EVALUATION PROTOCOL AND STATISTICAL ANALYSIS PLAN

Extension Evaluation of Alternative Provision Specialist Taskforces (APST)

RAND Europe, University of Westminster, FFT Education Datalab

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

YOUTH ENDOWMENT FUND

Extension evaluation of the Alternative Provision Specialist Taskforce (APST)

Addendum to the Statistical Analysis Plan and Protocol for the Evaluation of APST

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

Principal investigator(s): Dr Emma Disley

Project title	*Extension Evaluation of Alternative Provision Specialist Taskforces (APST)*
Developer (Institution)	*Department for Education (DfE)*
Evaluator (Institution)	*RAND Europe, University of Westminster, FFT Education Datalab*
Principal investigator(s)	*Dr Emma Disley* *Impact leads: Prof Richard Dorsett and Dave Thomson*
Protocol/SAP author(s)	*Natalie Picken, Dr Emma Disley, Prof Richard Dorsett, Dave Thomson, Dr James Merewood, Dr Kankan Zhang*
Trial design	*Quasi-experimental design (QED) using difference-in-differences method.*
Trial type	*Efficacy*
Evaluation setting	*State-funded alternative provision (AP) schools*
Target group	*Pupils in Year 7 to Year 11 attending state-funded AP schools*
Number of participants	*2,000 to 2,500 pupils per year in 22 AP schools*

Primary outcome and data source	***For pupils in Year 7-9:*** *Re-integration into a mainstream school after one year (National Pupil Database - NPD)* ***For pupils in Year 10:*** *Initial post-16 participation after two years (NPD + Individual Learner Record – ILR)* ***For pupils in Year 11****: Initial post-16 participation after one year (NPD + ILR)*
Secondary outcome and data source	***For pupils in Year 7-11:*** *Participation in education (NPD + ILR)* ***For pupils in Year 7-10:*** *Attendance (NPD)* ***For pupils in Year 7-9:*** *Attendance (NPD)* ***For pupils in Year 7-10:*** *Re-integration into a mainstream school (NPD)* ***For pupils in Year 10:*** *Re-integration into a mainstream school (NPD)* ***For pupils in Year 7-9:*** *Re-integration into a mainstream school (NPD)* ***For pupils in Year 10 and 11:*** *Key Stage 4 attainment in English and maths (NPD)* ***For pupils in Year 11:*** *Initial and Sustained post-16 participation (NPD + ILR)*

Table of Contents

1. Version

VERSION	DATE	REASON FOR REVISION
v.7.0	**20 December 2023**	**Revision following further feedback from YEF peer review.**
v.6.0	1 December 2023	Revision following feedback from YEF peer review.
v.5.0	30 October 2023	Revision following feedback from YEF and DfE.
v.4.0	25 September 2023	Revision following kick off meeting and preliminary analysis stage.
v.3.0	10 May 2023	Revision following feedback from Department for Education.
v.2.0	20 March 2023	Revision following feedback on original version from YEF.
v.1.0 [original]	03 March 2023	Original.

2. Background

In summer 2021, RAND Europe in consortium with University of Westminster (UoW) and FFT Datalab (FFT) was commissioned to carry out an independent evaluation of the delivery of the Alternative Provision Specialist Taskforces (APST) pilot between November 2021 and August 2023. This evaluation is funded by the Youth Endowment Fund (YEF). A statistical analysis plan (SAP), accessible on the YEF website[1], includes further information about this evaluation.

In January 2023, the DfE extended the delivery of the APST pilot by eighteen months. The APST pilot will now end in March 2025. The rationale for DfE extending the delivery of the APST programme, is that it would enable:

- At least 2 full academic years of full-scale delivery and impact evaluation;
- Better staff retention of specialists to maximise the integrity of the model over the evaluation term;
- A firm evidence-base to inform scale-up and AP reform;
- Limiting bias in the evaluation.

In June 2023, RAND Europe and the consortium was commissioned to carry out an evaluation of the first twelve months of the extended delivery period: covering the time between September 2023 and August 2024. The rationale for commissioning this evaluation was that an extension to the impact and process evaluation would provide the opportunity for:

- **The cohorts included in the current evaluation to be followed-up for a longer period of time,** thus exploring the extent to which any outcomes are sustained.
- **An additional cohort of pupils to be added to the evaluation.** This increases the ability to detect impact through greater sample size.
- **The impact of operating APST over a longer period of time to be measured.** Delays to the of implementation of APST mean that the current evaluation will only examine the impact of just over one full year of delivery (rather than two years as initially anticipated).
- **Greater understanding of how the APST model is embedded and evolves.** An additional year of process evaluation offers potential for learning lessons about how best to implement a multi-agency and child-centred approach, generating deeper insight into the APST model. This will be useful to inform future policy decisions and guide good practice around APST.

1 https://youthendowmentfund.org.uk/funding/who-we-fund/alternative-provision-specialist-taskforces-apst-department-for-education/

This document outlines the plans for the extension evaluation of APST. In this document we use the following terms:

- "**Current evaluation**" is used to describe the evaluation of APST as delivered between September 2021 and August 2023. This evaluation is covered by the current grant agreement between YEF and RAND and described in the SAP and protocol for the current evaluation. This is not the focus of this document.
- "**Extension evaluation**" is used to describe the evaluation of APST as delivered between September 2023 and August 2024. This is the focus of this document.

This document should be read in conjunction with the SAP and protocol for the current evaluation which may be found here: https://youthendowmentfund.org.uk/funding/who-we-fund/alternative-provision-specialist-taskforces-apst-department-for-education/.

3. Extension impact evaluation

This section uses the same subheadings as the SAP and protocol for the current evaluation. Information is included only when the approach is different to the approach taken to the current evaluation.[2]

3.1. Context: summary of the approach to impact evaluation in the current evaluation

A summary of the sample, outcomes, follow-up periods and cohorts in the current evaluation are presented in Table 1. For further information, please see the SAP and protocol for the current evaluation.

Table 1: Summary of the sample, outcomes, follow-up periods and cohorts for the current evaluation

Sample	**The current impact evaluation includes:** – **Cohort 1: All children in Years 7 to 11 in the 22 participating AP schools between September 2021 and August 2022** – **Cohort 2: All children in Years 7 to 11 in the 22 participating AP schools between September 2022 and August 2023**

[2] For more information about the current evaluation, please see the SAP and protocol for the current evaluation: https://youthendowmentfund.org.uk/funding/who-we-fund/alternative-provision-specialist-taskforces-apst-department-for-education/.

Outcomes and follow up periods	The current impact evaluation intends to assess the impact of attending a school with APST on students': – Initial post-16 destinations, re-integration into mainstream school, attendance for autumn and spring terms, KS4 attainment, participation in education **one year after** attending the school for **Cohorts 1 and 2.** – Sustained post-16 destinations and attendance for the whole year **one year after** attending the AP school for **Cohort 1.** – Social and emotional outcomes and conduct and hyperactivity **before and after** attending the school for **Cohorts 1 and 2.**

3.2. Summary of the differences in approach between the current and extension impact evaluations

A summary of the sample, outcomes, follow-up periods and cohorts in the extension evaluation are presented in Table 2. The extension evaluation will be conducted using the approach outlined in the SAP and protocol for the current evaluation with the following modifications:

- **The primary outcome for students in Year 10 will be different to the current evaluation:**
    - In the current evaluation, the primary outcomes are initial post-16 destinations after one year (Year 11) and re-integration into mainstream schools after one year (Years 7-10).
    - In the extension evaluation, the amended primary outcomes are initial post-16 destinations after one year (Year 11), **initial post-16 destinations after two years (Year 10)**, and re-integration into mainstream school after one year (Years 7-9).
    - See section 3.7 for further information.
- **Outcomes for students two years after attending the school will be explored (where data are available).** In the current evaluation, only outcomes for students one year after attending the school are included. See sections 3.7 and 3.8 for more information.
- **Only outcomes measured using administrative data sources will be included.** In the current evaluation, outcomes based on the Strengths and Difficulties Questionnaire (SDQ) are included. In the extension evaluation, outcomes based on the SDQ will not be included for any of the cohorts due to the high resource involved and high level of attrition. See sections 3.7 and 3.8 for more information.
- **An additional subgroup analysis of AP schools who continue APST for a third year will be carried out for the primary outcomes.** This is to examine if impacts

materialise after a longer period of implementation. See sections 3.7, 3.8, and 3.14 for more information.

