

**Project Title: Academic Mentors Evaluation Study Plan**

**Evaluator (institution): NFER, University of Westminster and Department for Education**

**Principal investigator(s): Pippa Lord**

<b>PROJECT TITLE</b>	Academic Mentors Evaluation
<b>EVALUATOR (INSTITUTION)</b>	NFER , University of Westminster and the Department for Education (referred to as the Evaluator)
<b>PRINCIPAL INVESTIGATOR(S)</b>	Pippa Lord
<b>STUDY PLAN AUTHOR(S)</b>	Veruska Oppedisano, Richard Dorsett, Rachel Hayes, Ben Styles, Roland Marden, Luke Bocock, Alice Phillips, Pippa Lord,
<b>STUDY DESIGN</b>	Programme evaluation involving a quasi-experimental design (QED)
<b>PUPIL AGE RANGE AND KEY STAGE</b>	5-16 years (All Key Stage)
<b>NUMBER OF SCHOOLS</b>	<p><b>Primary School Evaluation Sample:</b></p> <p>134 primary schools in total: RL Star Reading scores (Years 1-6) for 67 intervention schools compared with 67 matched comparison schools</p> <p><b>Secondary School Evaluation Sample:</b></p> <p>154 secondary schools in total: RL Star Reading scores (Years 7-10) for 77 intervention schools compared with 77 matched comparison schools</p> <p><b>Year 11 Population Sample:</b></p> <p>542 schools in total: Maths and English Year 11 levels for the 271 intervention secondary schools compared with 217 matched comparison schools</p>
<b>NUMBER OF PUPILS</b>	<ul style="list-style-type: none"> <li>• Evaluation Samples: <ul style="list-style-type: none"> <li>○ 18 pupils per school, <ul style="list-style-type: none"> <li>▪ 2,412 children primary literacy</li> <li>▪ 2,772 children secondary literacy</li> </ul> </li> </ul> </li> <li>• Population Sample: <ul style="list-style-type: none"> <li>○ 542 schools, 18 pupils per school, <ul style="list-style-type: none"> <li>▪ 4,878 children secondary English</li> <li>▪ 4,878 children secondary Maths</li> </ul> </li> </ul> </li> </ul>
<b>INTERVENTION</b>	National Tutoring Program Academic Mentors

<p><b>PRIMARY OUTCOME MEASURE AND SOURCE</b></p>	<p>Attainment in English in primary and secondary schools. Attainment in Maths and English in Year 11.</p> <p>Within primary schools this will be based on standardised assessments (Years 1-6).</p> <p>Within secondary schools this will be based on standardised assessments (Years 7-10) and on the teacher assessed data that are replacing GCSEs in 2021 if available on NPD (Year 11).</p>
<p><b>SECONDARY OUTCOME MEASURE AND SOURCE</b></p>	<p>Not applicable</p>
<p><b>SCHOOL ELIGIBILITY</b></p>	<p>Schools must meet one of the following criteria to be eligible:</p> <ul style="list-style-type: none"> <li>• an Income Deprivation Affecting Children Index (IDACI) of 40 or greater (40% of pupils live in the three most deprived deciles)</li> <li>• an IDACI of 35 to 40 and an Achieving Excellence Areas (AEA) score of 4 to 6</li> <li>• an IDACI of 30 to 35 and an AEA score of 5 or 6</li> <li>• an IDACI of 25 to 30 and an AEA score of 6</li> <li>• Schools with an above average pupil premium rate (average set at 23.5) if there is an available local academic mentor</li> </ul> <p>There is also a flexibility rule that allows up to 30% of schools to be recruited even if they don't meet the national eligibility criteria above. This flexibility rule is only be applied where a school meets certain conditions that suggest it is 'disadvantaged' despite it not meeting the IDACI criteria. This enables the programme to respond to other dimensions of disadvantage, e.g. rurality. In the past two years, 10% of schools without the eligibility criteria have been recruited.</p>

## Study Plan version history

VERSION	DATE	REASON FOR REVISION
1.0 [ <i>original</i> ]	13/12/2021	The publication of this version was delayed due to ongoing changes to the proposed methodology caused by cancellation of statutory assessments in academic year 2020/21.

## Table of contents

### Contents

Study Plan version history	2
Table of contents	3
National Tutoring Programme: Academic Mentors – summary	4
About the National Tutoring Programme:	7
About the Academic Mentors Pillar of the National Tutoring Programme: programme theory and design	10
About the evaluation	12
Research questions	13
Design overview	19
Participants and selection mechanism	20
Sample size calculations	21
Outcome measures and other data	23
Sampling	25
Primary outcome analysis	27
Secondary analyses	31
Missing data	33
Effect size calculation	34
Ethics	35
Data protection	35
Personnel	39
Risks	40
Timeline	41
References	42

## National Tutoring Programme: Academic Mentors – summary

Schools in the most disadvantaged areas need more support than ever as they deal with the impact of school closures due to the covid-19 pandemic. Teach First is supporting the recruitment, training and placement of the National Tutoring Programme (NTP) Academic Mentors (AM). Academic Mentors (AM) provide support tailored to individual schools needs with a focus on subject-specific work (both one-to-one and in small-groups), revision lessons and additional support for those not in school. Academic Mentors are mostly graduates with some experience in education or working with pupils. They will not all be qualified teachers but may be working towards an initial teacher training qualification or considering a career in education. They will receive a package of ongoing training delivered by Teach First, and will be managed by their school. The vast majority of the AM’s salary is funded by the government, but the mentors are employed directly by the school. The programme aims to support teachers and schools in providing a sustained response to the Covid-19 pandemic and to provider a longer-term contribution to closing the attainment gap. The programme is part of a wider UK-Government response to the pandemic.

**Table 1: TIDieR framework – programme summary**

Aspect	Description
<b>Programme</b>	National Tutoring Programme: Academic Mentors (referred to in the evaluation and this study plan as the ‘AM programme’).
<b>Why (rationale)</b>	<p>Research shows that pupils’ learning has been affected by school closures due to Covid-19 (Cullinane and Montacute, 2020; Coe, 2020). The programme is designed to provide additional support to schools to help disadvantaged pupils whose education has been most affected by school closures due to Covid-19.</p> <p>There is a large body of evidence that tutoring and small-group tuition is effective – particularly where it is targeted at pupils’ specific needs; and that it can be particularly effective for disadvantaged pupils (Dietrichson et al., 2017; Education Endowment Foundation, 2018b; a; Nickow, Oreopoulos and Quan, 2020; Torgerson et al., 2018). Despite the difference in name – ‘tutor’ for TP and ‘academic mentor’ for AM – both services aim to provide a similar tutoring service, focused on 1-1 and small group intervention.</p>

Aspect	Description
<b>Who (recipients)</b>	<p>State-maintained primary and secondary schools in England. Schools must meet one of the following criteria to be eligible:</p> <ul style="list-style-type: none"> <li>• an Income Deprivation Affecting Children Index (IDACI) of 40 or greater (40% of pupils live in the three most deprived deciles)</li> <li>• an IDACI of 35 to 40 and an Achieving Excellence Areas (AEA) score of 4 to 6</li> <li>• an IDACI of 30 to 35 and an AEA score of 5 or 6</li> <li>• an IDACI of 25 to 30 and an AEA score of 6</li> </ul> <p>Schools with an above average (23.5) pupil premium rate if there is an available local academic mentor</p> <p>The interventions should support pupils from disadvantaged households or those whose education has been disproportionately impacted by Covid-19.</p> <p>The programme is expected to reach 946 schools and 90,000 children</p>
<b>What (materials)</b>	<p>The National Tutoring Programme (NTP) is a key part of the Government’s COVID catch-up response for schools and forms part of the wider Catch-Up Funding. The overarching vision of the NTP is to improve academic outcomes of the most disadvantaged young people. The Academic Mentoring programme is one of the two tutoring services provided as part of the NTP.</p>
<b>What (procedures)</b>	<p>Academic Mentors (AM) provide support tailored to individual schools needs with a focus on subject-specific work (both one-to-one and in small-groups), revision lessons and additional support for those not in school.</p>
<b>Who (provider)</b>	<p>Academic Mentors are mostly graduates with some experience in education or working with pupils. They will not all be qualified teachers but may be working towards an initial teacher training qualification or considering a career in education. They will receive a package of ongoing training delivered by Teach First, and will be managed by their school.</p>
<b>How (format)</b>	<p>Academic mentoring will be delivered either online or face-to-face.</p> <p>Each eligible school can employ up to two mentors in the following subject areas:</p> <ul style="list-style-type: none"> <li>• Maths</li> <li>• English</li> <li>• Science</li> <li>• Humanities</li> <li>• Modern Foreign Languages</li> <li>• Primary – numeracy</li> <li>• Primary – literacy</li> </ul>
<b>Where (location)</b>	<p>Academic mentoring is expected to be delivered in schools (before, during and after school), in addition to usual teaching. In certain circumstances, mentoring can be delivered at home (for example, for pupils who are shielding/medically vulnerable and are accessing their school learning from home).</p>

Aspect	Description
<b>When and how much (dosage)</b>	AMs are recruited in three waves and can join a school in either October 2020, January 2021 or February 2021. It is expected that AMs will work with at least 50 pupils between September 2020 and July 2021, pro-rata for AMs that start during the academic year.
<b>Tailoring (adaptation)</b>	As a result of the national lockdown in January-March 2021, when schools only remained open to children of keyworkers and vulnerable pupils, the NTP made provision for a greater proportion of academic mentoring to be available for pupils online at home during that time.

