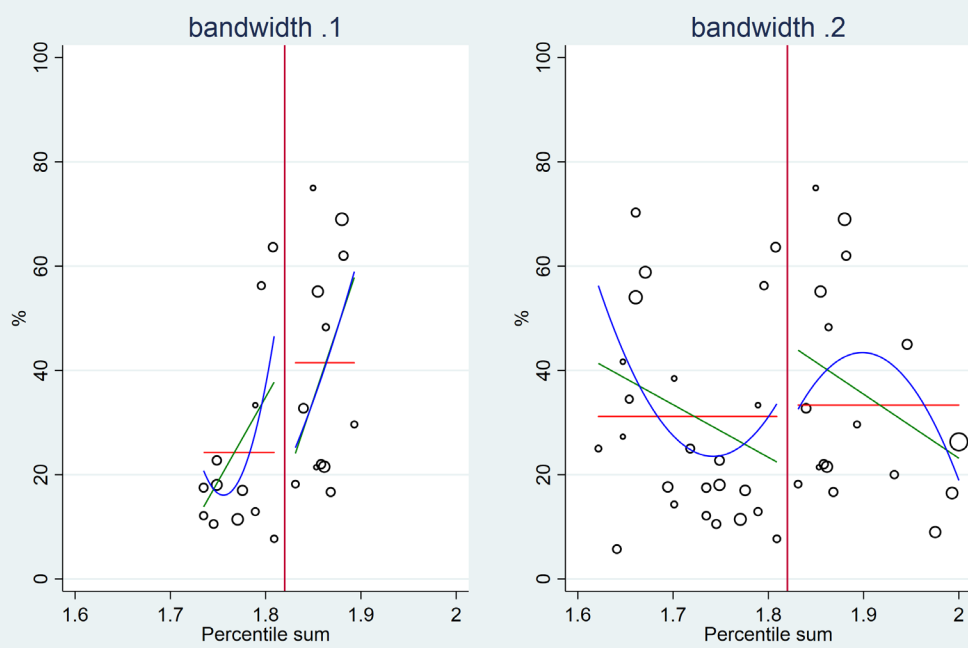


Fig 8: initial\_post16 for 2017 cohort

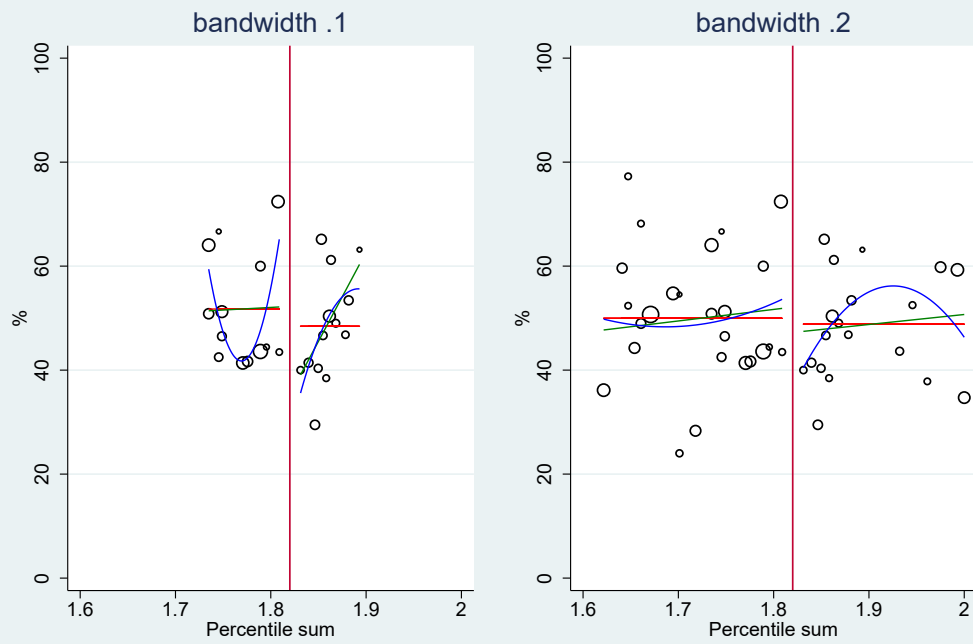


Fig 9: mains\_ext for 2017 cohort

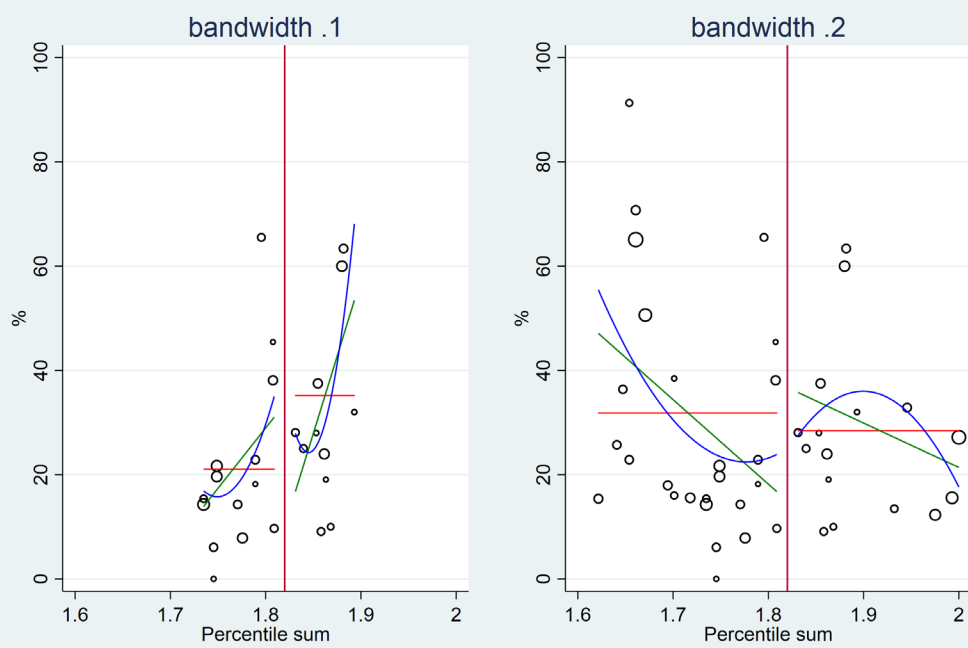


Fig 10: initial\_post16 for 2018 cohort

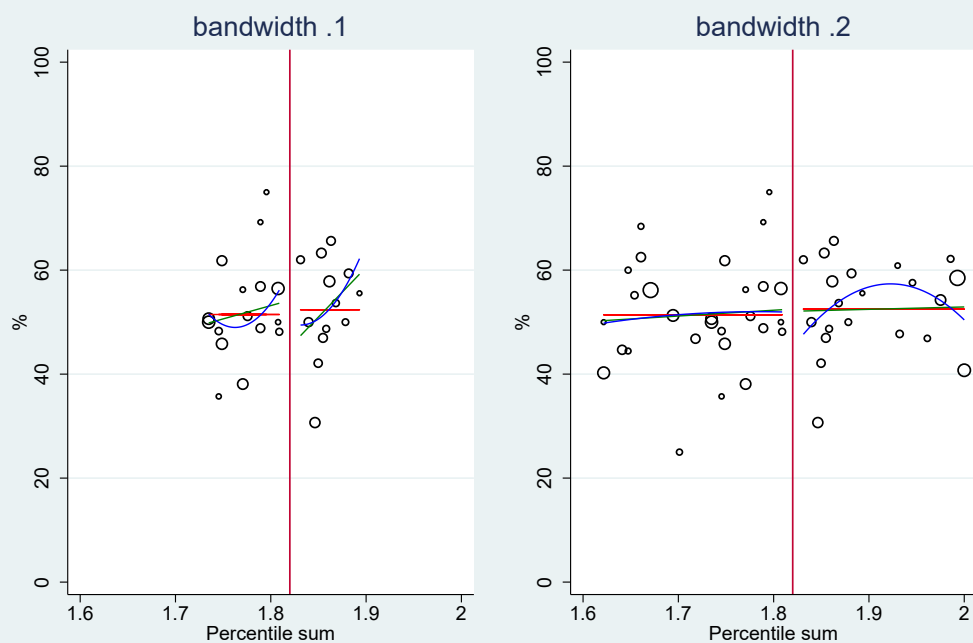


Fig 11: mains\_ext for 2018 cohort

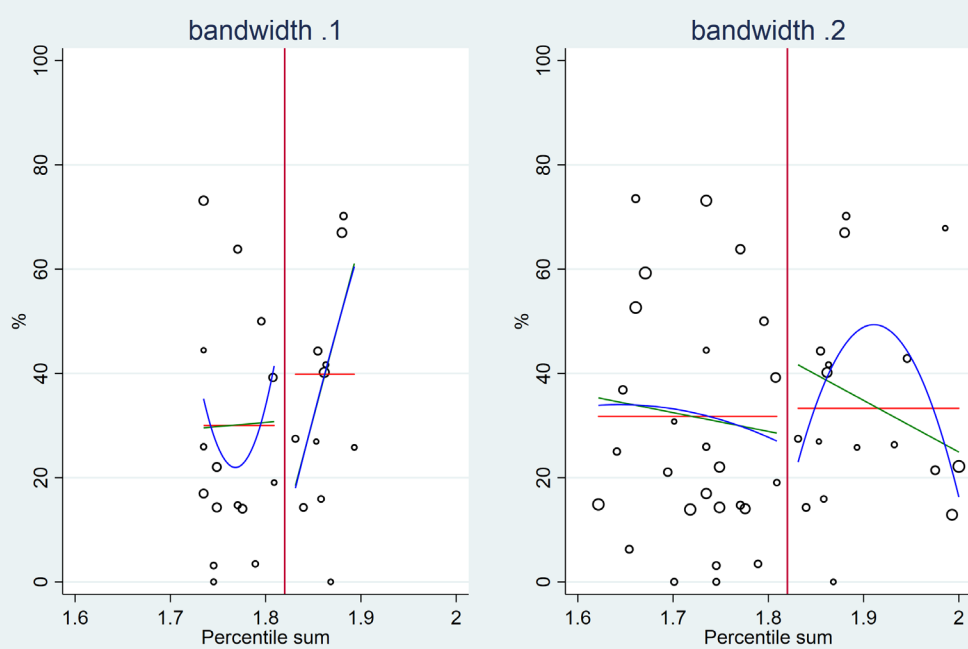




Fig 12: initial\_post16 for 2019 cohort

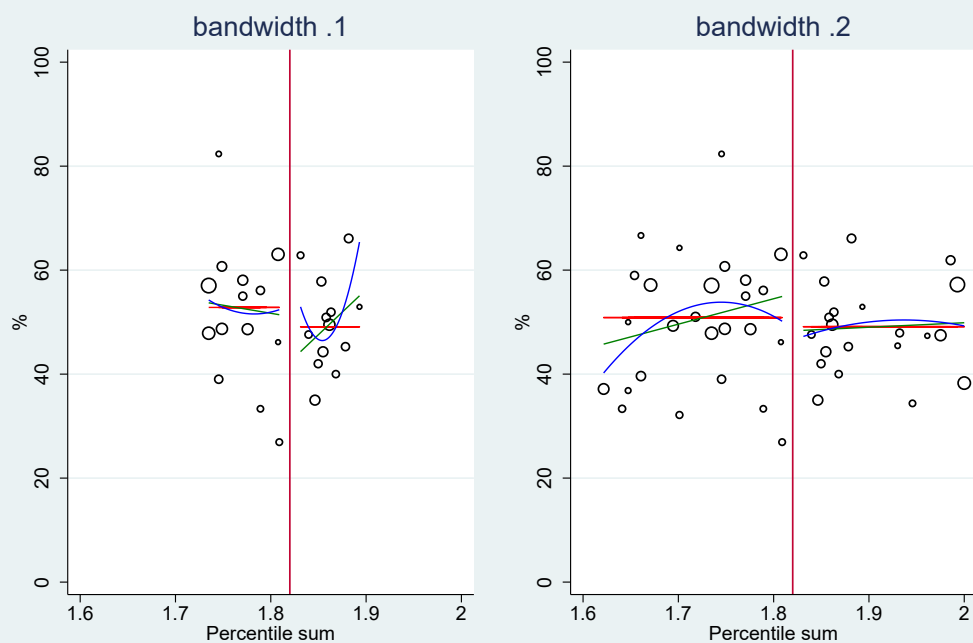
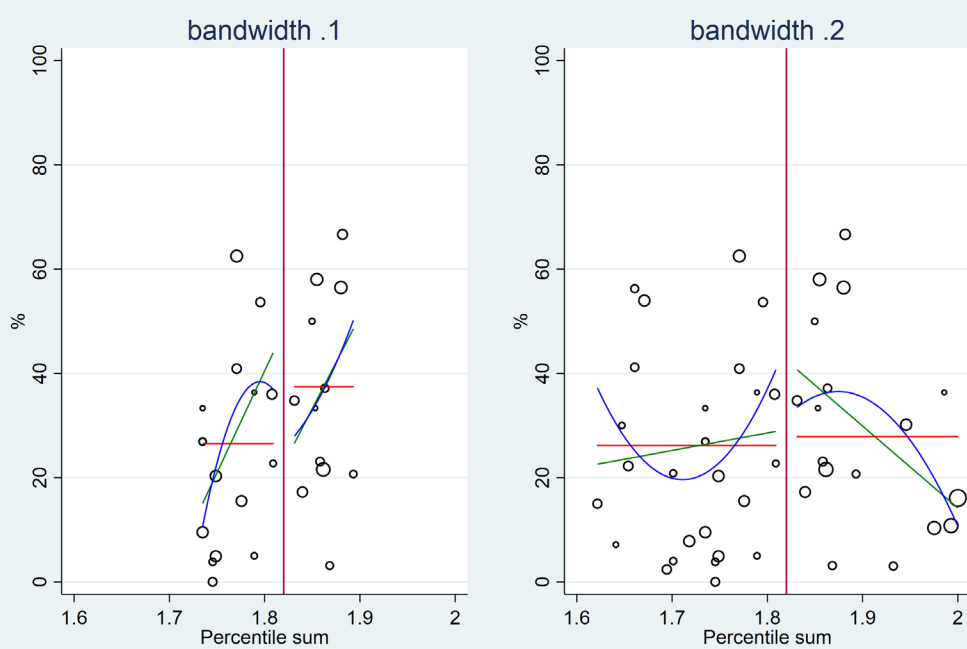