Table 2: The sample, outcomes, follow-up periods and cohorts for the extension evaluation

Sample	**The extension impact evaluation includes:** – **Cohort 1 and Cohort 2 (as in Table 1)** – **Cohort 3: all children in Years 7 to 11 in the 22 participating AP schools between September 2023 and August 2024**
Outcomes and follow up period	– The extension impact evaluation intends to assess the impact of attending a school with APST on students': ○ Initial post-16 destinations, re-integration into mainstream school, attendance for autumn and spring terms, KS4 attainment, and participation in education **one year after** attending the school for **Cohort 3**. ○ Sustained post-16 destinations and attendance for the whole year **one year after** attending the school for **Cohort 2**. ○ Initial post-16 participation, re-integration into mainstream school, attendance in autumn and spring term, KS4 attainment **two years after**[3] attending the school for **Cohorts 1 and 2**.

3.3. Research questions and objectives for the extension evaluation

The extension evaluation will explore 5 impact evaluation questions, organised by pupil outcomes (Table 3). Under each evaluation question, where applicable, we set out sub-questions based on the year group of pupils. We do not set out specific research questions for different cohorts using the same outcome indicator.

We outline below the changes between the evaluation questions for the extension evaluation and those set out in the SAP and protocol for the current evaluation.

- We amended the wording of the evaluation question about re-integration (EQ1 in Table 3) to reflect that reintegration is a primary outcome for Year 7-9 but not Year 10 pupils.
- We added one question to explore the initial post-16 destinations of Year 10 pupils two years after intervention (EQ2a), as this is proposed as the new primary outcome for this group.

[3] This means outcomes achieved two academic years after intervention. For example, for pupils in Year 11 at the time of intervention, this means their outcomes in Year 13.

- We added one question to explore the initial post-16 destinations for Year 11 pupils two years after intervention (EQ2d), as this is materially different to the outcomes examined in the current evaluation.
- We added one question to explore the attendance of pupils in Year 7 to 9 (EQ3b), to reflect the availability of outcome data for pupils two years after intervention.
- We amended the wording of the evaluation question about attainment (EQ4) to reflect the availability of KS4 attainment outcomes data for Year 10 pupils two years after intervention.
- We removed EQ19 and EQ20 from the current evaluation, as these relate to the Strengths and Difficulties Questionnaire (SDQ) that will not be pursued in the extension evaluation.

Table 3: Updated research questions for the extension impact evaluation[4]

EQ1[5] Re-integration	Pupils in Year 7-9	a) What is the difference in re-integration of Key Stage 3 (Year 7 to 9) pupils one year after intervention in the 22 participating AP schools in comparison to those pupils in comparison schools receiving business as usual?
	Pupils in Year 7-9	b) What is the difference in re-integration of Key Stage 3 (Year 7 to 9) pupils two years after intervention in the 22 participating AP schools in comparison to those pupils in comparison schools receiving business as usual?
	Pupils in Year 10	c) What is the difference in re-integration of Year 10 pupils one year after intervention in the 22 participating AP schools in comparison to those pupils in comparison schools receiving business as usual?
	Pupils in Year 7-10	d) What is the difference in re-integration of pupils in Year 7 to 10 pupils one year after intervention in the 22 participating AP schools in comparison to those pupils in comparison schools receiving business as usual?
EQ2[6] Post-16 outcomes	Pupils in Year 10	a) What is the difference in post-16 outcomes measured by initial destinations in Year 1 of post-16 (i.e. Year 12 or equivalent) of Year 10 pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?
	Pupils in Year 11	b) What is the difference in post-16 outcomes measured by initial destinations in Year 1 of post-16 (i.e. Year 12 or equivalent) of Year 11 pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?

[4] EQ19 and EQ20 as set out in the SAP for the current evaluation have been omitted as these relate to outcomes measured by the Strengths and Difficulties Questionnaire (SDQ).

[5] EQ1a is amended based on EQ15 in the SAP for the current evaluation. We have updated the wording of the question to reflect that re-integration is a primary outcome for Year 7-9 but not Year 10 pupils. EQ1c and EQ1d are the same as EQ15 as set out in the SAP for the current evaluation.

[6] EQ2a is a new evaluation question in the SAP for the extension evaluation. This question has been added to ensure that we investigate the new primary outcome for Year 10 students: the post-16 outcomes of Year 10 pupils. EQ2b and EQ2c are the same as EQ17 as set out in the SAP for the current evaluation. EQ2d is a new question for the extension evaluation to explore the initial post-16 destinations for Year 11 pupils two years after intervention. This has been added as this is materially different to the outcomes examined in the current evaluation.

		c) What is the difference in post-16 outcomes, measured by sustained destinations in Year 1 of post-16 (i.e. Year 12 or equivalent) of Year 11 pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?
		d) What is the difference in post-16 outcomes measured by initial destinations in Year 2 of post-16 (i.e. Year 13 or equivalent) of Year 11 pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?
EQ3[7] Attendance	Pupils in Year 7-10	a) What is the difference in attendance of pupils in Year 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?
	Pupils in Year 7-9	b) What is the difference in attendance of pupils in Year 7 to 9 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?
EQ4[8] Attainment	Pupils in Year 10 and 11	What is the difference in attainment, measured separately by Key Stage 4 Attainment, in English and Maths of Year 10 and 11 pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?
EQ5[9] Participation	Pupils in Year 7-11	What is the difference in participation in state-funded education of Year 7 to 11 pupils in schools receiving APST in comparison to those pupils in comparison schools receiving business as usual?

[7] EQ3a is the same as EQ16 as set out in the SAP for the current evaluation. EQ3b is a new question for the extension evaluation to reflect the availability of outcome data for pupils two years after intervention.

[8] EQ4 is amended based on EQ18 in the SAP for the current evaluation, to reflect the availability of KS4 attainment outcomes data for Year 10 pupils two years after intervention.

[9] This EQ is the same as EQ21 as set out in the SAP for the current evaluation.

3.4. Research design overview for the extension evaluation

The extension evaluation will follow the design used for the current evaluation and outlined in the SAP and protocol for the current evaluation: a quasi-experimental impact evaluation using administrative data from the National Pupil Database (NPD) linked to the Individualised Learner Record (ILR). Estimation will use a difference-in-differences methodology and an intention-to-treat design. All 22 AP schools in the current evaluation will remain in scope for the extension evaluation.

3.5. Participants

The intervention sample for the extension evaluation will be:

- All pupils in Years 7 to 11 in the 22 APs in the academic year September 2021-August 2022 (Cohort 1)
- All pupils in Years 7 to 11 in the 22 APs in the academic year September 2022-August 2023 (Cohort 2)
- All pupils in Years 7 to 11 in the 22 APs in the academic year September 2023-August 2024 (Cohort 3)

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 exposed to the intervention. The cohorts will include all pupils who are enrolled at AP schools in each year regardless of whether they are single or dual-main registered[10] and regardless of the length of time they are enrolled.

3.6. Outcome measure and other data

The primary and secondary outcomes that will be evaluated for each cohort in the extension evaluation are listed in Table 4 and Table 5. In these tables, **blue cells** indicate outcomes that will be included in the extension evaluation, **grey cells** indicate outcomes that are included in the current evaluation only, and **white cells** indicate outcomes that are not included in either evaluation.

[10] 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.