## About the National Tutoring Programme:

This section of the study plan sets out the current context in schools, the government-funding response, and the rationale for the NTP programme including evidence for academic mentoring. It then outlines the range of stakeholders involved in the programme (providers and participants). It highlights the key mechanisms and features of the programme and of mentoring that might affect or moderate pupils' learning outcomes.

### *Context in schools: Covid-19 and the impact on pupils' learning*

In response to the Covid-19 pandemic, the UK Government asked all schools in England to close in March 2020. Re-opening for some year groups was possible during June and July (where possible for Reception and Year 1, and for some Year 11s and Year 10s, as well as continued provision for children of keyworkers). All schools were asked to re-open from September 2020, and although schools now also have a duty to provide remote learning where needed, provision is still affected by the challenges of Covid-19 (for example, local closures, year groups or bubbles unable to be in school, teachers' teaching online at the same time as in class).

Research highlights that pupils are behind in their curriculum learning. In their review of the [impact of school closures on attainment](#), the Education Endowment Foundation (EEF), reported projections that school closures will widen the attainment gap between disadvantaged children and their peers (with a median estimate of widening by 36%), likely reversing progress made to narrow the gap since 2011 (Coe, 2020). According to the report of a national survey (weighted sample) of school leaders and teachers published in September 2020, teachers estimated that their pupils were behind in their learning with the average reported estimate being three months behind (Sharp et al., 2020) This issue was more acutely reported in the most deprived schools, and over half of teachers estimated that the learning gap between disadvantaged pupils and their peers had widened. Reasons relate to schools' varied provision of remote learning – a particular challenge for the most deprived schools, and for pupils from disadvantaged backgrounds, particularly around access to IT, having suitable spaces to learn in, and access to other learning resources (Cullinane and Montacute, 2020; Sharp et al., 2020; Hodgen et al., 2020).

The National Tutoring Programme (NTP) is a key part of the Government's COVID catch-up response for schools and forms part of the wider Catch-Up Funding. The overarching vision of the NTP is to improve academic outcomes of the most disadvantaged young people. The Academic Mentoring programme is one of the two tutoring services provided as part of the NTP. The NTP offers schools access to tutoring via two independent services: Tuition Partners (TP) and the Academic Mentoring (AM). With TP, schools select a tuition partner to work with from a list of approved tuition providers in their area. The tuition partner then provides tuition services for the school. AM is structured differently, but essentially provides a similar tutoring service. Teach First wholly manages the provision of tutors (referred to as 'academic mentors') to schools; recruiting, training and placing them in schools. Schools are not involved in the selection or initial training process of mentors but are given the opportunity to refuse a mentor. Unlike TP, the mentor then works in the school setting as a full-time employee of the school (under a short term contract).

Despite the difference in name – 'tutor' for TP and 'academic mentor' for AM – both services aim to provide a similar tutoring service, focused on 1-1 and small group intervention. It is important to note the distinction between 'academic mentoring' with this focus on tutoring and 'mentoring/coaching' that focuses on forming a close, personal relationship between mentor and mentee.

## ***Evidence for one-to-one and small group tuition***

In their review of the evidence on Covid-19 disruptions and the impact on attainment, the EEF suggested two key ways to support learning in these unprecedented times: i) to support effective remote learning to mitigate the extent to which the gap widens; and ii) sustained support to help disadvantaged pupils catch up. They particularly highlighted tuition as a route for providing support – in addition to high quality teaching and learning in the classroom. There is a large body of evidence that tutoring and small-group tuition is effective – particularly where it is targeted at pupils’ specific needs. The EEF toolkit pages on [one-to-one tuition](#) (Education Endowment Foundation, 2018a) and on [small group tuition](#) (Education Endowment Foundation, 2018b) show that both are effective interventions, and that training and support are important in the effectiveness of the tuition. Effect sizes vary across studies – with between three and six months additional progress being reported in studies of various one-to-one interventions; and in small group tuition the key finding across studies is that the smaller the group and the more aligned to pupils’ needs, the more effective the intervention.

Meta-analyses have shown that tutoring programmes yield consistent and substantial positive impacts on learning outcomes: the EEF Teaching and Learning Toolkit meta-analysis estimates the average effect size of tutoring to be 0.3 SD for small group tuition and 0.37 SD for one-to-one tuition; Nickow et al. (2020) found an overall pooled effect size estimate of 0.37 SD; Dietrichson et al. (2017) found a pooled effect size of 0.36 SD; and Ritter et al. (2009) found a pooled effect size of 0.30 SD.

Researchers also highlight that contributions to research on ‘recovery’ or ‘catch up’ should take into account lockdowns and absences throughout this academic year, patterns of recovery (Kuhfeld et al., 2020) and assumptions about different support strategies including online learning (Moss, 2020). Hence, any evaluation of mentoring should take into account wider context and ‘moderators’ that might affect the implementation of that mentoring. Moreover, research should also take into account that the counterfactual may be a very different ‘business as usual’ in the current climate – it is likely that pupils who do not receive AM will be provided with other forms of support by schools, and these may involve one-to-one or small group support (see funding response below).

## ***Particular benefits of tutoring for disadvantaged students***

There is evidence to suggest that the advantages of 1-1/small group tuition may be particularly relevant for disadvantaged students (Dietrichson et al., 2017; Torgerson et al., 2018). These students may suffer in the classroom due to comparison to their peers. A perceived sense of failure may result in low motivation and low self-efficacy, leading to poor learning outcomes. In contrast, teaching these students 1-1 or in homogenous small groups, allows favourable comparisons and allows teachers to communicate student improvements (Mischo and Haag, 2002). These incentives, in turn, help maintain high levels of motivation (Pintrich and Schunk, 2002).

## ***Who is involved in the AM programme? Providers and participants***

The following stakeholders are involved in the AM programme:



- Programme Managers at Teach First are leading the design and development of the AM programme in the academic year 2020–2021<sup>1</sup>. Programme Managers at Teach First are responsible for ensuring there is a rigorous recruitment and selection process of AMs, along with support and training throughout the academic year. Teach First will continue to coach the academic mentor through the year via the Curriculum and Training Leads.
- Academic Mentors: Whilst Teach First match AMs to schools, the AMs are employed directly by the school. AMs are responsible for delivering academic mentoring to participating pupils in schools with a focus on providing small group and one to one tutoring. AMs are recruited in three waves and can join a school in either October 2020, January 2021 or February 2021. It is expected that AMs will work with at least 50 pupils between September 2020 and July 2021, pro-rata for AMs that start during the academic year.
- Schools: State-maintained primary and secondary schools that serve disadvantaged communities throughout England are eligible to apply for an AM. In this context, disadvantaged communities is defined by the Income Deprivation Affecting Children Index (IDACI) and Achieving Excellence Areas (AEA) measures. Schools that have an above average pupil premium rate may also be able to apply for an AM if there is availability in their area.
- Pupils: Participating schools are able to identify which of their pupils they feel will most benefit from academic mentoring. Pupils can be in Year 1 – Year 11. Differently from TP, that targets PP pupils, AM does not have prescribed conditions on the characteristics of pupils who receive the intervention, although schools were encouraged to support pupils from disadvantaged households or those whose education has been disproportionately impacted by Covid. The programme is expected to reach around 50,000 pupils in the academic year 2020-21.

### *Context in schools: Spring 2021*

In January 2021, part-way through the AM programme, another national lockdown was announced, and schools were told to close to pupils other than those whose parent(s) were keyworkers, or who were identified as vulnerable pupils. All other pupils returned to home-schooling/remote learning. This had a significant effect on the delivery of academic mentoring, which had been planned to expand in earnest in January 2021. A number of key changes were made to delivery at this point by the NTP:

- most delivery moved online.
- the NTP agreed that during the lockdown, online mentoring could be delivered to pupils learning from home
- face to face mentoring was often only allowed within a defined ‘bubble’ i.e. a set group of pupils.

In addition to announcing that schools would close to many pupils, the government also announced that the summer exams – including GCSEs and Year 6 statutory assessments - would be cancelled. On 25<sup>th</sup> February 2021, it was confirmed that GCSEs would be awarded based on teacher assessed grades.

---

<sup>1</sup> Note, the design and delivery of the whole of the National Tutoring Programme in its first year will be led by a collaboration of five charities - the Education Endowment Foundation, Sutton Trust, Impetus, Nesta and Teach First - supported by the KPMG Foundation.

## About the Academic Mentors Pillar of the National Tutoring Programme: programme theory and design

This section of the study plan identifies the importance of certain delivery features and structures for effective learning. Academic mentoring is designed to address the effects of the loss of teaching time from school closures due to COVID-19. The clear message from the research is that academic mentoring needs to be sufficiently high quality with sessions having the right duration and frequency to achieve optimal results. It is recommended that tutors are knowledgeable in their subject area and trained in pedagogy, and that they deliver at least weekly sessions to students for a term or longer. Tutor subject knowledge and pedagogic expertise are commonly identified as important delivery elements for AM as well as the following structural characteristics, relationship with classroom learning, duration and frequency.