Fig 13: mains\_ext for 2019 cohort



Note: data points show school average outcomes. Those based on 10 observations or fewer are suppressed.

Another check on the validity of the RD design is to inspect whether covariates change either side of the cutoff. The intuition here is that if there is a difference in the value of covariates either side of the cutoff, we have less confidence that the RD estimate captures the effect of the treatment alone since it may also be capturing the effect of compositional differences.

The charts below (within Figure 5) consider a selection of covariates. For compactness (and in light of its superior performance shown already), only results for the wider bandwidth are shown. Furthermore, we only show results for year 11s and only consider the 2018 and 2019 cohorts. The format of the charts is as before.

The results are mixed. KS2 performance (both English and maths), FSM percentage and IDACI are very similar either side of the cutoff, sex and (for the 2019 cohort) SEN percentage show small differences while ethnicity (captured as percentage white British, wbri), prior exclusions and prior suspensions show substantial differences.

**Figure 5: Characteristics of schools' pupils and mean, linear and quadratic fits above and below cutoff, by bandwidth and cohort**

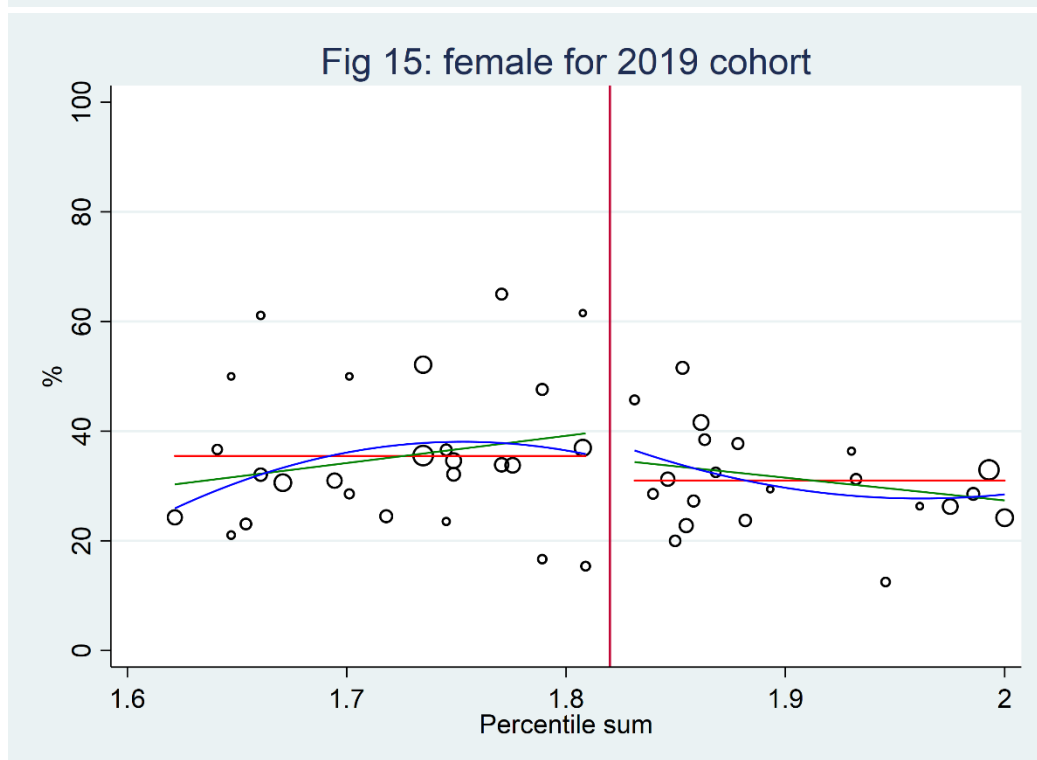
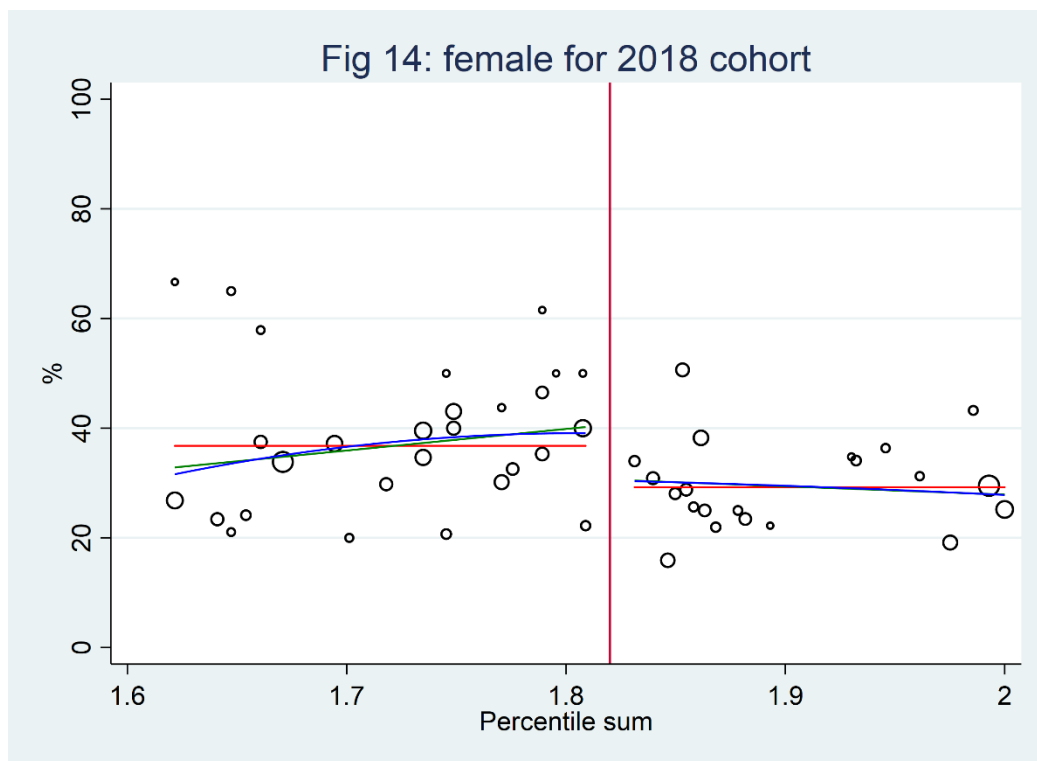