Table 4: Primary outcomes that will be explored in the extension evaluation

Outcome	Measure	Year Groups included	Cohort 1 (2021/22)	Cohort 2 (2022/23)	Cohort 3 (2023/24)
Re-integration	1 year after intervention	Years 7-9	Y	Y	Y
Post-16 participation	Initial destinations in Year 1 of Post-16 (i.e. Year 12 or equivalent, **1 year after** intervention)	Year 11	Y	Y	Y
Post-16 participation	Initial destinations in Year 1 of Post-16 (i.e. Year 12 or equivalent, **2 years after** intervention)	Year 10	Y	Y	

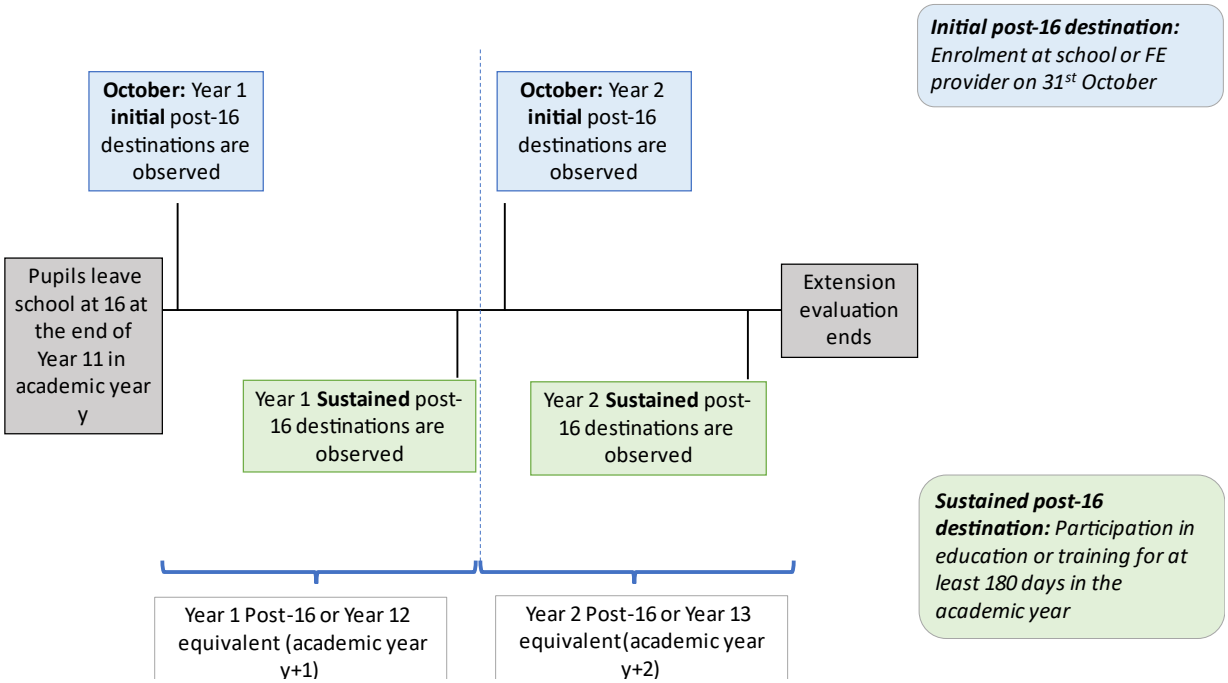
Table 5: Secondary outcomes that will be explored in the extension evaluation

Outcome	Measure	Year Groups included	Cohort 1 (2021/22)	Cohort 2 (2022/23)	Cohort 3 (2023/24)
Post-16 participation	Sustained destinations in Year 1 Post-16 (i.e. Year 12 or equivalent, **1 year after** intervention)	Year 11	Y	Y	
	Initial destinations in Year 2 of Post-16 (i.e. Year 13 or equivalent, **2 years after** intervention)	Year 11	Y	Y	
Re-integration	1 year after intervention	Years 7 to 10	Y	Y	Y
	1 year after intervention	Year 10 only	Y	Y	Y

	2 years after intervention	Years 7 to 9	Y	Y	
Attendance	1 year after intervention (whole year)	Years 7-10	Y	Y	
	1 year after intervention (autumn and spring term)	Years 7-10	Y	Y	Y
	2 years after intervention (autumn and spring term)	Years 7-9	Y	Y	
KS4 Attainment	1 year after intervention	Year 11	Y	Y	Y
	2 years after intervention	Year 10	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	Y	Y

Figure 1 shows the timeline for observing initial and sustained post-16 destinations.

Figure 1: Timeline showing initial and sustained post-16 destinations (Years 1 and 2) and the definition



3.7. Primary outcomes

In the extension evaluation, we are able to observe participants over a longer time period than the current evaluation. Consequently, data on post-16 destinations after two years has become available for pupils who participated in APST when in year 10. This is an outcome that is more directly related to the work of AP schools than re-integration, the primary outcome in the current evaluation: Year 10 pupils are less likely to re-integrate into mainstream school after attending an AP school than younger pupils and so AP schools may focus more on supporting Year 10 pupils into post-16 transition. Re-integration for Year 10 will be retained as a secondary outcome.

Given the availability of further data, we propose using slightly different primary outcomes for the current and extension evaluations:

- In the current evaluation, the primary outcomes are initial post-16 destinations after one year (Year 11) and re-integration into mainstream schools after one year (Year 7-10).
- In the extension evaluation, the primary outcomes will be initial post-16 destinations after one year (Year 11), initial post-16 destinations after two years (Year 10), and re-integration into mainstream school after one year (Year 7-9).

All primary outcomes for the extension evaluation are listed in Table 4.

A definition of the additional primary outcome is provided below. The definitions of other primary outcomes are in the SAP and protocol for the current evaluation.

Initial Post-16 participation for Year 10 in Year 2 (i.e., Year 12 or equivalent)

Measure: Enrolment at a school or FE provider on 31st October two academic years after attending an AP school in Year 10

Population: All pupils observed in School Census on roll at AP schools in England for 1 day or more in Year 10. Pupils observed at multiple AP schools are allocated to the first participating school at which they are observed, or the first non-participating school at which they are observed.

Years available: 2014 to 2025

Definition of the measure: For each pupil observed on roll at an AP school in academic year y, we will scan the School Census for those attending schools and the Individualised Learner Record (ILR) for those in further education (including work-based learning) on 31 October in year y + 2. Those in an education destination will be defined as those either 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 identified in the APST ToC (Impact 2, Annex D of the SAP and protocol for the current evaluation). Previous research conducted by members of the Consortium[11] shows that rates of initial participation among pupils who experience AP are low (fewer than 60% of pupils were observed to be participating among the 2018 cohort, for example).

The Consortium undertook preliminary analysis to examine pre-existing trends between participating and non-participating AP schools. Results are presented in Annex A. Briefly, the pre-existing trends exhibit a degree of volatility (see Table 1 of the Excel tables attached). Consequently, there is a risk that the parallel trends assumption on which the difference-in-difference specification is based does not hold. The consequences of this are discussed in section 3.13).

Effect size calculation

See the corresponding section of the SAP and protocol for the current evaluation.

[11] https://ffteducationdatalab.org.uk/2021/09/investigating-alternative-provision-part-2/

For all NPD outcomes the Consortium use a difference-in-differences specification over multiple years.

Preliminary analysis undertaken by the Consortium to test the amended primary outcome for Year 10 empirically estimated the MDES using the observed standard errors from placebo tests on pre-treatment data.

The table below presents the MDES for the amended primary outcome for Year 10. It is presented for:

a) the main specification, which uses all comparison AP schools.

b) the robustness specification (see the SAP and protocol for the current evaluation).

This mirrors the approach of the current evaluation outlined in the SAP and protocol for the current evaluation. (see Table 5 in the SAP and protocol for the current evaluation).

In Table 6, in both cases the MDES is below 0.2, indicating that the analysis is expected to achieve the power threshold typically required by the EEF for school trials[12]. Further information can be found in Annex A.

Table 6: 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)
Primary outcomes			
Initial post-16 participation	Year 10	0.08	0.12

3.8. Secondary outcomes

Secondary outcomes are described in the SAP and protocol for the current evaluation.