### *Tutor subject knowledge*

The literature suggests AM subject knowledge is beneficial for learning outcomes. Skilled teaching requires a complex interrelationship between knowledge of lesson structure and subject matter (Leinhardt and Greeno, 1986). Tutors with strong subject knowledge are more likely to be able to communicate that knowledge effectively to students. But learning can still occur where it is not present, for example, when tutors are peers or volunteers (Fantuzzo, King and Heller, 1992; Rogoff, 1990). Therefore, although AM subject knowledge should not be considered a prerequisite for tutorial learning it is clearly advantageous and preferable to it not being present at all.

### *Pedagogic expertise*

The techniques that AMs use to facilitate learning is widely acknowledged in the literature as important. In particular, academic mentoring that exploits the intimate environment offered by 1-1/small group tutorials is likely to be highly effective (Collins and Stevens, n.d.). In this sense, tutorials should be an interactive rather than a didactic experience between tutor and student (Lepper, Drake and O'Donnell-Johnson, 1997; Lepper and Woolverton, 2002). Tutors should make the tutorial a learning conversation in which students contribute much of the dialogue and the tutor intervenes appropriately to guide learning (Education Endowment Foundation, 2018a; McArthur, Stasz and Zmuidzinas, 1990; Merrill et al., 1992). Among the most important pedagogic principles identified is the idea of AMs managing conversations that encourage active learning from students (Chi et al., 2001). Ideally, students should be at the centre of these learning conversations, encouraged to explain their answers and ask questions and with tutors holding back from giving detailed explanations. AMs should also use this conversational style to probe students' understanding of content. For example, this could include AMs using comprehension-gauging questions rather than accepting student's own assessment of their understanding.

### *Structured format*

Aside from delivery, there is considerable focus in the literature on the most effective format for sessions. This relates to the frequency and duration of sessions as well as, when interventions take place in school, how sessions are coordinated with classroom learning. This latter issue is particularly relevant to the Academic Mentoring programme as all mentoring is directly managed by the school. The clear message in the literature is that the format and

coordination of sessions with classroom teaching has an important impact on the effectiveness of academic mentoring. In terms of format, short, regular sessions (30-40 minutes, three to five times a week) over a term or more appear to result in optimum impact. In terms of coordination, the close alignment of teaching to the classroom curriculum is strongly recommended.

### *Relationship with classroom learning*

An issue of concern in the literature is how targeted school interventions such as tutoring relate to wider school learning. Research suggests that learning is more effective when academic mentoring is linked with regular classroom teaching (Education Endowment Foundation, 2018a). However, experience has shown that academic mentoring in school can often be quite separate from classroom activities with relatively little connection between what students experience in and away from the classroom. In practice, this means it can be left to the student to make links between the coverage of the intervention and the wider curriculum coverage back in the classroom. Given that supported students are usually those who find accessing learning difficult in the first place, this presents an additional challenge.

The academic mentoring students receive should therefore be closely aligned with what is being taught in regular classes, for example, by providing remedial support on difficult topics. The coordination of academic mentoring and classroom teaching should be fostered by a close and supportive relationship between tutor and teacher.

### *Duration and Frequency*

Most studies demonstrate higher learning gains from extended periods of academic mentoring. For example, one study found that students receiving less than 20 hours tutoring scored 1 grade point higher than non-participants and those who had received more than 20 hours tuition scored 1.8 points higher than those who had no tuition (Smyth, 2008). Also the 20 week programmes Every Child a Reader and Every Child a Writer both showed larger achievement gains than the 10 hours of tuition provided through the Making Good Progress (Every Child a Chance Trust, 2009a; b). Studies suggest that intensive tutoring, where sessions are held several times a week tend to have greater impact (Elbaum et al., 2000)

## About the evaluation

### *Evaluation rationale and aims*

Given that national policy is aiming to support pupils' learning recovery in these unprecedented times, an evaluation of that support is important not only in terms of whether it is effective in supporting pupils' learning this year, but also in terms of how the programme as a whole is working, and any improvements needed to inform future tuition programme(s).

The evaluation aims to quantify the impact of the AM programme on pupil attainment/learning outcomes.

### *Evaluation design overview*

1. Impact evaluation – assessing the impact of mentoring on pupils' maths and literacy attainment outcomes, using a quasi-experimental design (QED) involving a comparison group.

When designing the evaluation, a number of issues were considered including: defining the research questions that could be answered; the appropriateness of a QED and how best to operationalise a comparison group design; scale and scope (note, the evaluation involves both population data for reporting on the monitoring data provided by AMs about all participants<sup>2</sup>, and evaluation sample data for analysing and reporting on impact<sup>3</sup>); burden on schools; use of NPD data and other assessment data; and the timescales of the programme and how best to provide formative feedback throughout the evaluation. These issues are discussed further where relevant.

---

<sup>2</sup> i.e. approx. 1,074 primary and secondary schools and 38,300 pupils.

<sup>3</sup> i.e. from a sample of schools that provide assessment data for the evaluation, involving 134 primary schools in English, 144 secondary schools in English; and 2412 primary pupils in English and 2592 secondary pupils in English (these numbers include AM and comparison group schools and pupils).

## Research questions

The research question for the evaluation is as follows (note further sub-research questions are detailed in Table 2):

**Impact RQ: What is the impact of AM on learning outcomes for pupils?** (this will be investigated through a number of estimators of impact, in English, in both primary and secondary schools and in Maths and English in Year 11).

The primary objective of the impact evaluation is to determine what difference, if any, is made by AM to attainment outcomes (maths and English). The impact evaluation uses a quasi-experimental design, due to the need to maximise reach to as many schools and pupils as possible, as soon as possible.

In primary and secondary schools, we will conduct the analysis on an evaluation sample of primary (years 1-6) and secondary (years 7-10) schools using standardised assessments with Renaissance Learning (RL). A third analysis will be conducted on the population year 11 pupils using the [teacher assessed GCSE grades](#) awarded in 2021. DfE has access to the RL assessment data and analysts from the University of Westminster will be seconded to the DfE to conduct the analyses in primary and secondary schools.

We will include analysis of attainment outcomes controlling for i) pupil characteristics, including gender, ethnicity, English as an additional language (EAL), year group, prior attainment, pupil premium (eligible vs. not), SEND vs. not, NTP tutoring received (vs. not); ii) school characteristics such as education stage (primary vs. secondary), Ofsted rating (high vs. low), proportion of FSM (high vs. low), and iii) other characteristics such as geography and Covid-19 hotspot<sup>4</sup>.

We will examine whether estimated impacts vary according to school characteristics (primary vs. secondary; high vs. low Ofsted rating; proportion of FSM; pupil characteristics (prior attainment; whether SEND; English as additional language; ethnicity; gender), geography and, if possible, whether the school is in a Covid-19 hotspot.

We will also run descriptive analyses to compare outcomes associated with different tutoring models among AM schools in the evaluation sample. The differences considered will include mode of delivery (online vs. face to face); timing of the session (during vs. after lessons); tutor:pupil ratio (1:1 to 1:10). We will also examine variation in outcomes by tutor characteristics (Qualified Teacher Status; teaching experience; shared gender with pupil/tutee).

The research questions are summarised in Table 2. There are six research questions. In five cases, there are two outcomes: English, primary and English, secondary. Research questions 1 to 3 involve analysis on a sample of AM pupils, and research question 4 involve the population of Year 11 pupils. The population analysis is only possible at secondary level due to changes in the summer exams. For each research question, the outcome/phase combinations are identified by a suffix: ep (English, primary), es (English, secondary), ey11 (English, year 11) and my11 (Maths, year 11). Hence RQ1.ep estimates the impact of AM availability on primary school PP pupils' English attainment. For each specific research question, Table 2 gives the outcome, the phase, the type of pupil for which impacts are

---

<sup>4</sup> We aim to include this however we are looking into sources of data for this variable. We currently plan to collect this information from the government website <https://coronavirus.data.gov.uk/>. We will also investigate whether we are able to identify school closures due to Covid-19 to use as a variable in the analysis.

estimated, the type of school used in the analysis and whether this uses the evaluation sample or the population of schools. The final column provides a brief description.