Fig 16: wbri for 2018 cohort

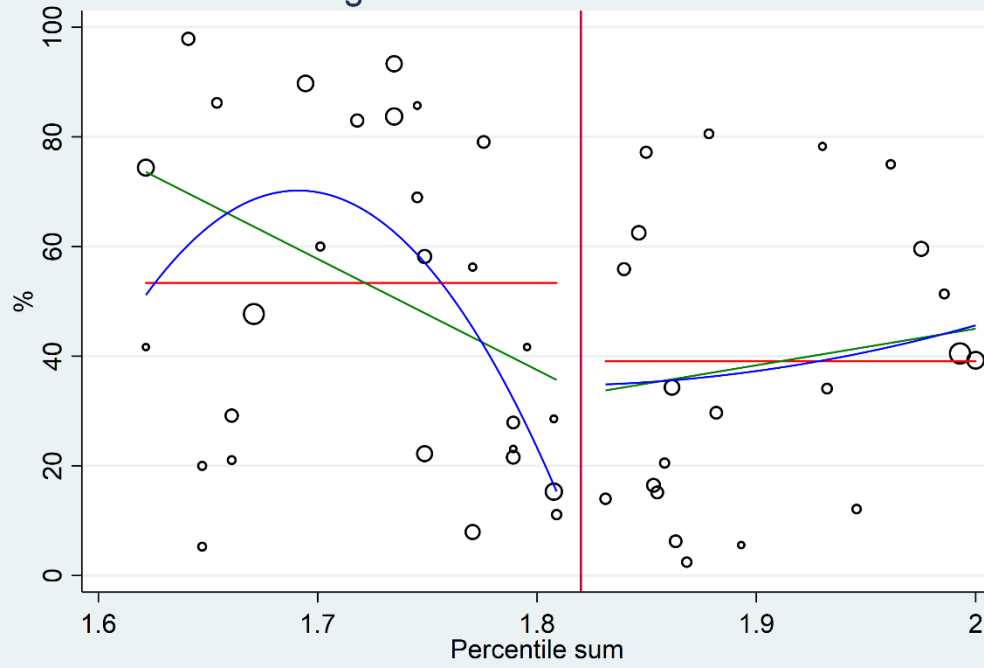


Fig 17: wbri for 2019 cohort

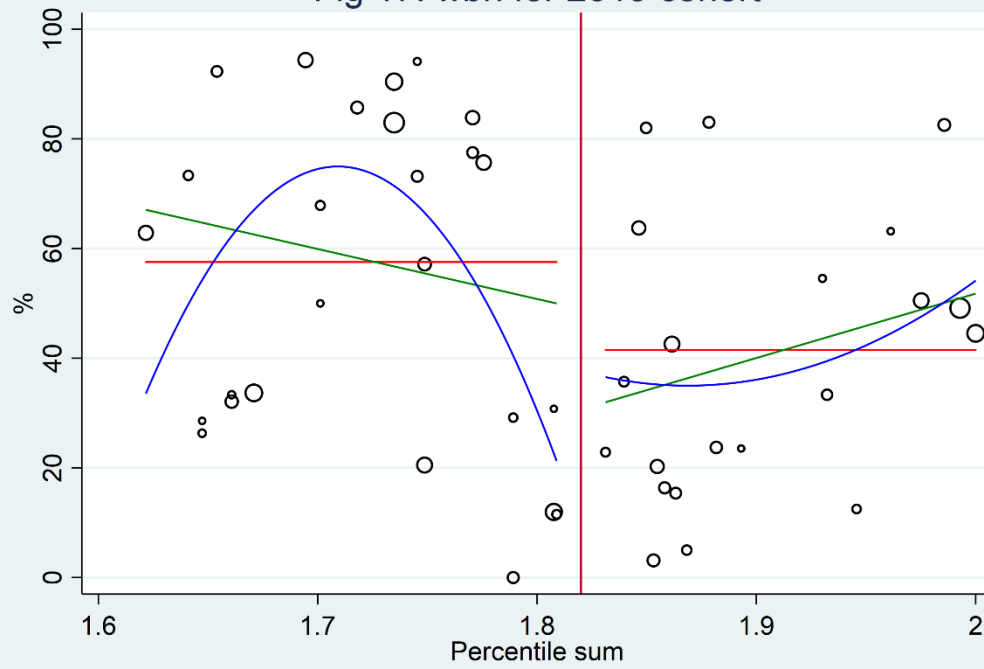


Fig 18: ks2\_mat for 2018 cohort

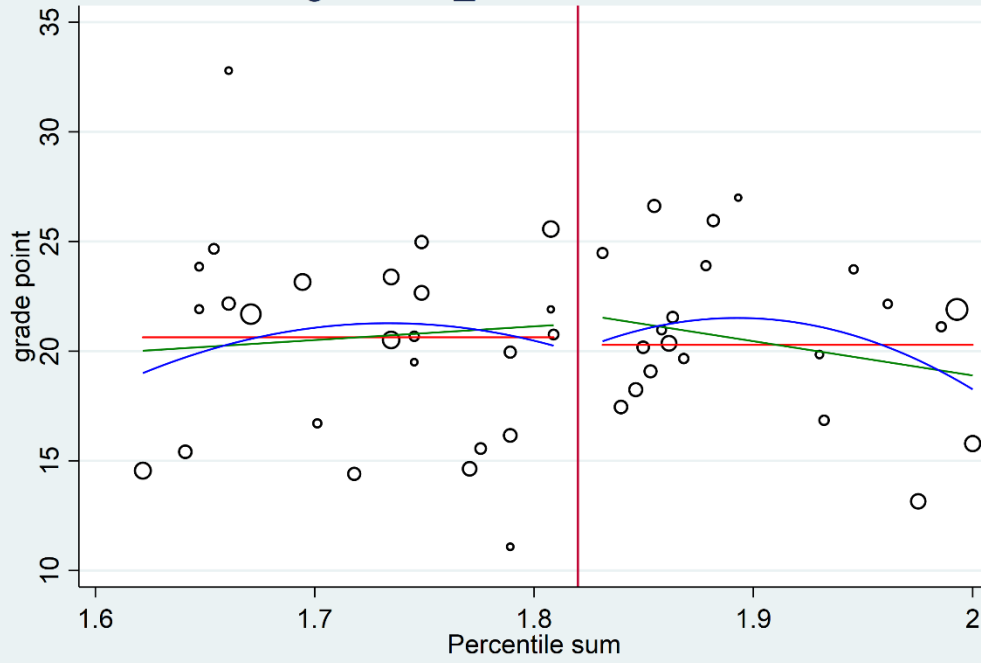


Fig 19: ks2\_mat for 2019 cohort

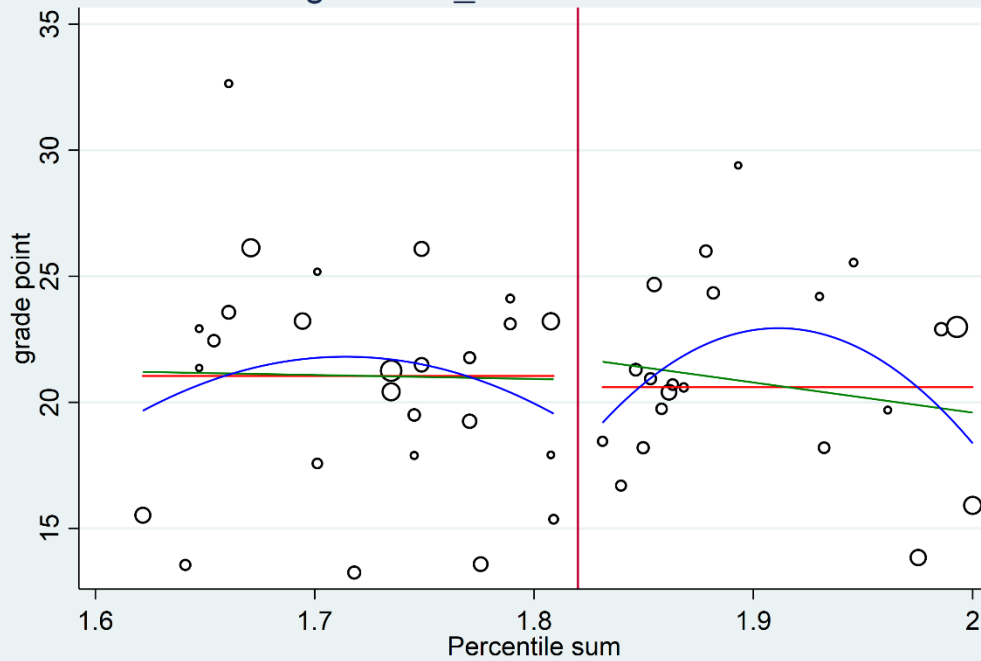


Fig 20: ks2\_eng for 2018 cohort

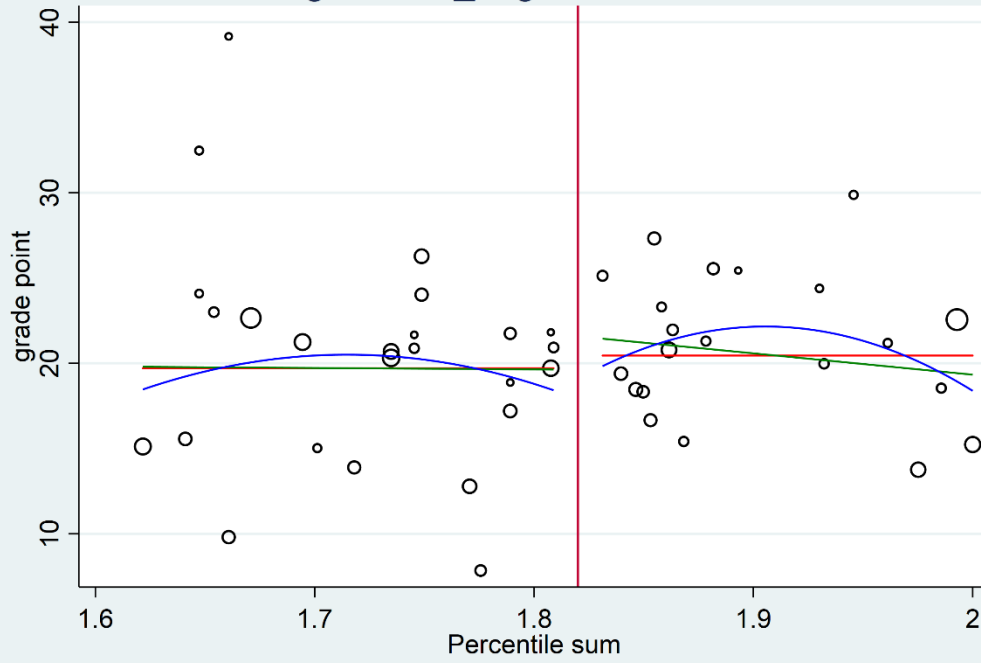


Fig 21: ks2\_eng for 2019 cohort

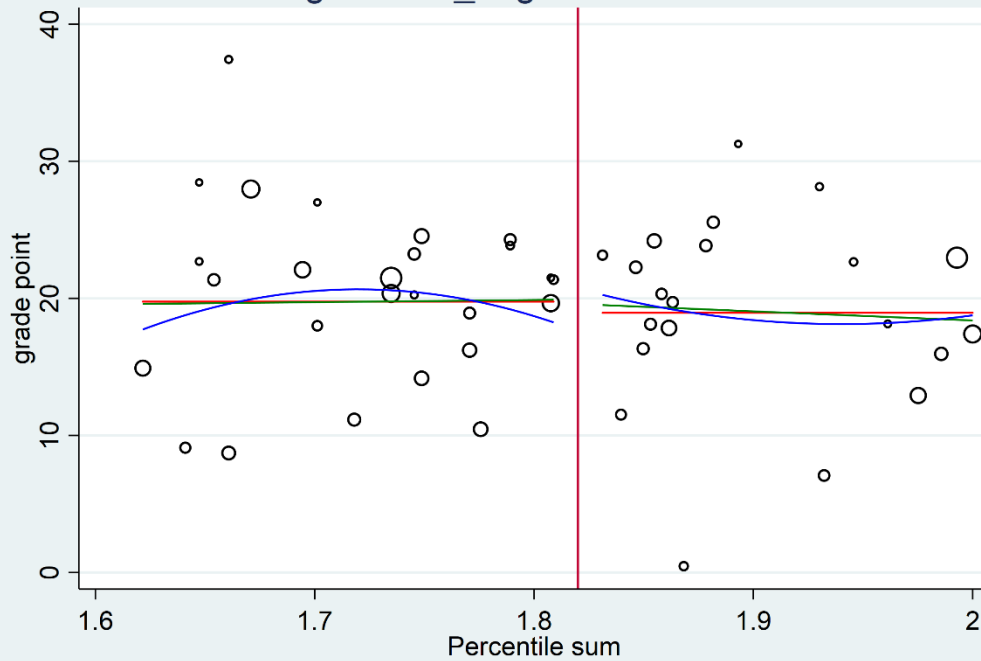


Fig 22: prior\_sen for 2018 cohort

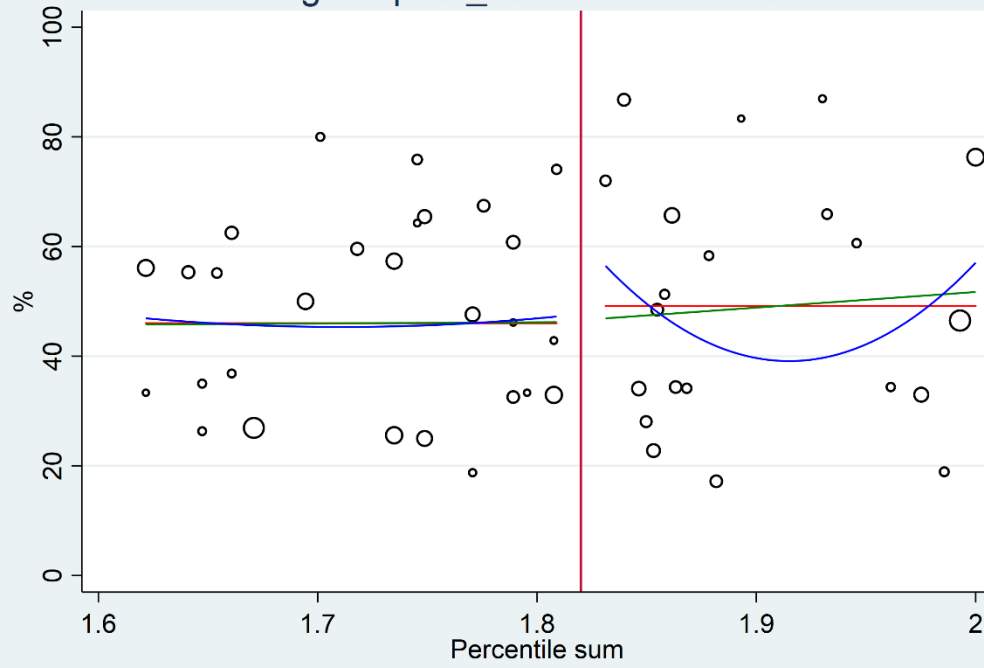


Fig 23: prior\_sen for 2019 cohort

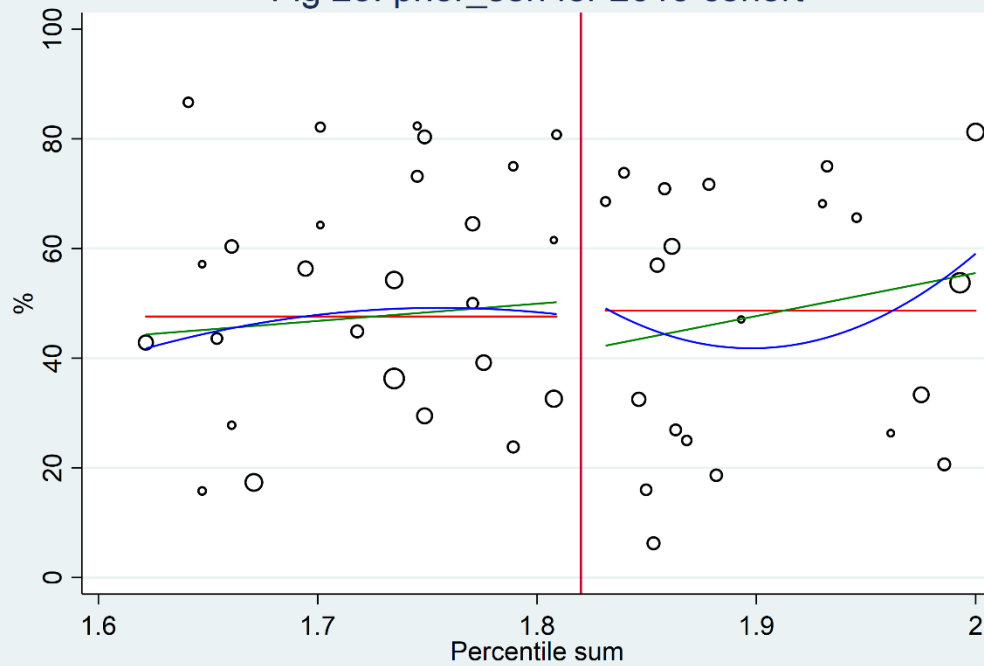


Fig 24: prior\_perm\_ex for 2018 cohort

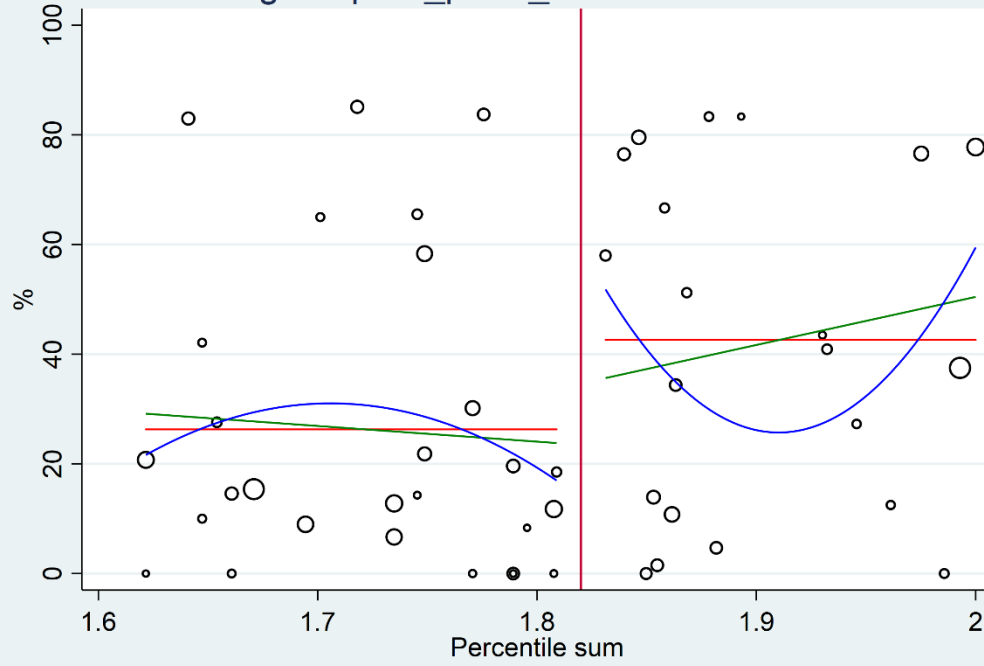


Fig 25: prior\_perm\_ex for 2019 cohort

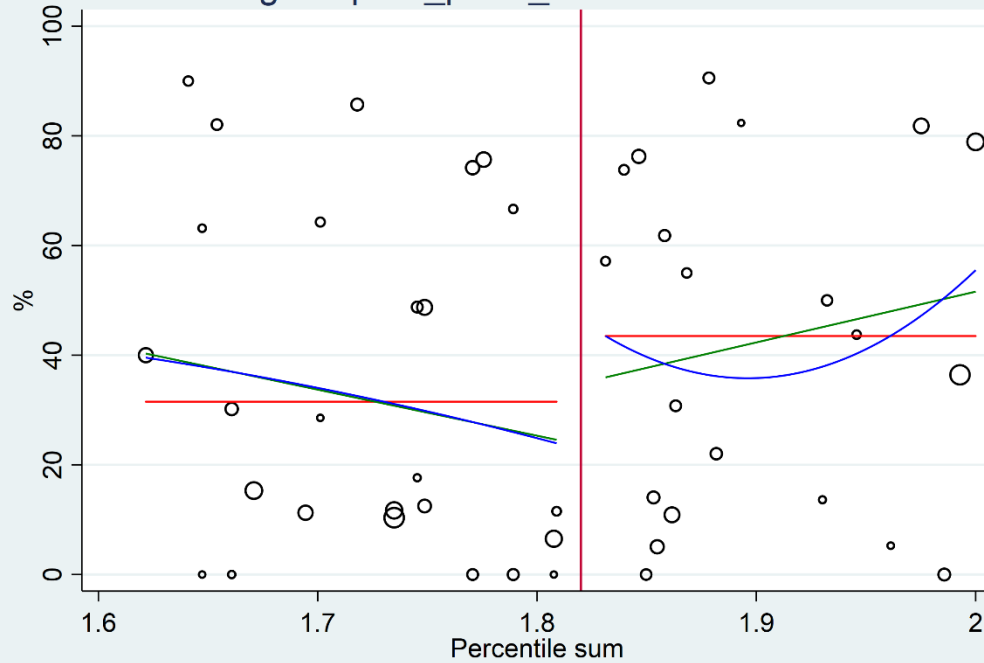




Fig 26: prior\_suspensions for 2018 cohort

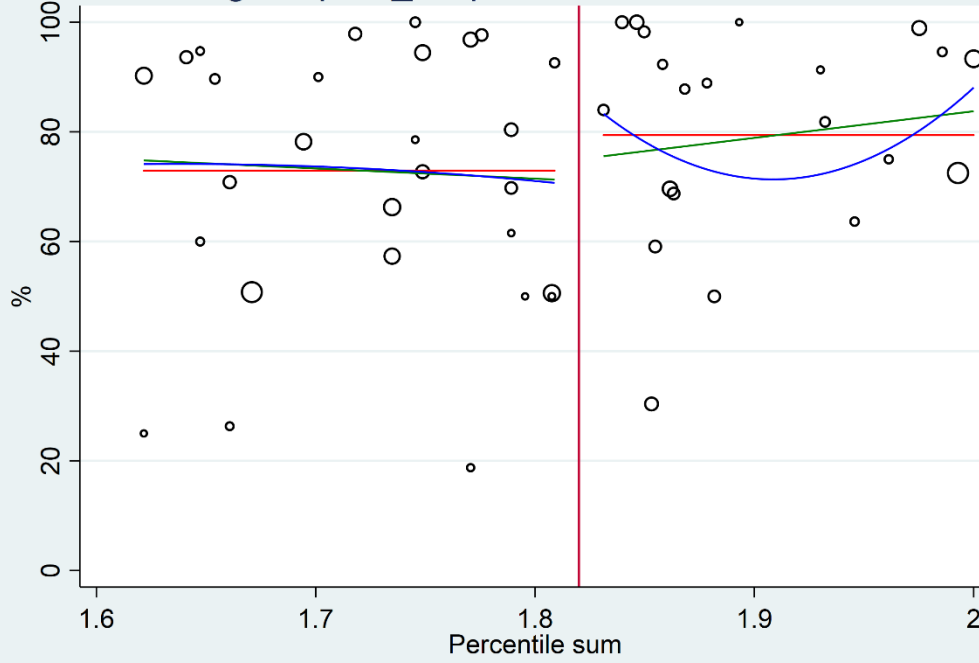


Fig 27: prior\_suspensions for 2019 cohort

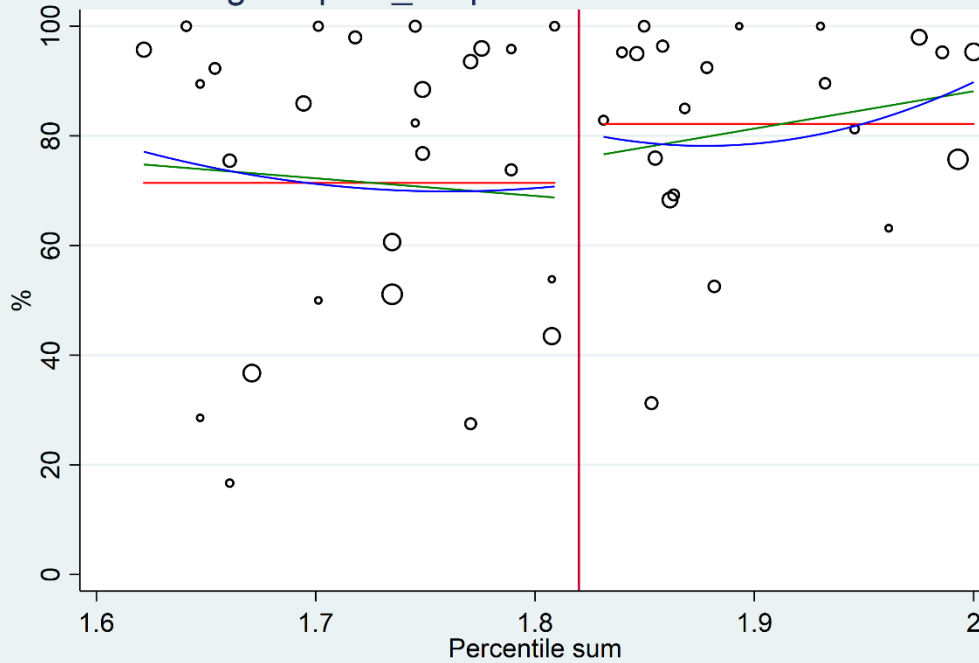


Fig 28: prior\_fsm\_percent for 2018 cohort

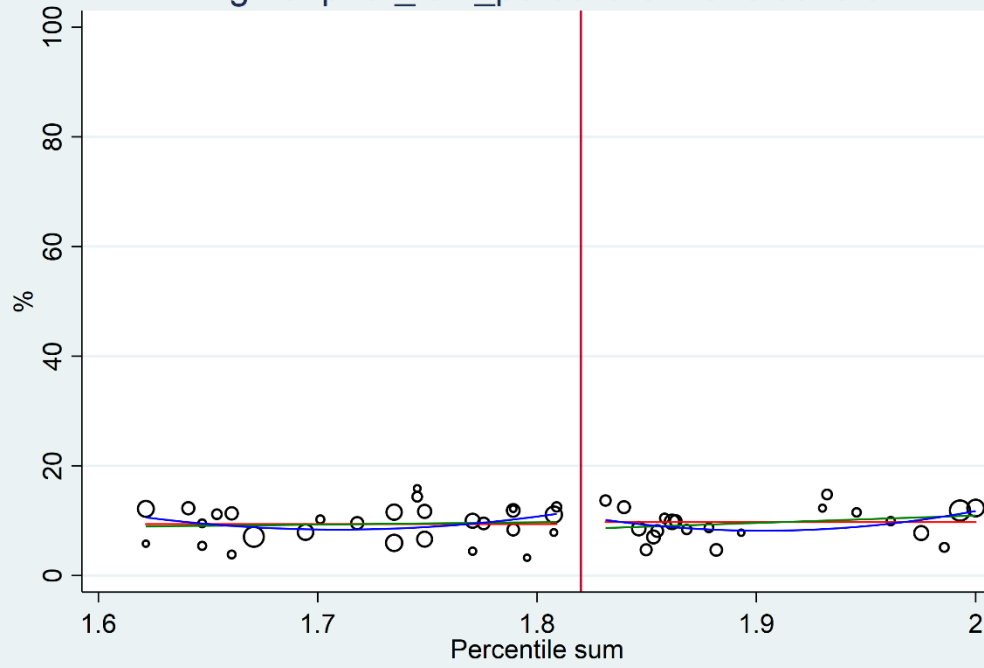


Fig 29: prior\_fsm\_percent for 2019 cohort

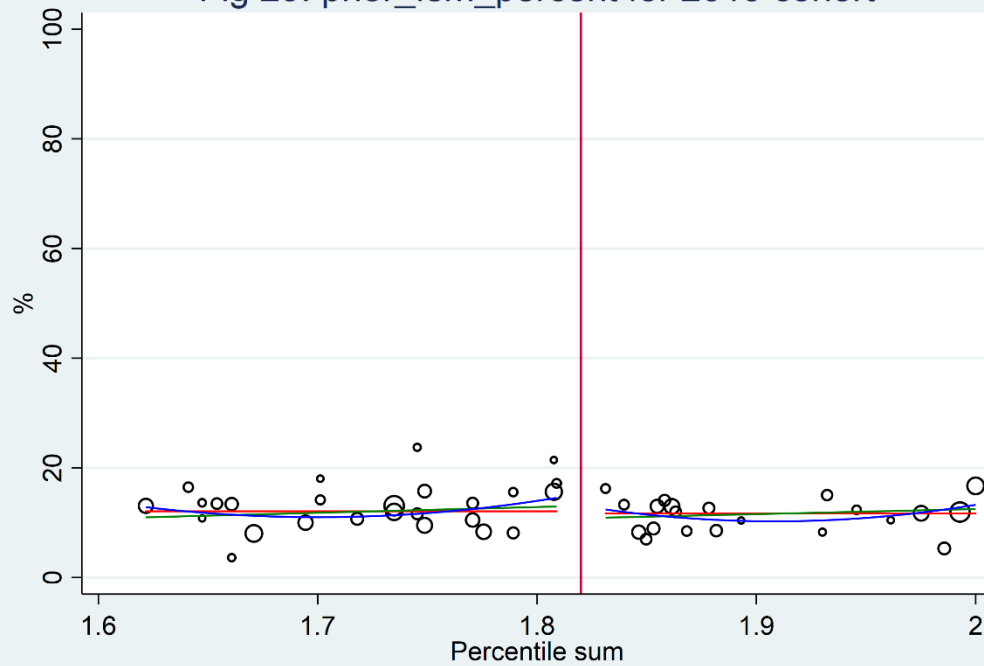


Fig 30: prior\_sen\_percent for 2018 cohort

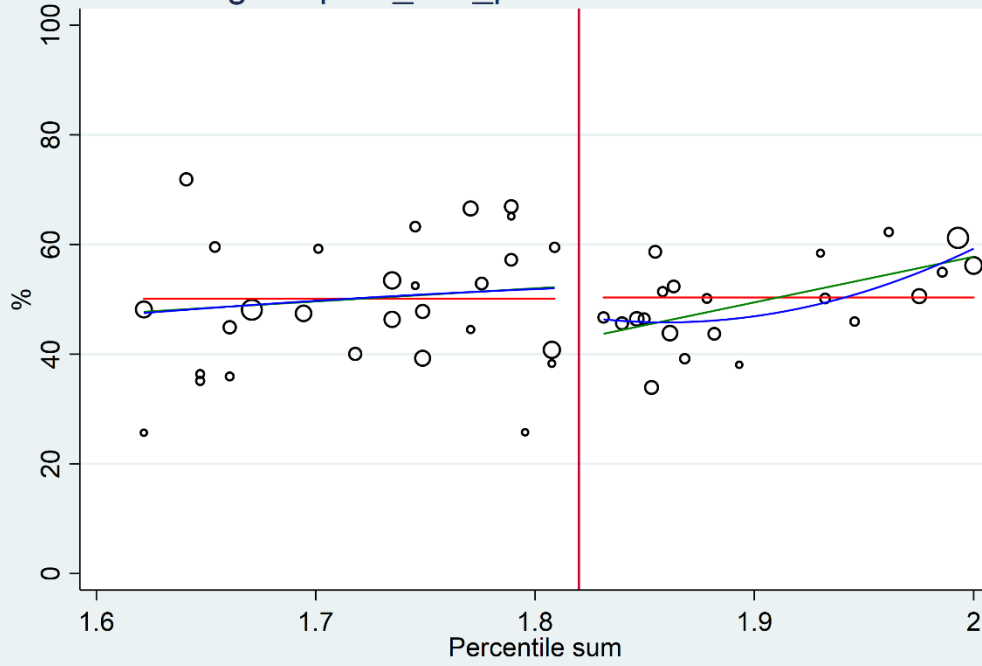
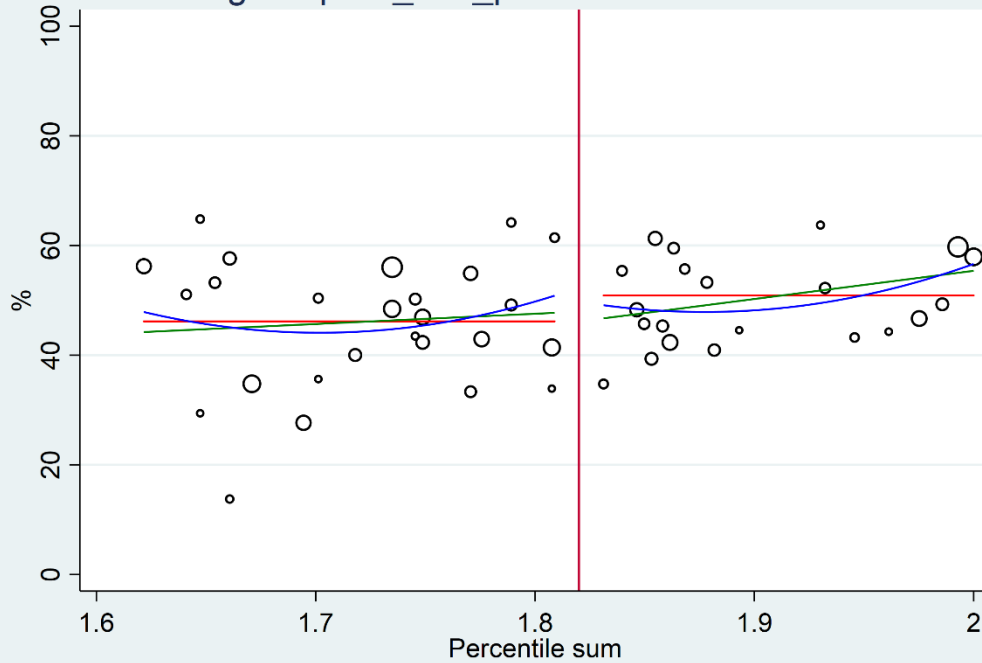
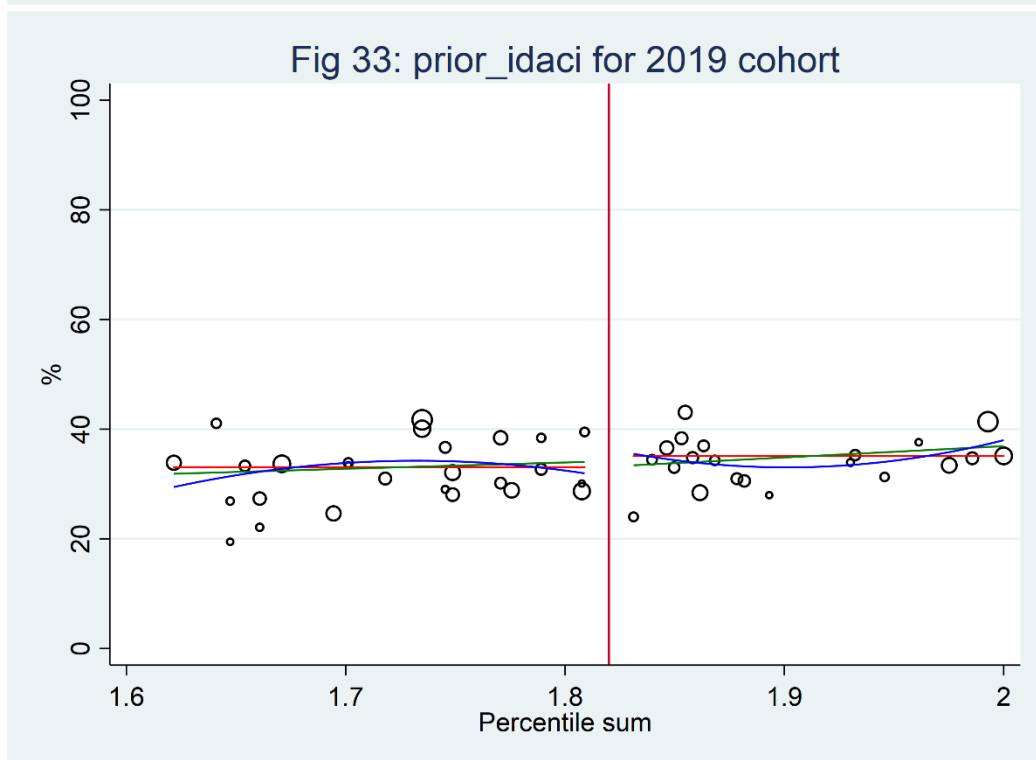
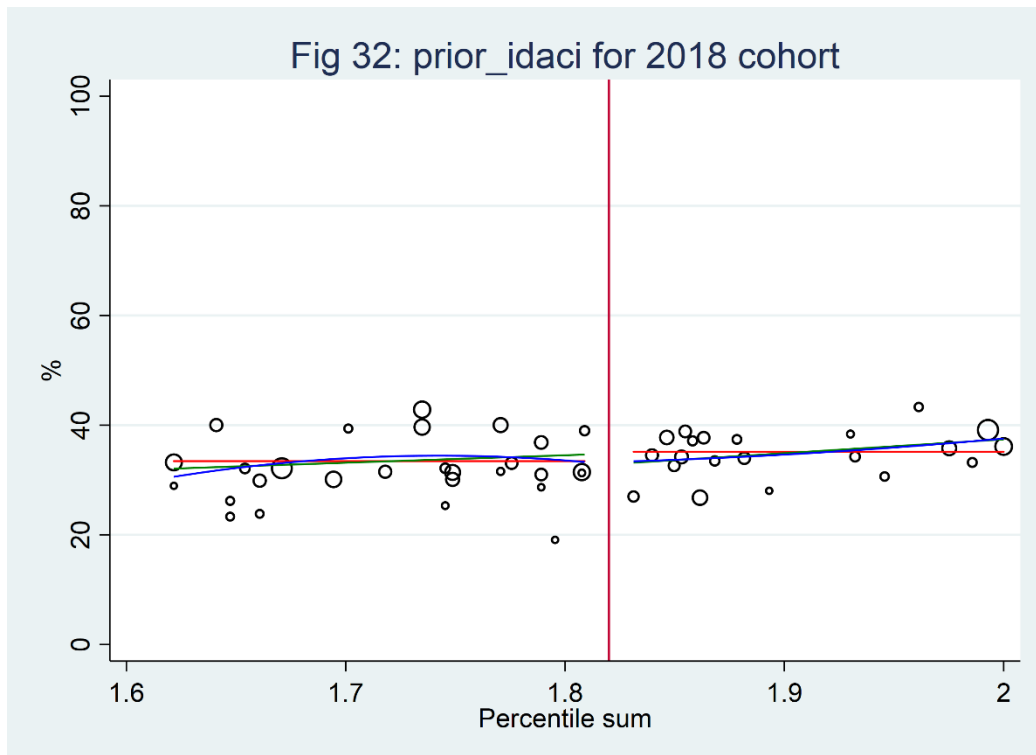


Fig 31: prior\_sen\_percent for 2019 cohort





Note: data points show average school outcomes. Those based on 10 observations or fewer are suppressed.