Four additional indicators are included in the extension evaluation. The definitions of these indicators are the same as in the SAP and protocol for the current evaluation but are either calculated over a different time period (after two years rather than after one year) or for a

[12] See Singh et al, 2023, *Improving power calculations in educational trials,* Education Endowment Foundation. Accessed on 30 October 2023: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-evaluation-reports-and-research-papers/methodological-research-and-innovations/improving-power-calculations-in-educational-trials

different population of pupils (Year 10 instead of Year 11). Definitions of these additional outcomes are provided below.

Initial Post-16 participation in Year 2 (i.e., Year 13 or equivalent)

Measure: Enrolment at a school or FE provider on 31st October two years after completing 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, or the first non-participating school at which they are observed.

Years available: 2014 to 2025

Definition of the measure: For each pupil observed on roll at an AP school in year y, we will scan the School Census for those attending schools and the Individualised Learner Record (ILR) for those in further education (including work-based learning) on 31 October in year y + 2. Those in an education destination will be defined as those either 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 identified in the APST ToC (Impact 2, see Annex D of the SAP and protocol for the current evaluation).

Re-Integration into mainstream in Year 2

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 two years after being enrolled at an AP school.

Population: All pupils observed in School Census on roll at AP schools in England for 1 day or more in Years 7 to 9. For each year, pupils who are observed at multiple AP schools are allocated to the first participating AP school at which they are observed, or the first non-participating school at which they are observed. No adjustment is made for pupils observed in multiple years.

Years available: 2014 to 2025

Definition of the measure: For each pupil observed attending an AP school in year y, we scan the School Census and the local authority alternative provision census in year y + 2 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 + 2 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 mainstream schools and less than 180 days in total in alternative provision are considered to have been re-integrated.

Rationale: An increase in re-integration of pupils into mainstream schools is an outcome identified in the APST ToC (OC12).

Attainment in Year 2

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 14 (Year 10). Pupils observed at multiple AP schools are allocated to the first participating school at which they are observed, or the first non-participating school at which they are observed.

Years available: 2014 to 2019, 2022 to 2025

Definition of the measure: As described in the SAP and protocol for the current evaluation for Year 11 pupils.

Rationale: Increased attainment is an outcome identified in the APST ToC. English and maths are 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,[13] consistent with the longer-term aim of APST to reduce propensity to be not in education, employment or training post-16.

Attendance in Year 2

Measure: Rate of attendance two years after being enrolled at an AP school.

Years available: 2014 to 2018, 2020-2025. Due to the COVID-19 pandemic resulting in schools being closed to the majority of state school pupils in the 2019/20 academic year, there is no endline absence data for the 2018 cohort. Data for the 2019 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 9. For each year, any pupils observed at multiple AP schools are allocated to the first participating school at which they are observed, or the first non-participating school at which they are observed. No adjustment is made for pupils observed in multiple years.

[13] https://ffteducationdatalab.org.uk/wp-content/uploads/2021/09/working_paper4.pdf

Definition of the measure: For each pupil observed attending an AP school in year y, we scan absence data for year y + 2. 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 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 APST ToC (OC10). Given the definition of authorised absence differs across schools, overall absence (authorised and unauthorised absence combined) is used as the measurement to ensure certainty. However, the Consortium have the following concerns about the quality of absence data in NPD (these are summarised in Appendix C of the published SAP and protocol for the current evaluation):

- There is the uneven impact of COVID-19 on absence, with London appearing to suffer less impact compared to other regions.
- Not all pupils appear in the absence data for the following year.
- There is variation among the group of pupils for whom absence is recorded.

3.9. Administrative datasets

Additional years of the datasets listed in Table 10 of the published SAP and protocol for the current evaluation will be included but no additional datasets are required.

3.10. Selection of the comparison group and identification assumptions

See the corresponding section of the published SAP and protocol for the current evaluation.

3.11. Primary analysis

See the corresponding section of the published SAP and protocol for the current evaluation.

3.12. Inference

See the corresponding section of the published SAP and protocol for the current evaluation.

3.13. Robustness checks

See the corresponding section of the published SAP and protocol for the current evaluation.

In addition, to assess sensitivity of the estimates to violation of the parallel trends assumption, the Consortium will make use of a recently-introduced development in the difference-in-difference literature. This literature provides an assessment of how big the failure of parallel trends would have to be in order to overturn a significant finding, allowing the Consortium to test how robust the findings of the analysis are to violations of the parallel trends assumption, where these violations exist (Annex A).

3.14. Subgroup analyses

See the corresponding section of the published SAP and protocol for the current evaluation.

In addition, the Consortium will interact the treatment effects for the primary outcomes with the following variable:

- School that continued to be funded to deliver APST in 2023/24 (yes/ no)

It may be necessary to arrive at some sort of assessment of sufficiency of funding for schools in 2023/24 compared to previous years in order to derive this variable.

3.15. Treatment effects in the presence of non-compliance

See the corresponding section of the published SAP and protocol for the current evaluation.

3.16. Missing data

See the corresponding section of the published SAP and protocol for the current evaluation.

4. Extension process evaluation

This section uses the same subheadings as the SAP and protocol for the current evaluation. Information is included only when the approach is different to the approach taken to the current evaluation.[14]

4.1. Research questions

The extension evaluation will explore 11 process evaluation questions organised within four objectives. These can be found in Table 7.

Throughout the analysis, the Consortium will consider any differences from the findings of the current evaluation.

The process evaluation questions for the extension are based on those used in the current evaluation (see Table 13 in the SAP and protocol for the current evaluation). A few alterations have been made to remove, refine, and add questions to reflect learning from the current evaluation.

Table 7: Proposed extension evaluation research questions (EQ6-15)

Objective 1	**Operation of APST:** *what APST "looked like" in schools, the barriers and facilitators experienced, and the differences between schools.*
EQ6	How do APST specialists work with children and young people and families as part of APST on a day-to-day basis? How does this differ between APST schools? How does this differ from APST delivery in academic years 2021-2 and 2022-23?
EQ7	What were the barriers and facilitators that affected operation of APST as planned? What, if any, adaptations were made to the operation of APST as planned?
EQ8	To what extent do APSTs account for and respond to diversity in students' ethnicities and genders?
EQ9	To what extent do AP schools have plans to continue delivering elements of APST after the end of the DfE-funded period (March 2025)? What are the barriers and facilitators to sustainability?

[14] For more information about the current evaluation, please see the SAP and protocol for the current evaluation: https://youthendowmentfund.org.uk/funding/who-we-fund/alternative-provision-specialist-taskforces-apst-department-for-education/.

Objective 2	**Partnership working:** *we will test the hypothesis that partnership working will lead to better support for students and families by trying to understand what partnership working involved, whether and how it made a difference, and what the barriers and facilitators were.*
EQ10	To what extent did APST specialists work in partnership with each other, with the wider AP school, and with local agencies? To what extent did this partnership working make a difference? What were the barriers and facilitators to partnership working? What adaptations were made?
Objective 3	**Outcomes**[15]**:** *we will explore stakeholders' perceptions of the impact that APST is likely to achieve on the outcomes listed in the theory of change.*
EQ11	To what extent and how was APST perceived by relevant stakeholders to contribute to the stated outcomes?[16]
EQ12	To what extent and how was 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 APST result in unintended consequences for any stakeholders during the length of the APST pilot?
Objective 4	**Lessons learnt for future policy and practice:** *we understand that the DfE intend to consider how to incorporate aspects of APST into good practice for all APs in future. We propose to identify cross-cutting lessons learnt about APST to help reflections on this point.*
EQ15	What are the transferrable lessons from the APST pilot for policymakers and practitioners?

[15] 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.

[16] These are set out in the Theory of Change, see Annex D to the current SAP & protocol evaluation.

4.2. Research methods

Table 8 outlines the planned research methods to answer the extension process evaluation questions. Further information about the research methods and the approach to conducting and analysing interviews, surveys, documentation and case studies can be found in the Implementation and Process Evaluation section of the SAP and protocol for the current evaluation.