**Table 2: Impact research questions**

questions						
Impact estimates						
<b>RQ1: What is the impact of AM availability on PP* pupils' attainment?</b>						
Research Question**	Outcome	Phase	Pupil type	School type	Sample/ population	Description
RQ1.ep	English	primary	PP	AM/non-AM	Sample	<p>The estimation method is based on matching/weighting and instrumental variables. This approach builds on the eligibility criteria. We will use two instruments in our estimator: the combination of the IDACI and AEA thresholds and the fraction of PP pupils above the average.</p> <p>We focus on PP pupils since they are expected to form a large proportion of the eligible group and can be identified within both the AM and comparison schools. Using PP pupils avoids the complication of pupil selection as a result of school decision and pupil choice. The drawback is that the resulting estimate relates to PP pupils only rather than to participants who received AM as a whole.</p> <p>This analysis will be on PP pupils in year groups where at least one pupil is in receipt of AM (and equivalent year groups in comparison schools).</p> <p>This analysis will use Renaissance Learning assessments administered at baseline as covariates for primary and year 7-10 secondary schools. Outcomes will also be from Renaissance Learning assessment.</p>
RQ1.es	English	secondary	PP	AM/non-AM	Sample	
<b>RQ2: What is the impact of AM availability on the attainment of pupils predicted to participate?</b>						
Research Question	Outcome	Phase	Pupil type	School type	Sample/ population	Description
RQ2.ep	English	primary	predicted participants	AM/non-AM	Sample	The estimation method is based on matching/weighting and instrumental variables. This approach builds on the eligibility criteria. We will use two

RQ2.es	English	secondary	predicted participants	AM/non-AM	Sample	<p>instruments in our estimator: the combination of the IDACI and AEA thresholds and the fraction of PP pupils above the average.</p> <p>Using NPD data, we estimate a predictive model of pupil participation within AM schools and use that to predict which pupils participate in both AM and [would participate in] comparison schools. By doing this, we aim to move closer to an impact on AM participants as a whole rather than PP pupils.</p> <p>This analysis will use Renaissance Learning assessments administered at baseline as covariates for primary and Year 7-10 secondary schools. KS2 SATs from NPD will be used for year 11. Outcomes will also be from Renaissance Learning assessment, except among year 11 for whom NPD data (teacher assessed GCSE) will be used.</p>
RQ2.ey11	English	secondary	predicted participants	AM/non-AM	Population	
RQ2.my11	Maths	secondary	predicted participants	AM/non-AM	Population	
<b>RQ3: What is the impact of the availability of AM on all pupils' attainment?</b>						
<b>Research Question</b>	<b>Outcome</b>	<b>Phase</b>	<b>Pupil type</b>	<b>School type</b>	<b>Sample/ population</b>	<b>Description</b>
RQ3.ep	English	primary	All pupils in years with AM	AM/non-AM	Sample	<p>The estimation method is based on matching/weighting and instrumental variables. This approach builds on the eligibility criteria. We will use two instruments in our estimator: the combination of the IDACI and AEA thresholds and the fraction of PP pupils above the average.</p> <p>We estimate impacts for all pupils in year groups receiving AM, regardless of whether they participate in AM. These estimates are likely to be smaller than RQ1 and RQ2 estimates since there is no attempt to identify pupils more likely to participate in AM and so its impact will be more diluted. Should AM be spread between a smaller number of schools extending eligibility to a larger proportion of their pupils, this dilution may be reduced. This estimator also captures the effect of spillover (peer) effects.</p> <p>This analysis will use Renaissance Learning assessments administered at baseline as covariates for primary and years 7-10 secondary schools. KS2 SATs from NPD will be used for year 11.. Outcomes will also be from Renaissance</p>
RQ3.es	English	secondary	all pupils in years with AM	AM/non-AM	Sample	
RQ3.ey11	English	secondary	all pupils in years with AM	AM/non-AM	Population	
RQ3.my11	Maths	secondary	all pupils in years with AM	AM/non-AM	Population	



						Learning assessment, except among year 11 for whom NPD data (teacher assessed GCSE) will be used.
<b>RQ4: What is the impact of AM availability on pupils' attainment in schools with values of the eligibility criteria just below and above the thresholds in Year 11?</b>						
<b>Research Question</b>	<b>Outcome</b>	<b>Phase</b>	<b>Pupil type</b>	<b>School type</b>	<b>Sample/ population</b>	<b>Description</b>
RQ4.my11	maths	year 11	PP pupils, predicted participants, all pupils	AM/non-AM	population, year 11	We estimate impacts for PP pupils, predicted participants and all pupils in all schools where year 11 is receiving AM in order to provide sensitivity analyses on the corresponding impacts for sampled schools (mirroring RQ1, RQ2 and RQ3). The population refers to all secondary schools in England (so all AM schools compared to all non-AM schools). This approach builds on the eligibility criteria.
RQ4.ey11	English	year 11	PP pupils, predicted participants, all pupils	AM/non-AM	population, year 11	We will use two instruments in our estimator: the combination of the IDACI and AEA thresholds and the fraction of PP pupils above the average.  This analysis will use the teacher assessed data that are replacing GCSEs in 2021. It will use KS2 scores as baseline covariates.
<b>RQ5: How does the impact of AM availability vary among PP pupils, by school and pupil characteristics?</b>						
RQ5.ep	English	primary	PP	AM/non-AM	sample	We revisit RQ1 to explore the extent to which impacts for PP pupils vary according to a number of school and pupil characteristics.
RQ5.es	English	secondary	PP	AM/non-AM	sample	This analysis will use Renaissance Learning assessments administered at baseline as covariates for primary and Year7-10 secondary schools.
<b>RQ6: How do outcomes vary among AM pupils, by model of tutoring?</b>						
<b>Research Question</b>	<b>Outcome</b>	<b>Phase</b>	<b>Pupil type</b>	<b>School type</b>	<b>Sample/ population</b>	<b>Description</b>
RQ6.ep	English	primary	Pupils receiving AM	AM	Sample	We examine how outcomes vary according to a number of aspects of AM-related factors :
RQ6.es	English	secondary	Pupils receiving AM	AM	Sample	1. School characteristics: Ofsted rating (high vs. low); proportion of FSM (high vs. low); type of school (academy/maintained); school size (by quartile). These

						<p>variables identify the context where AM is delivered and allow to analyse whether AM has been more effective in disadvantaged contexts.</p> <p>2. Pupil characteristics: prior attainment; pupil premium (eligible vs. not); SEND vs. not; English as an additional language, ethnicity and gender. These variables identify participants and allow to analyse whether AM has been more effective on disadvantaged pupils or children with specific demographics.</p> <p>3. Other: Covid-19 hotspot. School closure due to Covid-19 may affects learning disruption at the school level, and therefore the decision to use AM..</p> <p>Since these are only observed among AM schools, we present descriptive statistics rather than impact estimates.</p> <p>This analysis will use Renaissance Learning assessments administered at baseline as covariates for primary and Year 7-10 secondary schools.</p>
--	--	--	--	--	--	--

\* PP=Pupil Premium

\*\* ep=English, primary, es=English, secondary, ey11=English, year 11, my11=Maths, year 11

## Design overview

Table 3: Design overview

<b>Design</b>	Matching/weighting Instrumental Variable
<b>Unit of analysis (school, pupils)</b>	Pupils from schools qualifying for Pupil Premium
<b>Number of Units to be included in analysis (Intervention, Comparison)<sup>5</sup></b>	<p><b>Primary:</b> 134 total schools: 67 intervention schools in Literacy and 67 comparison schools. 134 total schools.</p> <p>For Literacy, primary: 2412 pupil premium pupils (1206 from intervention schools and 1098 from comparison schools)</p> <p><b>Secondary:</b> 154 secondary schools (77 intervention schools and 77 comparison schools), for Star Reading (144 total schools)</p> <p>For Star Reading, secondary: 2772 pupil premium pupils (1386 from intervention schools and 1386 from comparison schools)</p>
<b>Outcomes</b>	<b>Variable</b> Educational attainment
	<p>measure (instrument, scale, source)</p> <p>Attainment in English in primary and years 7-10, Attainment in English and Maths in year 11.</p> <p>In primary and years 7-10 this will be based on RL assessment, standardised.</p> <p>In year 11, this will be based on teacher assessed grades.</p>
<b>Baseline for outcome</b>	<b>Variable</b> Educational attainment
	<p>measure (instrument, scale, source)</p> <ul style="list-style-type: none"> <li>Standardised assessments in Star Reading at baseline for primary and Years 7-10 secondary school pupils</li> <li>KS2 SATs from NPD will be used for year 11</li> </ul>

<sup>5</sup> Depending on the method used, the number of units included in the analysis can differ from the pool of potential comparison units. For example, when using matching/weighting the pool of comparisons units could represent all schools in England, but only a certain number of units will be included in the analysis after a suitable match is found. Identifying the precise number of units included might not be possible at the design stage. In these cases Evaluators can speculate on the number of units that are expected depending on the method used.

## Participants and selection mechanism

### *A note on the sample*

Decisions around selection and recruitment have been informed by practical reasons, as well as the methodological design. We recognise that this study is being conducted in atypical circumstances (the ongoing Covid-19 pandemic) and we are conscious of the pressures and challenges that schools are facing. This is one of the main drivers of our decision to use schools that already administer RL tests, so as not to place additional burden on pupils and schools. Together with EEF we also recognise that for ethical reasons relating to the loss of teaching time due to Covid-19 it would not be appropriate to randomise, withhold or restrict access to the mentoring for schools and pupils that wish to receive it.

The planned population analysis is intended to serve as an important check against the analysis on the impact sample. The advantage of the population analysis over the evaluation sample analysis is that there is no selection involved in taking part in the evaluation other than choosing to become an AM school. The relative disadvantage is that it is limited to year 11, it does not use recent baseline assessments and the outcome measure is potentially less reliable this academic year than standardized assessments.

### *How schools join the AM programme: AM schools*

Schools are eligible to apply for an Academic Mentor if they have either:

an Income Deprivation Affecting Children Index (IDACI) of 40 or greater (40% of pupils live in the three most deprived deciles)

an IDACI of 35 to 40 and an Achieving Excellence Areas (AEA) score of 4 to 6

an IDACI of 30 to 35 and an AEA score of 5 or 6

an IDACI of 25 to 30 and an AEA score of 6

or if a school does not meet the criteria above, but has an above average pupil premium, set at 23.5%.