The fact that there are substantial differences for some characteristics is a cause for concern for the RD estimator. At the very least, regression should control for the influence of these variables. Differencing may be a more effective way to correct for possible confounding arising from compositional differences. We therefore propose a difference in RD (DRD) approach as a robustness

check. Adopting the linear mean specification for this and a bandwidth of 0.2 motivates the RDD hybrid approach discussed above.

## **Conclusion**

We undertook preliminary analysis of pre-treatment data for participating and non-participating schools in order to develop processes to create analytical datasets and inform the statistical plan.

This firstly involved developing several outcome measures (as described in the analysis plan sections on primary and secondary outcomes) that were relevant for the study population, in particular defining a measure of re-integration.

Secondly, we then examined pre-existing differences between participating and non-participating areas in outcomes. We examined differences with and without pupil-level controls.

Thirdly, we tested three different methodological options for estimating treatment effects: difference-in-differences, matching and regression discontinuity.

Finally, we wrote the code to estimate treatment effects and used this to calculate them for a “placebo” year to ensure that non-significant pre-treatment differences were recovered.

All three methods passed this latter test. We opted to pursue a strategy based on difference-in-differences for the analysis plan for two reasons. Firstly, DiD estimates based on all AP schools in the comparison group were estimated more precisely than those based on other methods. Secondly, time-invariant differences between schools can be netted out, thereby controlling for stable unobserved influences.

Analysis of pre-existing differences for each of the proposed outcomes revealed no apparent trends, except in the case of the Key Stage 4 maths outcome. Outcomes for participating schools tend to be lower than those of comparison schools.

As participating schools are located in areas with higher levels of serious violence, we explored an alternative difference-in-differences specification informed by the regression discontinuity design, the differences in regression discontinuity (DRD). This will serve as a test of robustness. We identified a preferred bandwidth of 0.2 for the area-level serious violence score used to select local authorities to participate in the programme. AP schools in local authorities within this cut-off will be selected as a comparison group.

## **Addendum 1 to Appendix A: Balance achieved for each outcome**

This appendix shows the balance achieved from the matching process for each outcome, for both CBPS and NN matching. The graphs included show the standardised mean difference between each of the matching variables, before and after matching. An absolute standardised mean difference of 0.2 or below is generally taken to indicate good balance. The dotted lines on the graphs indicate this level.