Table 8: Proposed research methods for extension process evaluation

Data source	Research method	Participants / data sources (type, number)	Timing	Rationale for inclusion
DS1	**Interviews**	Up to 3 Strategic Personnel at DfE at each timepoint. To take place virtually	Two timepoints: at the start of the academic year (Sep - Oct 2023) and at the end of the academic year (July - August 2024)	Understand what modifications (if any) the schools that are continuing delivery are intending to make to the APST model.
DS2	**Documentation review**	We will review the finalised sustainability plans provided by all schools continuing APST delivery to the DfE.	Start of academic year 23-24 (Sep - Oct 2023)	Understand what modifications (if any) the schools that are continuing delivery are intending to make to the APST model.
DS3	**Survey**	All SLT and all specialists at all schools continuing delivery. We will operate two surveys.	One time point: March 2024.	Understand operation of APST, barriers & facilitators, partnership working, perception of outcomes, and draw out lessons learnt.
DS4	**Interviews**	With the SLT lead at each APST school continuing delivery (up to 22 interviews). This will take place at one timepoint and be conducted virtually.	One time point: May - June 2024.	Understand operation of APST, barriers & facilitators, partnership working, perception of outcomes, and draw out lessons learnt.
DS5	**Case studies**	Three case studies, each focusing on one AP school that is continuing delivery of APST. Case studies will comprise interviews, document review, and observation.	One time point: May - June 2024.	Understand operation of APST, barriers & facilitators, partnership working, perception of outcomes, and draw out lessons learnt.
DS6	**Attending and observing ad hoc Programme wide meetings**	We have budgeted to attend up to two ad-hoc meetings of the SLT Leads / Project Coordinators in person to observe and gather data.	Ongoing	Understand operation of APST, barriers & facilitators, partnership working, perception of outcomes, and draw out lessons learnt.

4.3. Analysis

The approach to analysis is as described in the SAP and protocol for the current evaluation (Analysis section, Implementation and Process Evaluation).

5. Project management and quality assurance

RAND Europe is Consortium lead with responsibility for liaising with YEF and the DfE throughout the extension evaluation. RAND Europe and FFT and UoW will liaise closely to ensure that impact and process evaluations activities are aligned. The Consortium will build on existing relationships and ways of working in the current evaluation of APST.

The outputs of the evaluation will be subject to the YEF's peer review process. The Consortium will also apply some elements of the RAND Europe quality assurance process:

- The project leader, a peer reviewer within RAND or an expert advisor will review research tools (survey instruments, interview guides, and case study guides) and all outputs (slide decks, draft and final reports)
- All outputs from FFT and UoW will be reviewed by an experienced researcher at RAND Europe as well as by the Project Leader.

Where possible, internal RAND Europe quality assurance process will take place before the final draft of the consolidated report is shared with the YEF or DfE, but if not, this review will take place in parallel.

6. Data protection and ethics

6.1. Data protection

The evaluation will be conducted in compliance with UK GDPR and good practice in data protection. RAND Europe have undertaken an assessment of the data protection considerations at the outset of the project in conjunction with the Data Protection Officer. The approach to data protection will follow that set out in the SAP and protocol for the current evaluation.[17]

For the extension process evaluation: the data roles held in the current evaluation will be maintained with RAND Europe and DfE acting as independent data controllers of personal data (names, contact details, roles of SLT, Specialists, and strategic personnel working at the

[17] https://youthendowmentfund.org.uk/funding/who-we-fund/alternative-provision-specialist-taskforces-apst-department-for-education/.

DfE). The Consortium do not envisage that special category or sensitive data will be gathered the extension process evaluation.

For the extension impact evaluation: RE and DfE will act as independent data controllers and FFT and UoW acting as data processors.

6.2. Ethics

The Consortium are committed to ensuring that all research activities that involve human participation or personal data are undertaken in line with our ethical principles. The Consortium have already obtained ethical approval for the current evaluation from the RAND internal review board and from the University of Westminster. The extension evaluation of APST has received ethical approval from the RAND internal ethics review board.

7. Reporting

The **current evaluation** reports will be delivered as planned:

- A non-publishable report including findings from process and cost evaluations in November 2023.
- A non-publishable report including draft findings from the impact evaluation analysis in December 2024.
- A publishable final consolidated report including findings from the process, impact, and cost evaluations in June 2025.

The **extension evaluation** reports would be as follows:

- A non-publishable report including findings from process evaluation in November 2024 (it is envisaged that this report will be useful for YEF and DfE in considering ongoing development of APST)
- A publishable final consolidated report including findings from process and impact evaluations submitted for review by DfE and YEF in February 2026 and to be published in June 2026.

Box 1: How the current and extension evaluation final reports will be connected

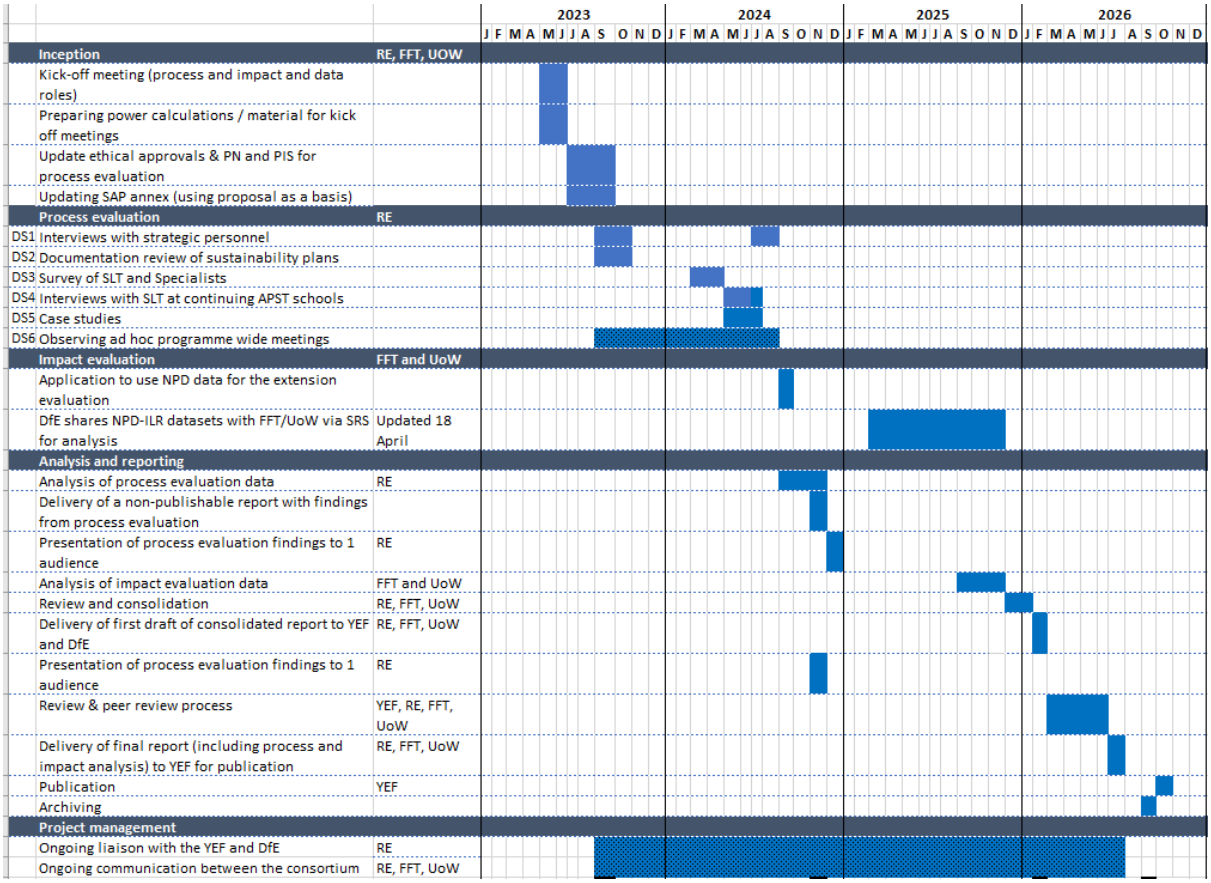
- The final consolidated report of the current evaluation, to be published by YEF in June 2025, will include findings relating to all the research questions for the current evaluation related to delivery of APST between November 2021 and August 2023. This report will state that an extension evaluation is being conducted and that publication of findings from the extension evaluation is expected in June 2026.
- The final consolidated report of the extension evaluation, to be published by YEF in June 2026, will include findings relating to all the research questions for the extension evaluation related to delivery of APST between September 2023 and August 2024.
- A reader of the final consolidated report of the extension evaluation report:
    - Will be aware that there was an earlier (the current) evaluation.
    - Will be aware of the key findings of this earlier (the current) evaluation: the extension evaluation report will include a summary of the key findings from the current evaluation.
    - Will understand where findings from the extension and current evaluation are the same or different. The extension evaluation report will refer to previous findings when relevant and will include reflections (in the conclusions), on the bigger picture, changes and continuities between the two evaluations