### *Population descriptive statistics*

Part of the evaluation will involve descriptive statistics of reach and spread of the provision overall – both in terms of the AM schools and AM pupils in the AM population. AM population data will be used to describe the number of schools and pupils involved in the AM programme, and their characteristics (such as Pupil Premium), as well as an overview of attendance at sessions. The population data is collected and recorded on an ongoing basis by AMs. The evaluator will collate, but not quality assure, the data that is provided. The Evaluator will match the population data to the NPD.

### *Recruitment to the evaluation sample*

AMs are recruited in three waves and can join a school in either October 2020, January 2021 or February 2021. Schools registered their interest in hosting an AM directly with Teach First

and shared key information about their needs and requirements (e.g. subject area, ideal candidate attributes, context about school). Teach First then did an eligibility check based on the contractual requirements. To be eligible for inclusion in the evaluation sample, schools must be using Renaissance Learning standardised assessments in English during the academic year 2020/21 with test scores being recorded using an online database. As part of the agreement between the school and Teach First, opt-out agreement to share this assessment data was sought when the school signed the partnership agreement with Teach First. A further opt-in agreement was sought within the RL online portal with in-product notifications alerting schools of this requirement and telephone calls to non-consenting schools to confirm they had withdrawn their opt-out consent.

In order for a school to take part in the intervention evaluation sample, the school must have at least one school year in which pupils are receiving mentoring and in which Renaissance Learning standardised assessments are being conducted in English and test scores are being recorded using an online database.

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

Comparison schools will be drawn from a sample of 220 schools who have expressed an interest in working with Teach First in the academic year 2021/22. Schools must be using Renaissance Learning standardised assessments in English during the academic year 2020/21 with test scores being recorded using an online database. As part of the agreement between the school and Teach First, opt-out agreement to share this assessment data was sought when the school signed the partnership agreement with Teach First. A further opt-in agreement was sought within the RL online portal with in-product notifications alerting schools of this requirement and telephone calls to non-consenting schools to confirm they had withdrawn their opt-out consent.

We will select a sample of non-AM comparison schools that shares similar characteristics to the AM schools in the evaluation sample. Schools will be matched on the characteristics in the sampling section. For each AM school signed up to the evaluation we will identify similar-looking non-AM schools, with the aim of selecting them as comparison schools. Only schools that use RL and have agreement in place to share these assessments data with DfE and TF will be eligible for inclusion.

An alternative option if we struggle with selecting non-AM schools, is the possibility of selecting comparison pupils using RL data on non-AM years in AM schools.

### **Sample size calculations**

We use cluster randomised trial power calculations to provide an indication of the MDES. We allow for clustering of pupils within schools. We note that the evaluation is not based on a randomised intervention but instead relies on a non-experimental approach. Hence, these power calculations should be viewed as approximations.

67 primary schools and 77 secondary schools are doing AM and have already agreed in sharing their RL data with the DfE and with Teach First. Control schools will be drawn from

the pool of 220 schools that have agreed to host a Teach First trainee teacher from September 2021, share their RL data and were not hosting an AM in the academic year 2020/21.

Table 4 contains the achieved sample sizes, i.e. pupils with: (1) baseline and (2) end-point assessment, and for whom we have (3) the required pupil data (for matching to NPD, tutoring attendance data). Based on the three data points, we have the following targets of:

- 134 primary schools
  - 134 primary schools (67 intervention and 67 comparison) for English
- 144 secondary schools:
  - 144 secondary schools (77 intervention and 77 comparison) for English

Assumed sample size is based on a conservative level of take-up within schools (we have assumed 18 pupils per school at primary and secondary<sup>6</sup>); this allows for schools with incomplete coverage of year groups, either in terms of tutoring or testing. It also reflects our intention to focus primarily on Pupil Premium pupils as the intervention targets disadvantaged children that can therefore be easily identified in control schools. Note that since our analysis focuses on disadvantaged pupils, we do not produce separate estimates for the FSM subgroup.

For both primary and secondary schools, we assume similar parameters for primary and secondary schools, noting that the EEF Catch-up Literacy and Catch-up Numeracy evaluations likewise assumed ICCs that were quite similar. We note that, among comparable EEF trials, the ICC among the achieved sample is sometimes higher than that assumed when designing the trial. For instance, with the Tutor Trust re-grant, the actual ICCs were 0.29, 0.17 and 0.23 for maths, reading and GPS (Grammar, Punctuation and Spelling), respectively rather than the assumed 0.19. Our assumptions regarding the ICCs are, if anything, conservative. (Allen et al., 2018) suggest ICCs of 0.07 at the end of KS1, 0.10 at the end of KS2 and around 0.15 at the end of KS4.

For primary Literacy, we estimate a MDES of 0.119 and for secondary Literacy an MDES of 0.114.

For the year 11 population analysis, there are 452 secondary schools doing AM. Based on previous estimates of useable data that the AMs provided, we assume 60% of secondary schools have usable attendance data, which leaves us with 271 schools. Power calculations suggest that with 271 AM secondary schools and 271 non-AM schools the MDES would be 0.095. This calculation used the R package 'PowerUpR'.

As noted above, these power calculations are based on a randomised experiment design. The analysis will use matching and IV regression. Assuming the identifying assumption for

---

<sup>6</sup> The number is based on the data on pupils receiving AM. On average, in primary schools, there are 36 pupils per school doing Literacy. About 50% of them are PP pupils which leave us with 18 pupils. We will need to review these numbers once the data collection is finalised.

matching holds, the reported MDES should be good approximations for the matching case. With IV, it is less clear that the reported power calculations will provide good approximations.

Table 4: Sample size calculations for estimation sample (post-attrition)

		Primary English	Secondary English
<b>Minimum Detectable Effect Size (MDES)</b>		0.119	0.114
<b>Pre-test/ post-test correlations</b>	level 1 (pupil)	0.70	0.70
	level 2 (school)	0.70	0.70
<b>Intracluster correlations (ICCs)</b>		0.15	0.15
<b>Alpha</b>		0.05	0.05
<b>Power</b>		0.8	0.8
<b>One-sided or two-sided?</b>		2-sided	2-sided
<b>Average cluster size</b>		18	18
<b>Number of schools</b>	Intervention	67	77
	comparison	67	77
	<b>Total</b>	134	154
<b>Number of pupils</b>	Intervention	1206	1386
	comparison	1206	1386
	<b>total</b>	2412	2772

## Outcome measures and other data

### *Key principles*

The QED aims to look at the impact of mentoring on attainment, as the purpose of the NTP AM is to support pupils to catch-up and reduce the amount of learning lost due to the COVID-19 pandemic and the restrictions on schools in 2020 and 2021.

We will use standardised English assessments provided by Renaissance Learning (RL) that have been standardised using a nationally representative sample of students in the recent past. Standardised tests yield standardised scores, which have a fixed mean and standard

deviation. This, in turn, means that if we align their scales using their national standardisation parameters, standardised scores from different year groups can be analysed together thus allowing impact to be measured across all year groups in each phase simultaneously.

In order to match the attainment data to the pupils, we will collect data about pupils in AM schools and non-AM comparison schools. Box 1 below outlines the data to be collected about different groups of pupils covered by the evaluation.

**Box 1: Data to be collected:**

**Named/NPD data collected for the population of Year 11 pupils (RQ4):**

Collect pupil data (name, DOB, UPN) from all year 11 pupils who have received AM in one of the Evaluation Schools via Teach First

This named data will be matched to NPD to collect the longer list of background variables for the population analysis

**For the impact evaluation sample (RQ1, RQ2, RQ3, RQ5, RQ6 [intervention only]) -  
Named/RL data collected from the intervention and comparison schools:**

**This applies to primary schools and to year 7-10 secondary school**

AM (intervention) schools: collect pupil data/assessment data/mentoring attendance data\* IF there is at least one pupil in the year group receiving tuition and the year has done/is doing standardised RL assessments.

Comparison schools: collect pupil data/assessment data from all year groups with standardised RL assessment data

**Use of de-identified data:**

We will access de-identified data from NPD for all y11 pupils in all English schools as part of the population analysis (RQ4). This will use the prior national curriculum assessment as the baseline and the teacher assessed data that are replacing GCSEs in 2021 as the endpoint (i.e. for y11 use Y6 for BL and Y11 as EP). We will need it from all schools to establish a comparison.

We would then add in the named data about the AM pupils so we know which pupils in Y11 received mentoring. Therefore the population analysis in the robustness check would be a hybrid of named (covered above) and de-identified data.

***Baseline measures***

For year groups 1-10, the baseline measures used will be the RL standardised test taken by pupils in the autumn term 2020 (or early spring term 2021, so long as it is before the start of AM in the school). This will include tests in literacy, analysed separately by phase.

Agreements have been made with RL for them to share the test results (overall score) with the DfE for schools that give permission to do so. RL will share data for the entire year group in participating intervention schools if at least one pupil in that year group has received academic mentoring. The University of Westminster is seconded to the DfE to access these data in the analysis of the intervention.



## **Primary outcome**

For Year 11 we will use the alternative agreed by the Department of Education (teacher assessed grades).