As described in the main text, we use a two-step method: we begin by filtering out schools located in LAs with a below average level of the serious violence measure. Then we match the remaining schools using NN or CBPS.

This approach was chosen because of the extremely high standardised mean difference (SMD) in the serious violence measure before matching. The SMD for this variable is 16.5 – the highest absolute standardised mean difference for any other variable is 1.6.

Because the SMD for this variable is so high, it has a strong impact on some of the matching models, leading to matches that improve the SMD for this variable but do not achieve good balance for many of the other matching variables.

Excluding the serious violence measure from the matching process entirely is not an option as the variable was an important part of the selection process. Therefore, we prefer to filter out schools in LAs with below average levels of serious violence before matching. For matches including historical outcomes related to attainment and post-16 participation, this results in matches with much improved balance in the other matching variables, at the expense of weaker matches with respect to serious violence. However, the two-step approach is less successful for matches including historical outcomes relating to re-integration and absence.

Our preference is to use the two-step method for all matches, for the sake of simplicity, although an alternative approach could be to use a two-step method for outcomes related to attainment and post-16 participation, and a one-step method for outcomes relating to re-integration and absence.

The plots below show the balance obtained from both the two-step method and from a one-step method that omits filtering the potential comparison schools before matching, and instead includes the serious violence measure as a matching variable.

We have omitted SMDs of >5 from the plots to improve legibility – this excludes the SMD for the serious violence measure before matching.

## Nearest neighbour matching

### Attainment, post-16 participation

Figure 6: Two-step method (preferred)

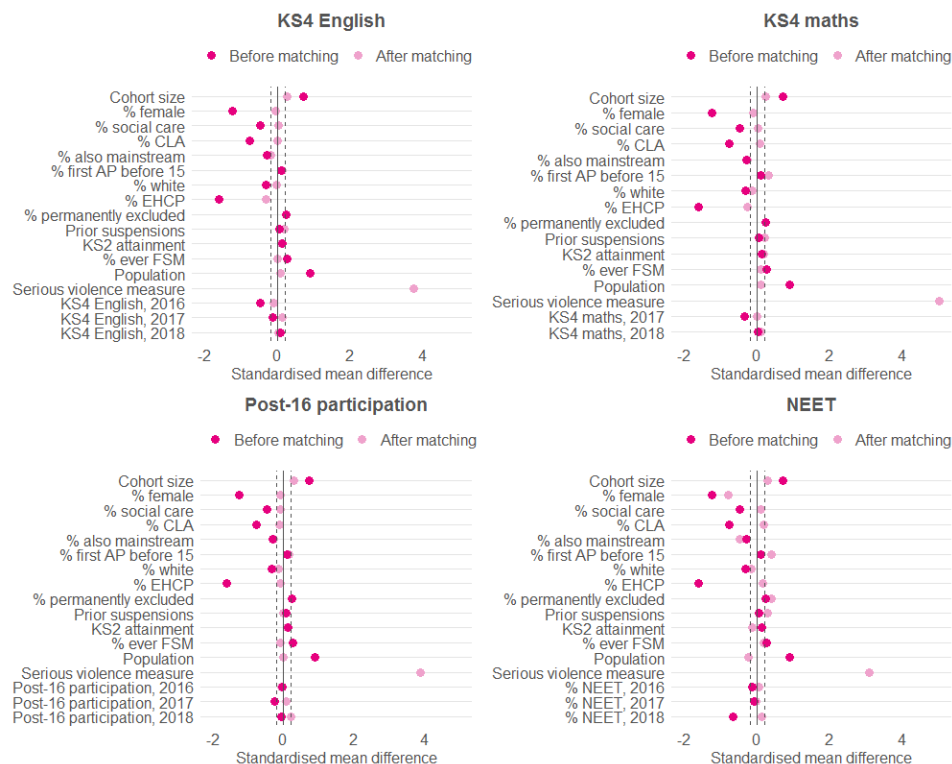
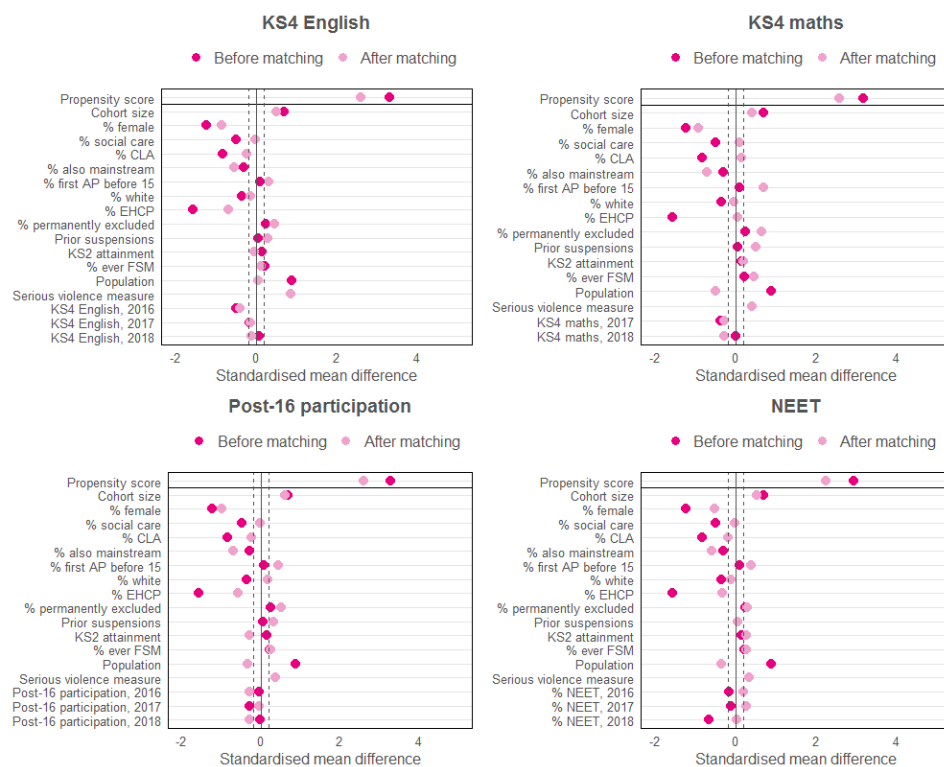
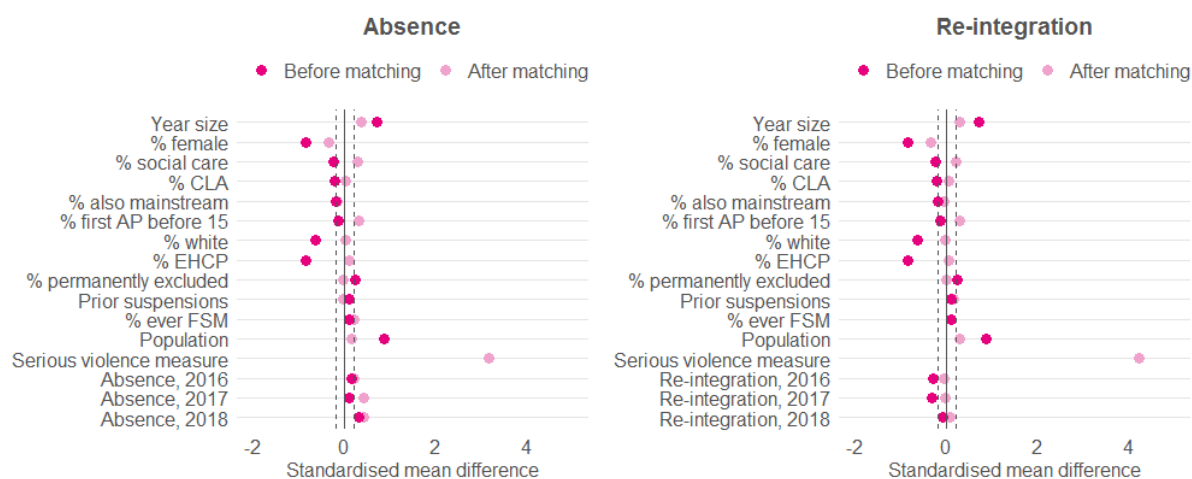


Figure 7: One-step method

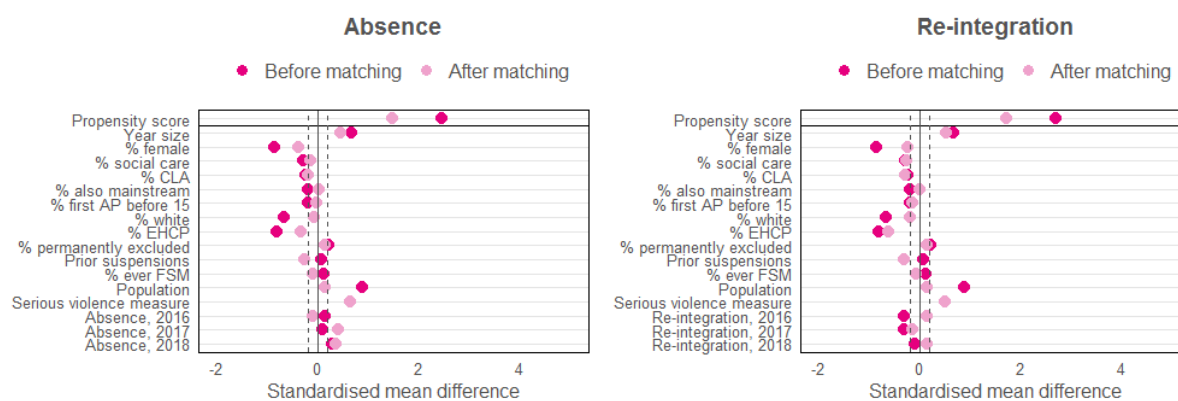


## Re-integration, absence

**Figure 8: Two-step method (preferred)**



**Figure 9: One-step method**





## CBPS weighting

### Attainment, post-16 participation

Figure 10: Two-step method (preferred)