No formative feedback to schools or the DfE is planned as part of the extension evaluation. A presentation of key findings from the process evaluation to schools is planned. The exact timing is confirmed but will be after November 2024.

8. Timeline

Figure 2 outlines the proposed timeline for the extension evaluation.

Figure 2: Proposed timeline for extension evaluation



Annex A: Preliminary analysis to inform decision to change the primary outcome for Year 10

A.1. Introduction

As discussed in section 3.6, the extension evaluation includes an amended primary outcome from the current evaluation for Year 10 students.

The decision about primary outcomes in the current evaluation was informed by preliminary statistical analysis. The results of that analysis are included in the SAP and protocol for the current evaluation.[18]

The Consortium has conducted preliminary statistical analysis to investigate the amended primary outcome for the extension evaluation. This annex presents the results from the analysis.

As for the current evaluation, results of the preliminary statistical analysis are provided for two specifications:

- (a) The main specification: in which all AP schools in *all local authorities* not participating in ASPT form the comparison group.
- (b) The robustness specification (known as the regression discontinuity design hybrid): in which all AP schools in the *18 non-APST local authorities with the highest levels of serious violence* form the comparison group.

A.2. Data

Preliminary analysis for the extension evaluation uses the same data sources as the preliminary analysis conducted for the current evaluation (as set out in Annex A of the published SAP and protocol for the current evaluation.

Pupils in scope for the preliminary analysis are all those enrolled at an AP school (pupil referral units and AP free schools and academies) in Year 10. In the extension evaluation, pupils' initial post-16 destinations are observed at the end of October when pupils are in Year 12, as defined in the SAP and protocol for the current evaluation.

Difference-in-differences models are estimated with a placebo treatment year in 2017/18. Pupils' initial Year 12 destinations are observed in the 2019/20 academic year (the most recent ILR data available for the preliminary analysis). Models are fitted with and without

[18] https://youthendowmentfund.org.uk/funding/who-we-fund/alternative-provision-specialist-taskforces-apst-department-for-education/.

statistical controls, the latter being those shown in Table 9 of the SAP and protocol for the current evaluation.

A.3. Descriptive statistics

Table 1 of the attached Excel workbook presents summary data on cohort sizes and rates of Year 10 pupils subsequently observed in education in October of Year 12. Participation rates in APST schools were slightly higher than comparison schools in two of the five years shown, and lower in the other three years, particularly 2016.

Table 2 of the attached Excel workbook analyses pre-existing trends between APST schools and comparison schools using both a) the main specification and b) the robustness specification. Models are fitted both with and without statistical controls.

Participation was significantly lower among APST schools for the 2016 cohort across all specifications.

Although neither the main specification nor the robustness specification show any particular trends, the differences in participation rates between APST schools and comparison schools are erratic. This motivates the decision to explore the possible impact of non-parallel trends in pre-intervention data on estimated treatment effects using the simulation below.

This is needed because the proposed estimation approach relies on the assumption of parallel trends (that outcomes in participating APST schools evolve in a similar way to those in comparison schools). Essentially, the proposed estimation approach applies the observed trend in outcomes in comparison schools to the pre-intervention outcomes in participating schools to provide an estimate of what outcomes would have been in participating schools in the absence of the intervention (i.e. counterfactual outcomes). If the assumption of parallel trends does not hold, this approach will not be able to provide an unbiased estimate of the counterfactuals and so the overall impact estimates will likewise be biased.

A.3.1. Power Calculations

As in the SAP and protocol for the current evaluation, the Consortium empirically estimate minimum detectable effect sizes (MDES) for a placebo treatment year using the difference-in-differences specifications. These assume an alpha level of 0.05. This is shown in Table 9. Further information on estimated model parameters is presented in Table 3 of the attached Excel workbook.

Table 9: Empirically observed MDES for initial post-16 participation

(1)	(2)	(3)	(4)

Outcome	Year group (cohort)	Main specification (all comparison schools)	Robustness specification (comparison AP schools in the 18 non-participating local authorities with the highest levels of serious violence)
Initial post-16 participation	Year 11	0.08	0.12
Initial post-16 participation	Year 10	0.12	0.17

The first row of Table 9 restates, for the purpose of comparison, the corresponding MDES from Table 5 in the SAP and protocol for the current evaluation.

The second row of Table 9 presents the MDES for the Year 10 cohort. For both specifications (i.e. columns 3 and 4), the MDES for the Year 10 cohort falls below 0.2, indicating that the analysis is expected to achieve the power threshold typically required by the EEF for school trials.

Compared to the Year 11 cohort results, the MDES for the Year 10 cohort are roughly half as big again. This suggests the analysis will be less sensitive for the Year 10 cohort.

A bigger concern is whether year-on-year trends in the outcome level in APST schools run parallel to those in comparison APST schools. As explained above, the difference-in-differences estimator relies on the assumption of parallel trends holding in the pre-intervention period. For the year 10 cohort, a joint significance test of no interaction between year and being enrolled at a school in an APST area was statistically significant (F statistic<0.05). In other words, we cannot rely on the assumption of parallel trends. This is in contrast to the Year 11 cohort, where the same test was not significant.

A.3.2. Mean outcomes for schools close to the APST cutoff

The sum of serious violence and hospital admission percentile scores was used by the DfE to identify APST areas at the start of the programme. Those CSPs/LAs with a combined score greater than 1.82 were selected as APST areas and those with combined scores below 1.82 were not selected as APST areas. This discontinuity can be exploited for estimation purposes. The intuition is that CSPs/LAs 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. This is the regression discontinuity, or RD, estimator. Analogous to difference-in-differences, a difference in RD (DRD) involves deducting RD estimates from pre-intervention years to correct for fixed bias.

The graphs that follow concentrate on AP schools close to the cutoff and examine whether outcomes in previous years show a change at the cutoff. We define 'closeness to the cutoff' 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).
- Whether the sum of the percentile scores has 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 AP 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.

Figure 3 plots the mean outcome for a school against its 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 Year 10 pupils at a school who participated in Year 12. Pairs of graphs are shown for each successive cohort, from 2014-2018. For each cohort, 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.