For all other year groups, in each school we will use the same RL test in literacy used at baseline taken during the summer 2021 term.

## **NPD data for evaluation sample**

We will ask the NPD team to match in named data to our population sample pupil list via URN, UPN, name and DOB. In addition to the results of the test that is replacing GCSEs, we will request the variables listed below.

In order to establish comparison groups, we will request de-identified data for all pupils born between 1<sup>st</sup> September 2003 and 31<sup>st</sup> August 2015 including their final KS2 and/or KS4 pupil data from 2014/15 until 2020/21 (including those with endks=0 or schres=0).

We will collect the following variables:

- Basic data for matching to NPD (name, date of birth, Unique Pupil Number - UPN)
- Background characteristics such as gender, ethnicity, socio-economic status
- Information on pupil performance / attainment
- Special educational needs
- Attendance / exclusion
- Interactions with social services

For year 11 data, we will adjust for multiple testing using the simulation approach of Clarke, Romano and Wolf (2020), as implemented by the Stata program `rwolf` ado.

## **Sampling**

### **Propensity Score Matching**

We will use matching to construct from the pool of eligible non-AM schools that use RL and are recruited in the Training Programme that Teach First will deliver from September 2021. Schools recruited in this Training Programme have sharing agreements with DfE and TF to access RL tests results. The comparison group will be selected from this pool of schools, so that it is similar in important regards to the AM schools in the evaluation sample. This assumes that sufficient school characteristics can be observed to control for selection into AM (the 'selection on observables' or 'conditional independence' assumption). It is this type of selection that (Weidmann and Miratrix, 2021) consider, providing evidence that simple matching approaches may work well for this purpose.

Our sampling approach exploits the eligibility thresholds to identify the sample of schools that would be eligible to receive AM. Schools are eligible to apply for an Academic Mentor if they have either:

1. an Income Deprivation Affecting Children Index (IDACI) of 40 or greater (40% of pupils live in the three most deprived deciles)
2. an IDACI of 35 to 40 and an Achieving Excellence Areas (AEA) score of 4 to 6
3. an IDACI of 30 to 35 and an AEA score of 5 or 6
4. an IDACI of 25 to 30 and an AEA score of 6
5. Or if a school does not meet the criteria above, but has an above average pupil premium, set at 23.5%.

The national eligibility criteria for the Training Programme are the same as the AM national eligibility criteria. In place of the pupil premium threshold, there is a flexibility rule that allows up to 30% of schools to be recruited even if they don't meet the national eligibility criteria above. This flexibility rule is only be applied where a school meets certain conditions that suggest it is 'disadvantaged' despite it not meeting the IDACI criteria. This enables the programme to respond to other dimensions of disadvantage, e.g. rurality. In the past two years, 10% of schools without the eligibility criteria have been recruited.

Our matching approach is set out in detail in the steps below:

1. Identify AM and non-AM schools in each different categories of each eligibility criteria (IDACI score, AEA and proportion of FSM eligible pupils used as proxy for pupil premium).
2. Estimate a probit regression of AM participation conditioning on the IDACI score, the AEA score, and the FSM population within each one of the categories.
3. Using the estimation results, predict the probability of being an AM school (the propensity score) within each category.
4. Working through the AM schools in reverse order of their propensity score, match each one to its nearest non-AM school without replacement, subject to the propensity scores being sufficiently close, within each category. This can be done in Stata using the psmatch2 ado file by issuing the command: "psmatch2 treatment\_variable xvariables, descending noreplacement". We choose to match without replacement since the aim of matching is to identify schools to recruit into the evaluation sample.
5. Save the matched non-AM schools to a file; let us call this file NonAM1.
6. Repeat steps 8 times, each time saving a file of matched schools and then removing those selected from the pool of potential comparator schools (i.e. if a school in NonAMi is matched then this will be removed for groups m>i). This results in 8 files NonAM1, ,,,, NonAM8.
7. Schools in the matched samples that are testing for the relevant year groups and subject will be selected in the comparison group starting with NonAM1 and working down the list.

Matching will use the AM schools and the pool of schools in the Training Programme. We emphasise that matching will be carried out as a means of identifying the sample and therefore is prior to outcomes being known. Also, note that the matching will be conducted separately for primary and secondary schools.

We plan to augment the evaluation sample with some ineligible schools, as that would allow us to use the eligibility criteria to achieve identification of the impact of AM. To achieve this, we would include some TP and non-TP Research Champion schools that use RL data. Since these are not a random sample, we will construct weights to restore sample representativeness.

To test whether propensity score matching controls for school selection, we will conduct placebo tests for the AM schools in earlier years using the data from the NFER Register of Schools. If the selection of the control sample controls adequately for unobserved factors, we do not expect to find any significant difference in attainment between AM and control schools prior to the intervention. We will perform a placebo test at the outset as a way to inform our approach and sampling. We will then perform a second placebo test after constructing the sample of matched comparison schools to check the similarity of the two groups of schools before the intervention. The placebo testing will be done for each of the three preceding years (except 2020 for which there are no KS2 and GCSE available and 2019 for which there is no data available), using results for KS2 and GCSE to demonstrate similarity of the achieved match in the past.

## Primary outcome analysis

We will use matching/weighting estimates and IV regression, described below.

### *Matching/weighting: What is the impact of AM availability on PP pupils' attainment? (RQ1)*

The sample of non-AM schools identified using the matching-based sampling approach described above may have some differences from the AM schools. This could arise particularly if recruitment is difficult such that the achieved comparison group is made up of schools which are less well-matched. To address this, we will construct weights to bring the achieved non-AM sample into line with the AM sample. Furthermore, working with the achieved sample, we can construct these weights using both pupil- and school-level variables. This allows us to more tightly control for differences between the AM and non-AM schools. In addition to the variables listed in the table above, we will also include English as an additional language, special educational needs and interaction with social services (all variables taken from the NPD).

We choose a weighting approach rather than a second matching stage because, having collected test data, we want to avoid any sample loss. We will construct weights using entropy balancing. This can be implemented in Stata using the ebalance ado file. It has the advantage of automating the process of balance checking and thereby reducing the scope for researcher bias. We emphasise that this weighting step will be carried out before outcomes are known.

Since RQ1 focuses on PP pupils, the weights will be calculated only for the PP pupils. We will report a comparison of the characteristics of AM schools in the evaluation sample with:

- 1) all non-AM schools;

- 2) the matched sample of comparison schools that forms the basis for recruitment;
- 3) those effectively recruited for the study as comparison schools;
- 4) those effectively recruited for the study as comparison schools, weighted.

To estimate impacts, we will regress the outcome on an AM indicator, baseline measure of the outcome of interest and year group indicator. All school level variables used for the matching will be used as controls. Pupil-level controls will include background variables, such as gender, ethnicity, EAL, special educational needs, geography and Covid-19 hotspot. Residuals will be clustered at the school level to account for any common school-specific unobservable component. Regression will be based on pupils in the AM schools and their matched comparators, using the weights derived using entropy balancing. The software used to run the model is Stata.

The coefficient on the AM indicator will represent the estimated treatment effect, on an 'intention to treat basis'. For RQ2 and RQ3, the regression will be estimated on the sample of predicted AM pupils and all pupils, respectively.

***Instrumental variable: What is the impact of AM availability on PP pupils' attainment? (RQ1, RQ4)***

We will use IV techniques to provide estimates of AM that do not rely on the selection on observables assumption. This approach builds on the eligibility criteria defined in the previous section and is based on the intuition that, for schools close to the thresholds, eligibility is as good as randomly assigned. An implication of this is that, for such schools, eligibility influences AM participation but does not otherwise materially affect outcomes. This in turn suggests that eligibility can be used as an instrument in an IV analysis of the impact of AM on outcomes. Since we know some eligible schools will not participate in AM and some ineligible schools will participate, we will implement our estimator using IV (which we note is also robust to the possibility that some ineligible schools may participate). In fact, we can use two instruments in our estimator: the combination of the IDACI and AEA thresholds and the fraction of PP pupils above the average. In its simplest form, the IV estimator is simply the treatment-control difference in attainment divided by the treatment-control difference in participation. The resulting estimate is interpreted as the impact of AM availability among pupils in schools close to the eligibility thresholds.

In practice, we will operationalise this by estimating a 2-stage least squares regression of the outcome on an indicator of whether the school is eligible for AM (instrumented using the two eligibility criteria) and the same variables used in the linear regression. The coefficient on the AM coefficient will be the impact estimate.

The approach relies on the eligibility criteria having an impact on participation. If this is not the case, IV estimates can be unreliable. This is the problem of weak instruments. A preliminary analysis on the data of schools recruited in the AM intervention until March shows that AM schools are more likely than Non-AM schools to satisfy at least one of the two eligibility criteria.

Figure 1 shows eligibility on the grounds of IDACI (x-axis) and AEA (y-axis). Pink dots identify eligible schools and blue dots identify non-eligible schools. Figure 2 shows participation in AM on the grounds of IDACI (x-axis) and AEA (y-axis). Pink dots identify AM schools who are not eligible according the IDACI/AEA and according to the PP criteria and blue dots identify AM schools who are eligible according to the IDACI/AEA criterion. The graph shows that the majority of the AM schools satisfy the eligibility as defined by the two indexes.