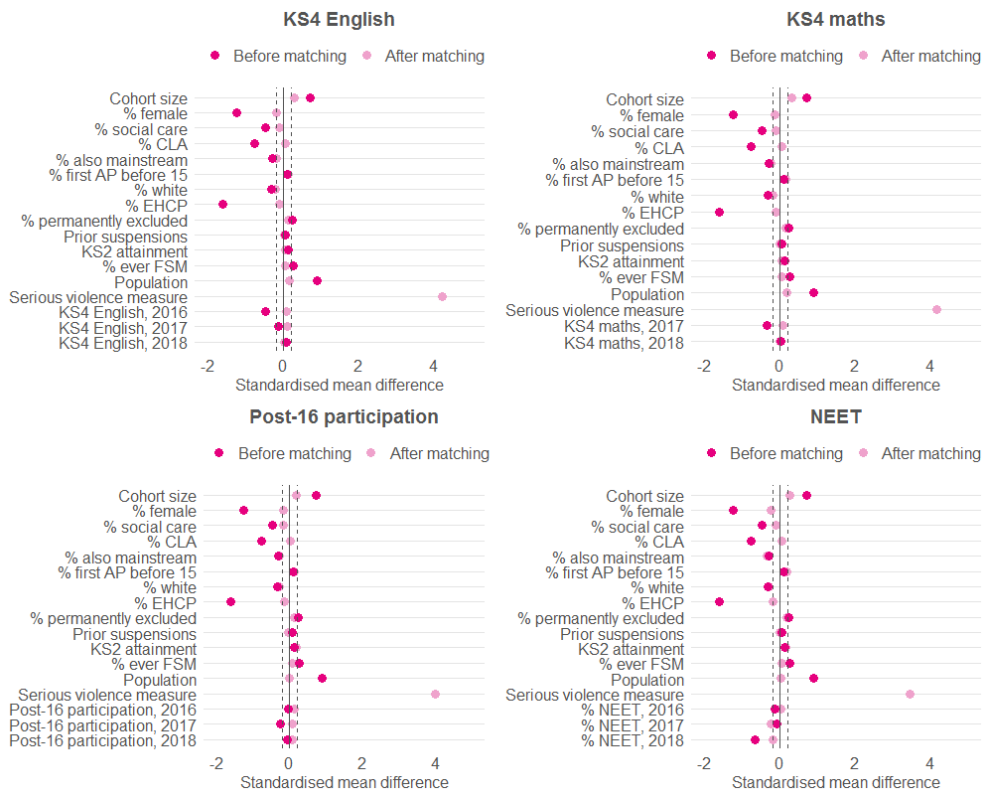
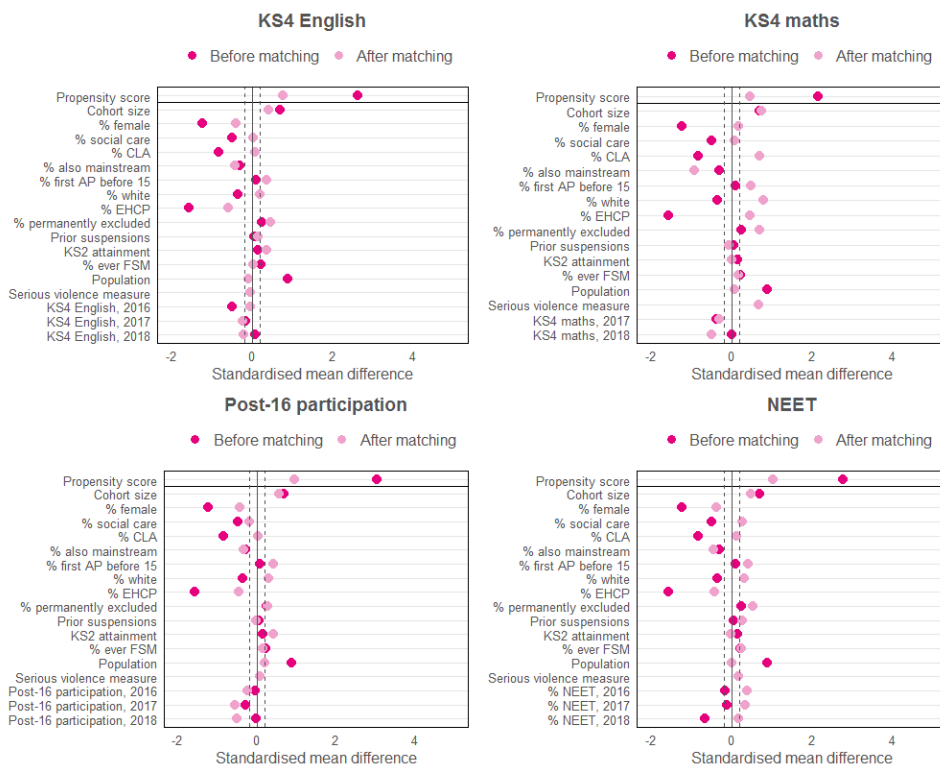
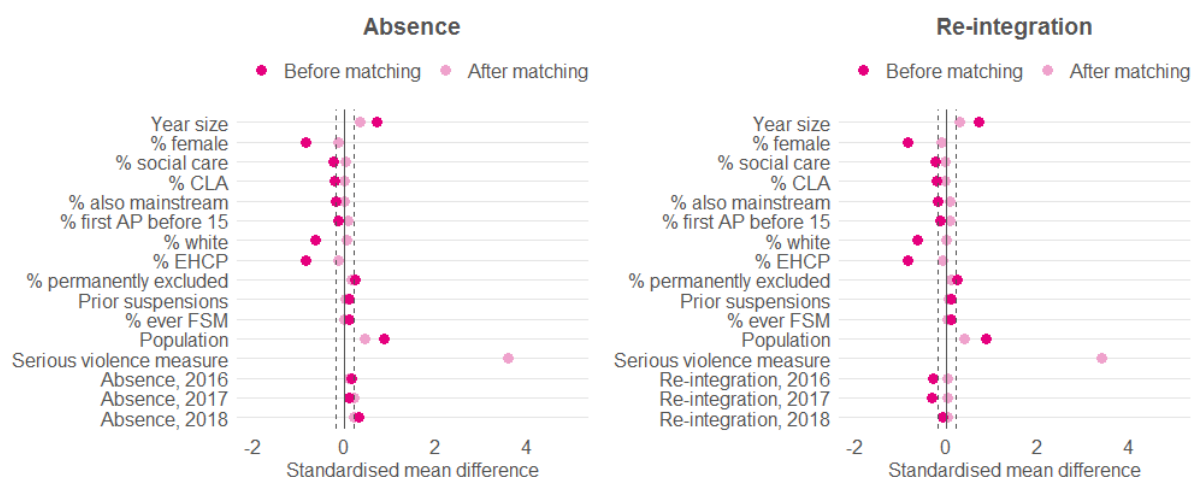


Figure 11: One-step method

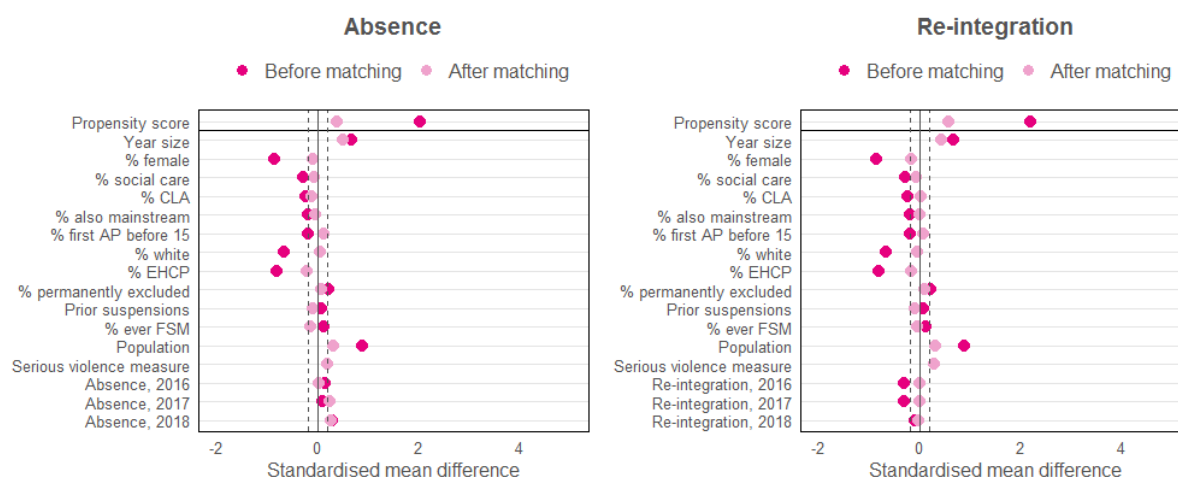


## Re-integration, absence

**Figure 12: Two-step method (preferred)**



**Figure 13: One-step method**



## Addendum 2 to Appendix A: Data tables

Please find attached in the Excel tables.

## **Appendix B - Review Process**

We consider that this study is a **prospective study that requires recruiting comparison schools**:

- It is based on interventions that have not been fully implemented
- Where evaluation activities are planned in intervention and comparison schools which require collecting new data.
- Where comparison schools need to accept taking part of the study.

We understand that the review process is as follows:

- These study plans will be reviewed by YEF and an additional Peer Reviewer before the framework (e.g., matching approach) to select comparison units is used.
- Comparison units have been selected as described in Appendix C.

## Appendix C - Extra information regarding the impact evaluation

### C.1 Identifying Comparison Schools

By the end of September 2021, 21 comparison AP schools<sup>55</sup> in serious violence hotspots have been selected to participate in the Alternative Provision Specialist Taskforce (APST) pilot.

In order to examine the impact of APST on pupil outcomes measured using the Strengths and Difficulties (SDQ) instrument, we identified a set of comparison schools for each participating school.

The process of selecting comparison schools was made difficult by:

- A lack of readily available published data about AP schools and the pupils they serve.
- In addition, published data on pupil numbers excludes subsidiary-registered pupils
- Knowing little (if anything) about how (or even if) outcomes measured by SDQ or SRDS differ between participating and other schools prior to the start of APST
- Heterogeneity among AP schools in the needs of pupils they serve
- By definition, participating and non-participating AP schools differ according to the serious violence measured used to select schools for the pilot.

In a separate project, FFT Education had undertaken research on alternative provision using the National Pupil Database (NPD). As a result of this, a range of measures for each AP school in England were available<sup>56</sup>. These measures can be used to find AP schools that are similar to the participating schools. However, there are some caveats:

- The data used covered the period up to the end of 2018/19. Some of this may be out of date.
- Some schools may have closed, merged or opened since the data were created. These will be out-of-scope.
- The measures were created for a different purpose; they may not be the most relevant measures for this purpose.

In addition, we use the serious violence measures used to select local authorities to participate in the programme. However, whereas counts were used by DfE for this purpose

---

<sup>55</sup> An additional school was recruited in January 2022

<sup>56</sup> <https://ffteducationdatalab.org.uk/2021/09/investigating-alternative-provision-part-one/>

(irrespective of the population size of the local area), we convert them into rates per 10,000 in the population.

The measures used are:

**Table 20: Measures used to select local authorities to participate in the programme.**

<b>All_pupils</b>	Total number of pupils registered at the setting in the 2018/19 academic year (including subsidiary registered pupils)
<b>Pct_ks4</b>	The percentage of all_pupils in 2018/19 who were aged 14 or 15 on 31/08/2018
<b>Pct_main</b>	The percentage of all_pupils in 2018/19 who had a registration type of 'current' or 'main'
<b>Pct_perm_ex</b>	The percentage of all_pupils in 2018/19 who had been permanently excluded prior to joining the AP school
<b>Abs_average</b>	3-year average absence rate at the setting 2017 to 2019
<b>Post16</b>	The percentage of pupils who completed KS4 at the setting in 2017/18 who were observed in education in Autumn 2018
<b>Hospital_rate</b>	Number of hospital admissions for assaults with sharp objects per 10,000 in the population
<b>SV_rate<sup>57</sup></b>	Number of serious violence offences in 2020 per 10,000 in the population

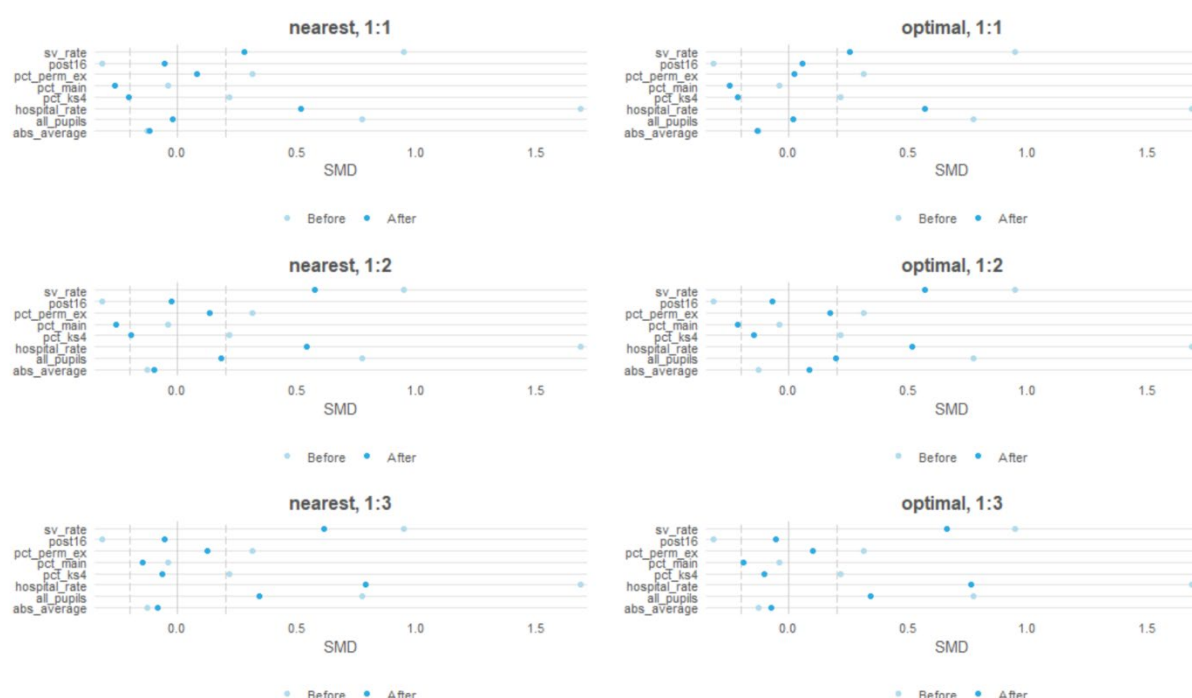
Three AP schools lacked the Post16 variable. In two cases because of small numbers of Key Stage 4 pupils. In the other case, the school was relatively new. Consequently, we ran a matching process in two stages. In stage 1, we matched schools with complete data. In stage 2 we matched the remaining schools.

We used the MatchIt package in R to match schools. We tried both nearest neighbour and optimal matching (without replacement in both cases). The Love plot below shows the balance in observed characteristics before and after matching, excluding the serious violence measure. The matching improves the balance between participants and non-participants but not perfectly. Ideally all measures would be between -0.2 and 0.2 SMD (standardized mean differences). Nearest neighbour matching tends to work a little better.