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 the RD estimate is close to zero.**

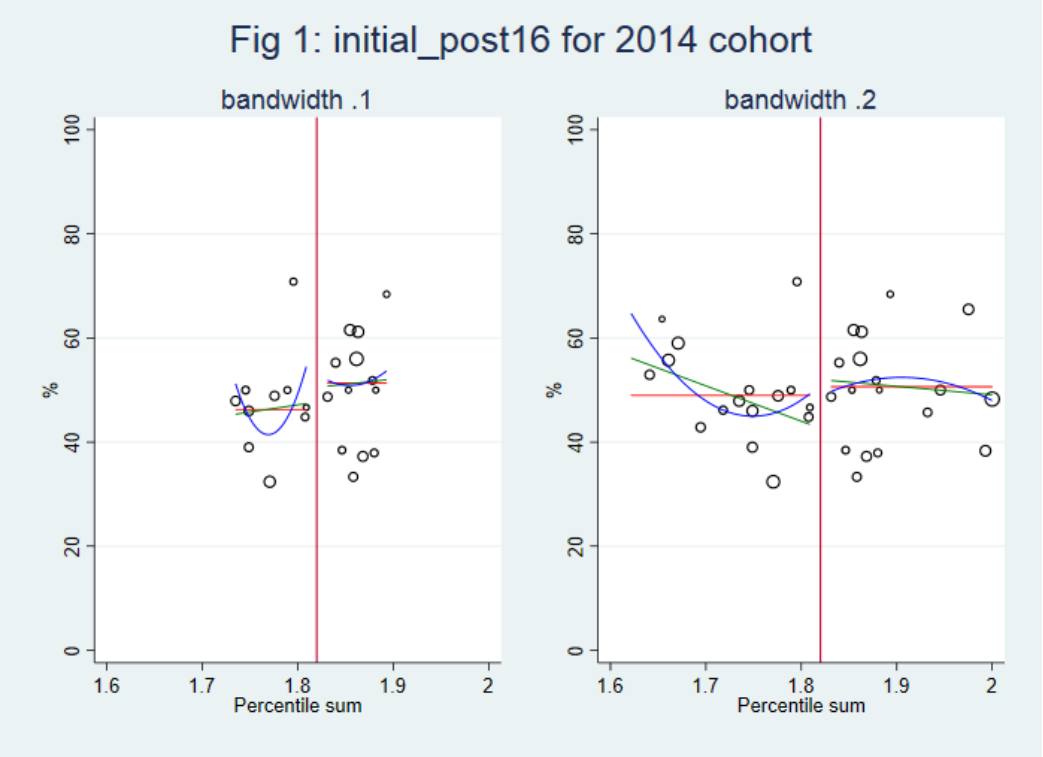
Summarising the results using the narrower bandwidth (0.1) definition:

- Local linear regression (green lines) performs best (in the sense of providing estimates closest to zero).
- The performance of local means (red lines) is more erratic. This is true also of the local quadratic regression (blue lines), which also appears over-sensitive in the sense that it changes it changes rapidly in the region of the cutoff.

Summarising the results using the wider bandwidth (0.2) definition:

- The picture is more mixed.
- The local means (red lines) approach slightly outperforms the local linear regression (green lines) approach but this is not consistent across cohorts.
- The local quadratic regression (blue lines) performs at least as well as the other methods but the fact that the curvature of the lines varies so much year-on-year cautions against relying on this specification in a DRD framework.

Figure 3: Mean outcome against percentile sum with different bandwidths (0.1 on the left, 0.2 on the right) for 2014 – 2018 cohorts