Figure 3 shows eligibility on the ground of the second criterion, by plotting the distribution of schools according to the fraction of PP pupils. The red line at 0.25 identifies the average value of fraction of PP in the sample. The majority of schools in the non-AM sample lie on the left of the threshold, while the majority of the schools in the AM sample lie on the right of the threshold. The two graphs provide evidence that the eligibility criteria actually can predict AM participation.

Having two instruments, we can perform the Sargan overidentification test to test for the exogeneity of the instruments. The test, whose null hypothesis is that the instruments are appropriately independent of the error process, allows to evaluate the validity of the instruments.

The nature of our instruments is such that the estimator has a regression discontinuity design. For this, we require that the variables conferring eligibility -IDACI/AEA and PP proportion - are continuous at the thresholds. We also require that schools cannot manipulate their eligibility status. This requirement is likely to be met since it is determined by IDACI, AEA and PP proportion, all of which are not under the control of the schools (certainly at the time of the programme implementation).

To estimate impacts, we will regress the outcome on an AM indicator, a baseline measure of the outcome of interest and year group indicator. Pupil-level controls will include background variables, such as gender, ethnicity, EAL, FSM and special educational needs. Standard errors will be clustered at the school level to account for any common school-specific unobservable component. Regression will be based on pupils in the AM eligible and ineligible schools, using the sample weights. The software used to run the model is Stata. The coefficient on the AM indicator will represent the estimated treatment effect, on a 'treatment on the treated' basis at the school level and on an 'intention to treat' basis at the pupil level.

Year 11 impacts will be estimated for the population of secondary schools. For primary and secondary schools, we plan to augment the evaluation sample with some ineligible schools, as that would allow us to use the eligibility criteria to achieve identification of the impact of AM. To achieve this, we would include some TP and non-TP schools. Since these are not a random sample, we will construct weights to restore sample representativeness. Half of the sample will be composed by eligible schools and half by ineligible TP schools. We will choose TP schools with values of the eligibility criteria just below the thresholds. If this is proven challenging for the IDACI and AEA scores, we will select ineligible schools on the basis of the fraction of PP.

Figure 1: IDACI and AEA scores for AM (pink) and non AM (blue) schools

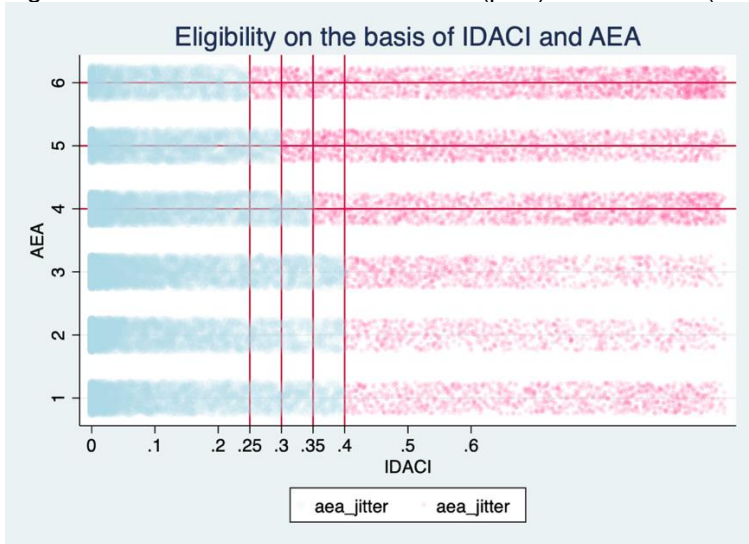


Figure 2: Participation in AM on the basis of the IDACI and AEA scores

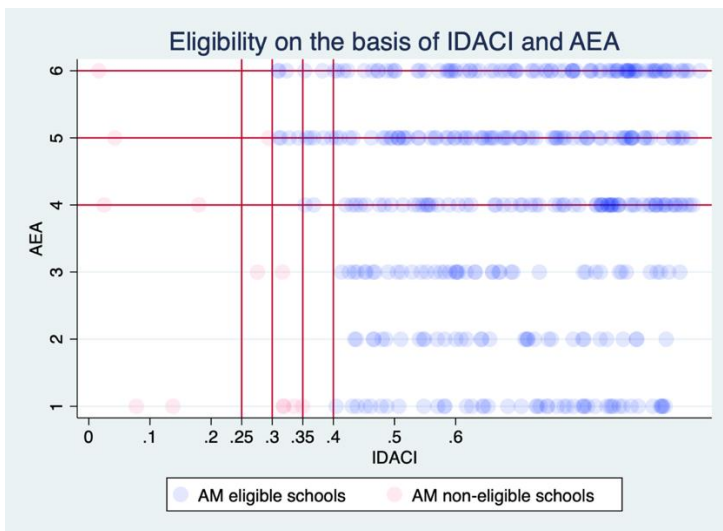
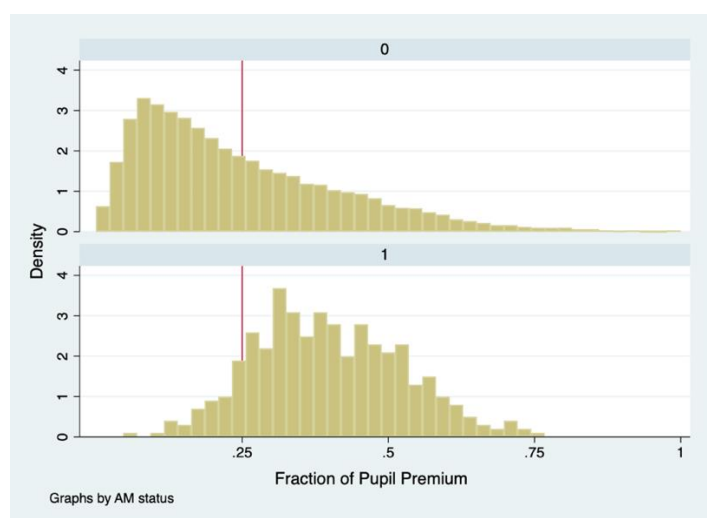


Figure 3: Distribution of PP across AM/non AM schools



### **Inference**

Uncertainty will be conveyed using confidence intervals. We note that with the linear regression impact analysis, no account is taken of uncertainty arising from the matching and weighting, which we regard as pre-processing steps. For year 11 data, we will adjust for multiple testing using the simulation approach of Clark et al. (2020), as implemented by the Stata program `rwolf.ad0`.

### **Secondary analyses**

#### ***What is the impact of AM availability on the attainment of pupils predicted to participate? (RQ2)***

Our approach to the primary analysis provides an estimate of the impact on a subgroup of the population, PP pupils, which may not coincide with the group of children who will receive the intervention. Differently from TP, that targets PP pupils, AM does not have prescribed conditions on the characteristics of pupils who receive the intervention. Preliminary data on the intervention show that about 50% of pupils receiving AM are pupil premium.

RQ2 involves an alternative approach to approximating the group of children who may receive AM. It involves modelling the probability of pupil participation in AM schools, using various markers of disadvantage recorded in the NPD and attainment (socio-economic status measured by FSM/PP, special educational needs, interaction with social service, prior attainment, English as first language and ethnicity). The results will be used to predict participation in both AM and non-AM schools in the evaluation sample. Having done this, we will follow a similar approach to that of RQ1 but instead of selecting PP pupils, we will select predicted AM pupils (where, in non-AM schools, predicted AM pupils are those who would be predicted to participate were AM available). Impacts will be estimated as (weighted regression-adjusted) comparison of outcomes among predicted eligible pupils in AM and matched non-AM schools.



### ***What is the impact of the availability of AM on all pupils' attainment? (RQ3)***

As another means of understanding the overall effect of AM, a third analysis will focus on attainment of all pupils (rather than PP pupils or predicted AM pupils) in year groups with AM in AM schools compared with all pupils in comparison year groups in non-AM schools in the evaluation sample. These estimates are likely to be smaller than RQ1 and RQ2 estimates, as the AM impact will be more diluted. This estimator also captures the effect of spill over (peer) effects. The purpose of this is to capture the overall impact of AM. The regression analysis will control for the same school level and pupil level characteristics mentioned in the primary analysis.

### ***Moderator analysis: How does the impact of AM availability vary among PP pupils, by school and pupil characteristics? (RQ5)***

Moderator analysis will be conducted through interaction terms on the following categories individually of:

1. School characteristics<sup>7</sup>: Ofsted rating (high vs. low); proportion of FSM (high vs. low); type of school (academy/maintained); school size (by quartile). These variables identify the context where AM is delivered and allow to analyse whether AM has been more effective in disadvantaged contexts.
2. Pupil characteristics: prior attainment; pupil premium (eligible vs. not); SEND vs. not; English as an additional language, ethnicity and gender. These variables identify participants and allow to analyse whether AM has been more effective on disadvantaged pupils or children with specific demographics.
3. Other: Covid-19 hotspot<sup>8</sup>. School closure due to Covid-19 may affect learning disruption at the school level, and therefore the decision to use AM.

Estimates are based on the schools in the evaluation sample. Since our estimation approach focuses on PP pupils and predicted AM pupils, the issue of compliance does not arise.