---

<sup>57</sup> Not available for the Greater Manchester area. In the matching I use the rate for Birmingham for all local authorities in Greater Manchester.

**Figure 14: Love plot showing balance in observed characteristics between APST and comparison group schools**



For both methods we tested 1:1, 1:2 and 1:3 matching. The latter was preferred despite balance tending to be slightly worse in order to increase the chances of recruiting at least one comparison school for each participant.

In the event, one of the three matches (using NN matching) was recruited for 15 of the participants.

In 6 cases reserve matches had to be recruited from the pool of hitherto unmatched schools. To do this, we re-ran the matching algorithm on a reduced dataset containing the 6 participating schools plus all other AP schools that had not already been selected as one of three comparison schools for another participating school.

## C.2 Issues with absence data in NPD

In the preliminary analysis, we noted two problems with absence data.

The first is that not all pupils appear in the absence data for the following year. This may occur if they emigrate, move into home education or die, for instance. Consequently, they have zero sessions of possible attendance.

Secondly, there is variation among the group of pupils for whom absence is observed in the total number of sessions recorded. Selected percentiles of the distribution are shown in Table 21 below. While half of pupils have between 297 and 319 possible sessions, 12.5% have no sessions whatsoever. A further 10% have between 1 and 196 sessions. 2% of pupils have 400 or more sessions. These are likely to be error (e.g., multiple pupils being allocated the same PMR or dual-registered pupils being counted in the absence data of 2 schools simultaneously).

**Table 21: Selected percentiles of the distribution of student population.**

Year	Percentile							Pupils
	5	10	25	50	75	90	95	
2014	0	0	196	302	309	312	332	12191
2015	0	0	216	308	314	318	334	12719
2016	0	0	246	306	310	314	330	13408
2017	0	0	242	298	305	310	318	14166
2018	0	0	244	302	308	310	318	14615
All years	0	0	232	302	310	314	324	67099

In models with absence as an outcome, there are two possible approaches to follow. The first is that we use the unadjusted absence rate for each pupil. So, if a pupil is only observed to be on role for one session and they were absent for that session, they would have an absence rate of 100%. However, we also try a weighted approach where each pupil is weighted using the following formula:

$$\frac{\min(300, \text{possible\_sessions})}{300}$$

### C.3 Points scores used in attainment measures

**Table 22: Points scores used in attainment measures**

Qualification	Grade	Points
GCE AS level	A	67.5
GCE AS level	B	60
GCE AS level	C	52.5
GCE AS level	D	45
GCE AS level	E	37.5
GCE AS level	X	0
GCE AS level	U	0
Basic Skill at Level 1	P	12.5
Basic Skill at Level 1	X	0
Basic Skill at Level 1	U	0
Basic Skill at Level 2	P	23

Basic Skill at Level 2	Q	0
Basic Skill at Level 2	U	0
Basic Skill at Level 2	X	0
Functional Skill at Entry Level	3	7
Functional Skill at Entry Level	2	6
Functional Skill at Entry Level	1	5
Functional Skill at Entry Level	X	0
Functional Skill at Entry Level	U	0
Functional Skill at Entry Level	F	0
Functional Skill at Level 1	P	12.5
Functional Skill at Level 1	Q	0
Functional Skill at Level 1	X	0
Functional Skill at Level 1	F	0
Functional Skill at Level 1	U	0
Functional Skill at Level 2	P	23
Functional Skill at Level 2	Q	0
Functional Skill at Level 2	U	0
Functional Skill at Level 2	F	0
Functional Skill at Level 2	X	0
GCSE Full Course	*	58
GCSE Full Course	A	52
GCSE Full Course	B	46
GCSE Full Course	C	40
GCSE Full Course	D	34
GCSE Full Course	E	28
GCSE Full Course	F	22
GCSE Full Course	G	16
GCSE Full Course	X	0
GCSE Full Course	Q	0



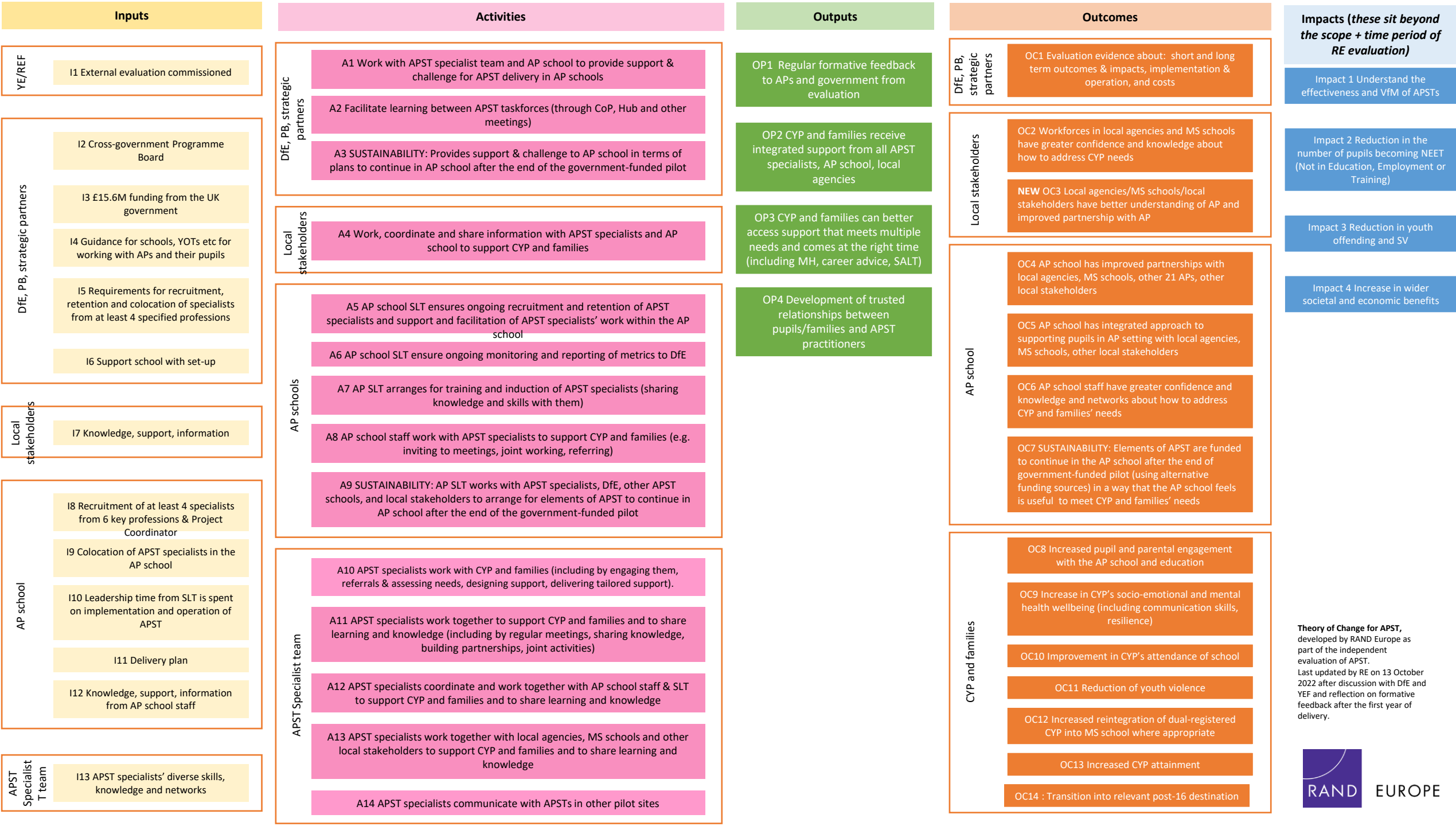
GCSE Full Course	U	0
ELQ D/M/P grading A	3D	1.85
ELQ D/M/P grading A	3M	1.75
ELQ Band A	3	3.5
ELQ Band A	2	3
ELQ Band A	1	2.5
ELQ Band A	X	0
ELQ Band B	3	7
ELQ Band B	2	6
ELQ Band B	1	5
ELQ Band B	U	0
ELQ Band B	Q	0
ELQ Band B	X	0
ELQ Band C	3	14
ELQ Band C	2	12
ELQ Band C	1	10
ELQ Band C	X	0
ELQ Band C	F	0
ELQ Band C	U	0
ELQ Band C	Q	0
ELQ Band D	3	28
ELQ Band D	2	24
ELQ Band D	1	20
Key Skill at Level 1	P	18.75
GCSE (9-1) Full Course	9	60
GCSE (9-1) Full Course	8	55
GCSE (9-1) Full Course	7	51
GCSE (9-1) Full Course	6	47
GCSE (9-1) Full Course	5	43

GCSE (9-1) Full Course	4	39
GCSE (9-1) Full Course	3	34
GCSE (9-1) Full Course	2	28
GCSE (9-1) Full Course	1	19
GCSE (9-1) Full Course	U	0
GCSE (9-1) Full Course	X	0
GCSE (9-1) Full Course	Q	0
Edexcel Certificates	*	58
Edexcel Certificates	A	52
Edexcel Certificates	B	46
Edexcel Certificates	C	40
Edexcel Certificates	D	34
Edexcel Certificates	E	28
Edexcel Certificates	F	22
Edexcel Certificates	G	16
Edexcel Certificates	X	0
Edexcel Certificates	U	0
Level1/2 certificates	*	58
Level1/2 certificates	A	52
Level1/2 certificates	B	46
Level1/2 certificates	C	40
Level1/2 certificates	D	34
Level1/2 certificates	E	28
Level1/2 certificates	F	22
Level1/2 certificates	G	16
Level1/2 certificates	X	0
Level1/2 certificates	Q	0
Level1/2 certificates	U	0
International GCSEs (interim)	X	58

International GCSEs (interim)	*	58
International GCSEs (interim)	U	58
International GCSEs (interim)	A	52
International GCSEs (interim)	B	46
International GCSEs (interim)	C	40
International GCSEs (interim)	D	34
International GCSEs (interim)	E	28
International GCSEs (interim)	F	22
International GCSEs (interim)	G	16
International GCSEs (interim)	Q	0
Graded Drama Music Lit Speech	7M	29.5
Graded Drama Music Lit Speech	6D	27
Graded Drama Music Lit Speech	5D	18
Graded Drama Music Lit Speech	4D	15
Graded Drama Music Lit Speech	2M	5.5
Graded Drama Music Lit Speech	3P	5
Grade 1 Drama Music Lit Speech	M	8
Grade 2 Drama Music Lit Speech	D	14
Grade 2 Drama Music Lit Speech	M	11
Grade 3 Drama Music Lit Speech	M	14
Grade 3 Drama Music Lit Speech	P	10
Grade 5 Drama Music Lit Speech	D	36
Grade 5 Drama Music Lit Speech	M	32
Grade 6 Drama Music Lit Speech	D	54
Grade 6 Drama Music Lit Speech	M	45
Grade 7 Drama Music Lit Speech	P	43
Grade 8 Drama Music Lit Speech	D	86
Cambridge International Certificate Level 1/Level 2	*	58
Cambridge International Certificate Level 1/Level 2	A	52

Cambridge International Certificate Level 1/Level 2	B	46
Cambridge International Certificate Level 1/Level 2	C	40
Cambridge International Certificate Level 1/Level 2	D	34
Cambridge International Certificate Level 1/Level 2	E	28
Cambridge International Certificate Level 1/Level 2	F	22
Cambridge International Certificate Level 1/Level 2	G	16
Cambridge International Certificate Level 1/Level 2	X	0
Cambridge International Certificate Level 1/Level 2	Q	0
Cambridge International Certificate Level 1/Level 2	U	0
Cambridge International Level 1/Level 2 (9-1) Certificate	7	51
Cambridge International Level 1/Level 2 (9-1) Certificate	6	47
Cambridge International Level 1/Level 2 (9-1) Certificate	5	43
Cambridge International Level 1/Level 2 (9-1) Certificate	4	39
Cambridge International Level 1/Level 2 (9-1) Certificate	3	34
Cambridge International Level 1/Level 2 (9-1) Certificate	2	28
Cambridge International Level 1/Level 2 (9-1) Certificate	1	19
Cambridge International Level 1/Level 2 (9-1) Certificate	X	0
Cambridge International Level 1/Level 2 (9-1) Certificate	U	0

## Appendix D: Theory of Change





[youthendowmentfund.org.uk](https://youthendowmentfund.org.uk)



[hello@youthendowmentfund.org.uk](mailto:hello@youthendowmentfund.org.uk)



[@YouthEndowFund](https://twitter.com/YouthEndowFund)

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413

---