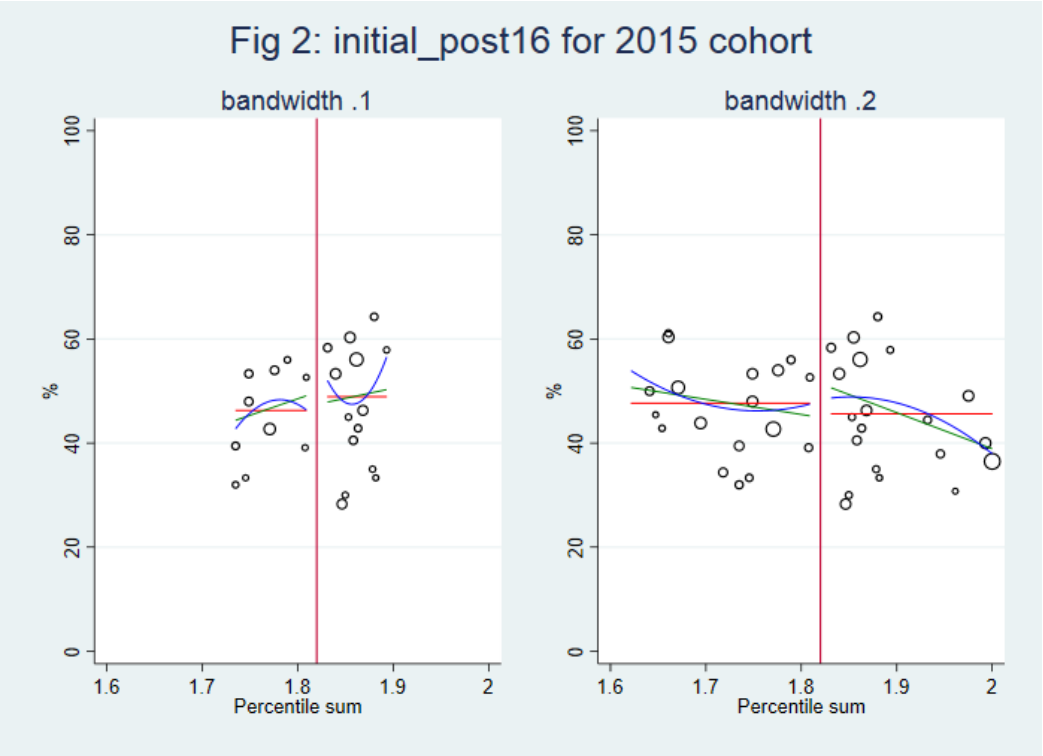


Fig 3: initial_post16 for 2016 cohort

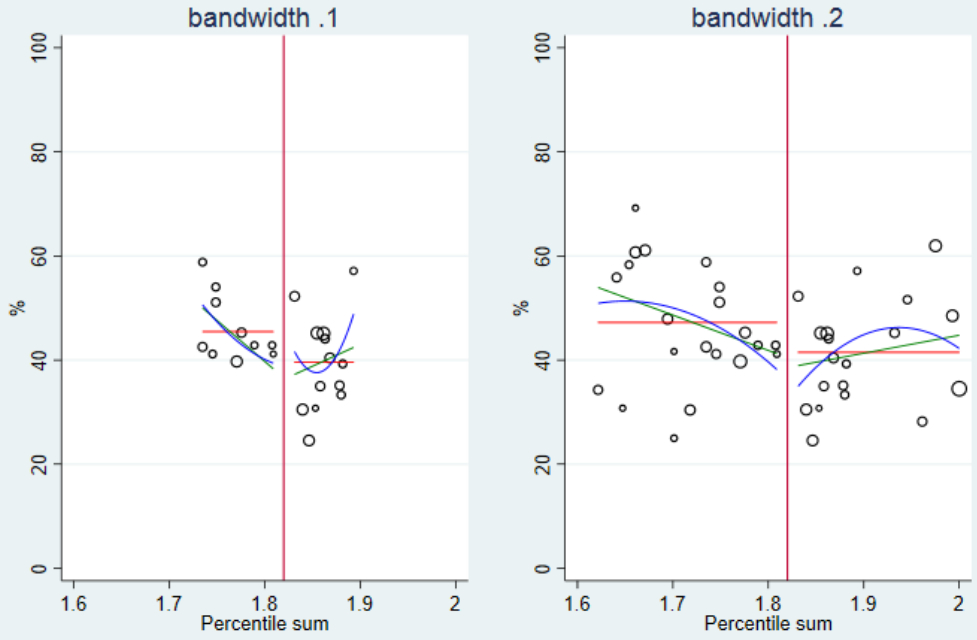
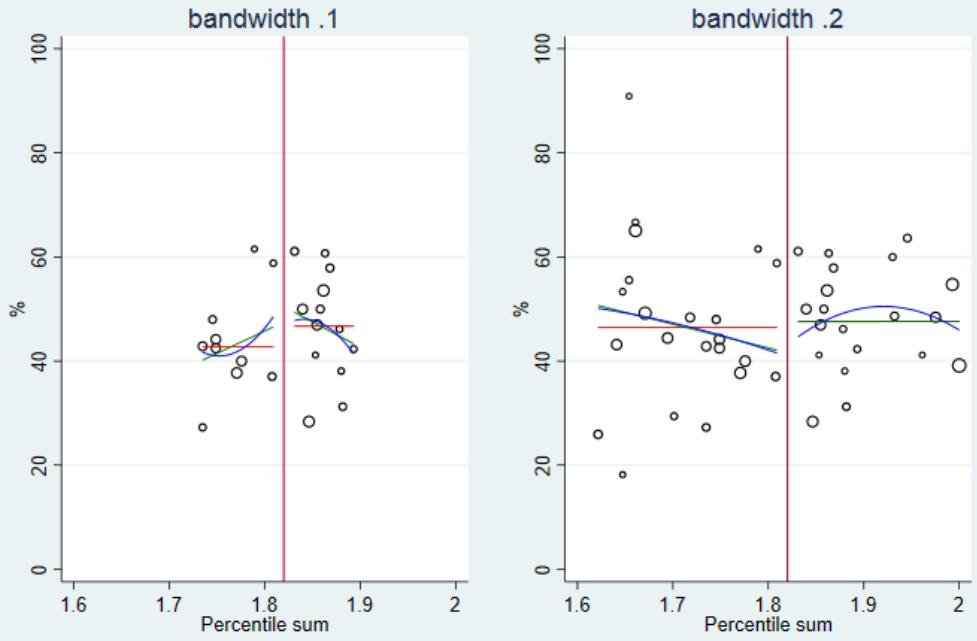
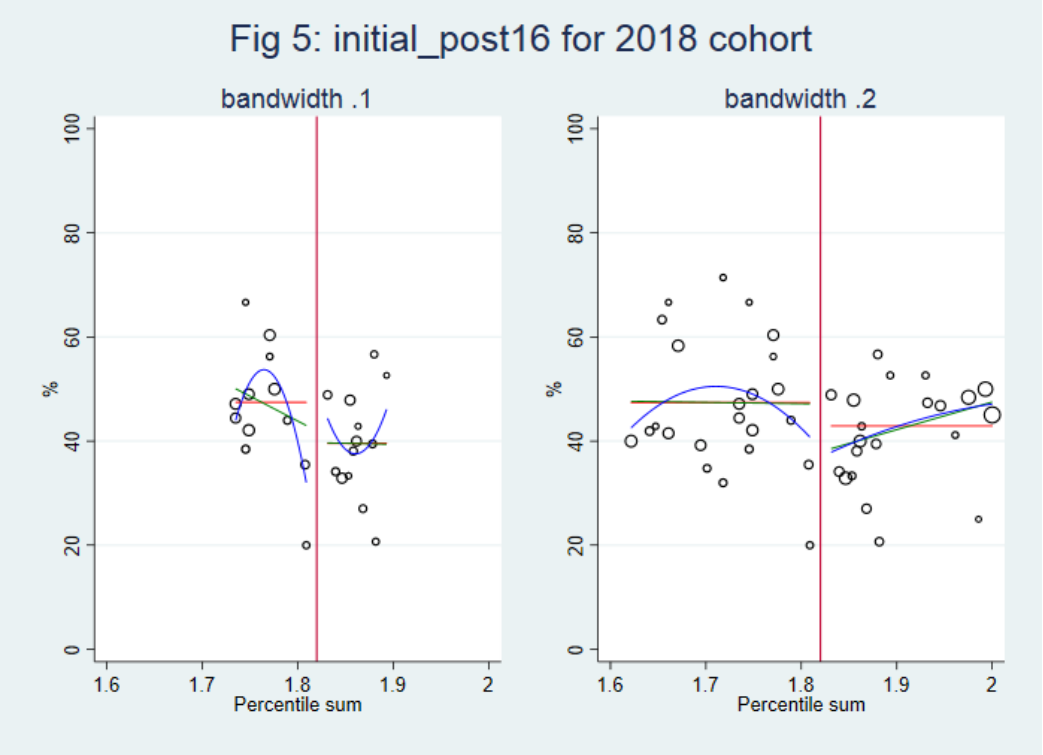


Fig 4: initial_post16 for 2017 cohort





Note: data points show school average outcomes. Those based on 10 observations or fewer are suppressed. In several cases it seems as if there are only two lines on one or both sides. This arises when two lines are very similar and appear to overlap; in fact, there are always three lines. The blue line represents the local quadratic regression, the green line the local linear regression, and the red line the performance of local means.

A.3.3. Simulation of treatment effects

We can adjust mean outcomes above the cutoff to simulate a treatment effect. Table 10 summarises the results of bootstrapping estimates of different effect sizes, without controls (upper panel) and with controls (lower panel).

Taking Panel A (without controls) first, the results of RD estimation are shown both under local means (lm) and local linear regression (llr). Each cell in the table represents the number of times a particular estimate was found to be statistically significantly different from zero for the 2018 cohort. In the first row, the effect size is zero. We would expect to find a significant impact in about 5% of replications. In fact, the rate of false positives is substantially higher in all cases. This can be understood from Figure 3, which shows the actual outcome to be appreciably lower above the cutoff. Successive rows relate to simulated impacts of 10, 15 and 20% of the standard deviation of the outcome, respectively. As such, they indicate the power of the estimation approach. Neither the lm nor the llr approach performs well. The DRD results use two cohorts – 2017 and 2018 – to estimate impacts. This also performs poorly, perhaps unsurprising since Figure 3 point to a lack of consistency across the cohorts in estimation bias.

The results in Panel B (with controls) show that controlling for pupil characteristics does little to alter the conclusion that neither the RD nor the DRD approaches reliably detects effects.

Table 10: Bootstrapped estimates for RD and DRD with different effect sizes, without controls (panel a) and with controls (panel b), using local means (lm) and local linear regression (llr)

	RD		DRD	
Effect size	lm	llr	lm	llr
Panel A: Without controls				
0	18%	16%	8%	40%
0.1	1%	2%	1%	7%
0.15	4%	0%	1%	2%
0.2	28%	0%	7%	1%
Panel B: Controls				
0	21%	50%	9%	26%
0.1	2%	14%	1%	5%
0.15	8%	6%	4%	2%
0.2	36%	2%	10%	1%
	Note: estimation uses bootstrapping on a		*Note: estimation uses bootstrapping on a*	
	sample of 1,890 pupils (200 reps)		*sample of 3,575 pupils (200 reps)*	

A.4. Conclusion

The conclusions that the Consortium draw from the preliminary statistical analysis are:

- A difference-in-differences analysis of post-16 participation for the Year 10 cohort is likely to be adequately powered.
- The Consortium have concerns around whether the assumption of parallel trends is satisfied.
- The Consortium recommend proceeding with the planned difference-in-differences analysis but we recommend adding an additional sensitivity test (see box below), to

be conducted at the time of the final impact analysis for the extension evaluation, which will mean that any caveats around the final results of the evaluation can be added and the results appropriately interpreted. For consistency with the current evaluation, we propose conducting DRD estimation as a further sensitivity analysis.

To get some sense of the sensitivity of the estimates to this violation, we will make use of a recently-introduced methodological innovation provided by Rambachan and Roth (2023).[19] Their HonestDiD approach provides an assessment of how big the failure of parallel trends would have to be in order to overturn a significant finding. It also gives confidence intervals that reflect this uncertainty.

Practically, our recommended approach follows Roth et al. (2023):

- Using results from the impact estimation, construct an event study plot showing stability of outcome differences between APST and non-APST areas in each year for which we have data
- Present diagnostics of the power of the pre-test (Roth, 2022)
- Conduct a sensitivity analysis to show the robustness of the conclusions to non-parallel trends (Rambachan and Roth, 2023).

More broadly, we further recommend that this approach be applied to estimators of all outcomes (as listed in the protocol for the current evaluation as well as in the extension).

Annex B: References

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

Rambachan, A. and Roth, J. (2023) *A more credible approach to parallel trends* Review of Economic Studies

Roth, J. (2022) *Pretest with caution: Event-study estimates after testing for parallel trends* American Economic Review: Insights

[19] Rambachan, A. and Roth, J. (2023) *A more credible approach to parallel trends* Review of Economic Studies

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413