### ***How do outcomes vary among AM pupils, by model of tutoring? (RQ6)***

A descriptive analysis (using the data collected via templates for the above impact analysis) will compare outcomes associated with different tutoring models and moderators among AM schools in the evaluation sample. We do not propose any impact analysis within RQ6 since we cannot observe the counterfactual treatment model among non-AM schools. Instead, this element of the analysis will summarise mean attainment among participating pupils in AM schools according to the model of tutoring they experience. In particular, we will look at:

---

<sup>7</sup> As highlighted in the previous footnote, we may need to review/amend our approach to the school-level variables in NPD, once we know which/how many school-level variables we are able to match into the NPD.

<sup>8</sup> We aim to include this however we are looking into sources of data for this variable. We currently plan to collect this information from the government website <https://coronavirus.data.gov.uk/>. We will also investigate whether we are able to identify school closures due to Covid-19 to use as a variable in the analysis.

1. The intervention: mode of delivery of completed sessions (online vs. face to face); timing of the session (during vs. after lessons); tutor:pupil ratio (1:1 up to 1:10).
2. Tutors: QTS qualification/not; shared gender with pupil/tutee.

## Missing data

The section describes how we deal with missing values at follow up. We would not expect to find missing values at school level, as we are recruiting schools with a testing regime in place. If a school that drops from the programme is using the online repository of an assessment provider or is providing tutoring to year 11, then we could still explore assessment outcomes, unless the school also withdraws from the evaluation and requests that the data is not used. If the pupil drops out from AM, we would observe their assessment anyway. We would miss the observation for students missing the test, and it would be important to understand why the child missed the test.

We expect 15-20% of students missing the test, taking into account the possibility of children being sick or isolated on the day of the test. We will explore the extent of missingness by counting the observations for which the assessment variables are missing, and the pattern of missingness in the outcome variables.

To explore the pattern of missingness, we will run a logistic regression on the probability of dropping out at follow-up which includes individual observable background characteristics and the baseline outcomes.

The logistic model will provide evidence on whether dropping out is correlated with observable characteristics and baseline testing. The student could miss the test because of sickness or isolation, in which case baseline testing would not be correlated with the probability of drop out. We would then safely assume that the observation is missing completely at random. Focusing only on the sample of complete cases would not bias the estimates. If missingness is correlated with having achieved a low score at baseline and this factor is associated with substantive model outcome, then controlling for baseline attainment should address the issue. If low scores at baseline and other observable characteristics are associated with substantive model outcome, missing data will be imputed using multiple imputation (MI), a statistical technique which uses the distribution of observed data to estimate a set of plausible values for missing data. The missing values are replaced by the estimated plausible values by the estimation of multiple datasets. The results obtained from each dataset are combined using Rubin's rules to create a "complete" dataset (Schafer, 1999). Results with MI will be reported in addition to the headline impact estimates.

If the baseline attainment score is correlated with the probability of dropping out but does not completely explain attrition, we will assume missing not at random (MNAR). As an additional sensitivity analysis, we would employ the strategy in Carpenter *et al.*, (2007), which consists of using importance sampling to re-weighting the MI parameter estimates. The approach consists in the following procedure: impute under missing at random (MAR) and obtain parameter estimates for each imputed data set. The overall MNAR parameter estimate is a

weighted average of these parameter estimates, where the weights depend on the assumed degree of departure from MAR. Results with re-weighted MI will be reported in addition to the main impact estimates.

## Effect size calculation

Estimates will be presented as effect sizes, calculated using the Hedges' g formula. Formally, the effect sizes are calculated as follows:

$$g^* = \frac{\Gamma((n_T + n_C - 2)/2)}{\sqrt{(n_T + n_C - 2)/2} \cdot \Gamma((n_T + n_C - 3)/2)} \cdot \frac{\beta_T}{\sqrt{\frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{n_T + n_C - 2}}}$$

where  $n_T$  is the number of treatment group observations,  $n_C$  is the number of control group observations,  $\Gamma()$  is the gamma function,  $\beta_T$  is the regression coefficient on the dummy variable indicating membership of the treatment group,  $S_T^2$  is the variance of the outcome variable among the treated group and  $S_C^2$  is the variance of the outcome variable among the control group.

## Ethics

The evaluation went through ethical approval at project start up on 29th September 2020 – at a meeting where all members of the Evaluator team were present. This ethics checklist is a key process within NFER's Code of Practice (CoP), and any issues raised are escalated to CoP group. All items on the checklist met with approval and did not need to be raised. A copy of the checklist is in Appendix A. Note, at the time of writing this study plan, the University of Westminster are seeking ethical approval from their ethics committee. We will update the study plan with the outcome of this as soon as this is known, and address any issues accordingly.

All participants take part in the evaluation activities with informed consent.

All participants (parents, and KS4 pupils, tutors, school staff and AM staff) are provided with a privacy notice relevant to processing their (or their child's) data. Participants can withdraw from data processing at any time during the evaluation – and instructions are provided in the privacy notice for how to inform their school, Teach First and/or Evaluator that they do not wish their data to be processed.

## Data protection

### *Data protection statement and GDPR compliance*

The Evaluator will be compliant with the Data Protection Act 2018 (DPA) and General Data Protection Regulation (GDPR). NFER has ISO27001 and Cyber Essentials Plus certifications and registration with the Information Commissioner's Office. Other members of the consortium have equivalent accreditations to demonstrate their compliance with DPA and GDPR.

To carry out the evaluation, it will be necessary to use and share personal data about pupils (both those who take up the offer and those who do not), as well as key staff members at participating schools and AM staff delivering the mentoring, so that they can be asked about delivery.

The Evaluator has put in place appropriate measures to prevent pupils' personal information from being accidentally lost, used or accessed in an unauthorised way, altered or disclosed. In addition, each organisation involved will limit access to pupils' personal information to their staff members who have a business need to see it. Any data shared between the school, Teach First, EEF, the Evaluator and DfE will be via secure portal.

### *Legal bases*

To make the use of pupils' data in the evaluation lawful, the Evaluator has identified specific grounds, known as a legal basis, for its processing. The legal basis available depends on the type of organisation, and these are outlined below.

EEF and NFER have identified the following legal basis for processing personal data:

GDPR Article 6 (1) (f) which states:

*Legitimate interests: the processing is necessary for your (or a third party's) legitimate interests unless there is a good reason to protect the individual's personal data which overrides those legitimate interests.*

We have carried out a legitimate interest assessment, which demonstrates that the evaluation fulfils the Evaluator's core business purposes (undertaking research, evaluation and information activities). It has broader societal benefits and will contribute to improving the lives of learners by providing evidence for about the most effective ways of providing catch-up tuition. The evaluation cannot be done without processing personal data but processing does not override the data subject's interests.

The University of Westminster have identified the following legal basis:

GDPR Article 6 (1) (e) which states:

*Public task: the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law.*

A separate legal basis is identified for processing special data. The legal basis for processing special data for the evaluation of Teach First is:

GDPR Article 9 (2) (j) which states:

*Archiving, research and statistics (with a basis in law): processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.*

### **Linking to NPD and use of Secure Research Service (SRS)**

NFER will securely submit the pupil data to the National Pupil Database (NPD) team to be matched to the pupil data held on NPD. The University of Westminster will access the matched NPD data for analysis through the SRS secure online system. The SRS system does not allow users to remove or copy data from its servers.

The project meets the Office for National Statistics "five safes" in the following ways:

- Safe people: all researchers accessing the project's data via the SRS are Accredited Researchers and hold a 'basic disclosure' certificate that is no more than 2 years old
- Safe projects: the project meets the conditions for accessing personal level data. A full request to the NPD team will be submitted, outlining the appropriate and ethical use of the data, and the public benefit of the research (to contribute to the evidence base on tutoring, and inform future tutoring programmes). It has broader societal benefits and will contribute to improving the lives of learners by providing evidence about the most effective ways of providing catch-up tuition. The evaluation cannot be done without processing personal data but processing does not override the data subject's interests. The research team and the EEF are committed to publishing the results of the study.

- Safe settings: all researchers working on the NPD data will only access the data via the SRS secure online system. Our organisations will apply for safe room connectivity to have SRS remote connectivity access.
- Safe outputs: All outputs will be checked by the ONS team to ensure that the outputs do not allow identification of individuals. Outputs will be checked against the Intended Permitted Outputs and be subject to standard ONS disclosure rules.

Safe data: the data request includes data variables of identifiability risk level 3 (PMR), as the DfE will match the data we collect with the NPD data. The PMR (meaningless identifier) replaces the UPN when the data are matched and then archived to minimise the risks of identification. Our researchers will only analyse de-identified data in the SRS.

The Parent (and KS4 Pupils with wording suitably adjusted) Privacy Notices contain the following information about personal data collection and linking to NPD

- Teach First will collect some personal data about your son/daughter directly from their school, including name, date of birth, UPN, year group. They will also record any attendance at mentoring sessions.
- The Evaluator will also collect pupil background details and assessment data from the school's commercial test provider. The Evaluator will use your son/daughter's UPN to obtain further background information (for example their gender, ethnicity, household proximity to school, eligibility for pupil premium, free school meals, English as an additional language, has a social worker or is a looked after child, has special educational needs, a disability or has fallen behind or is at risk of falling behind in their school work) from the NPD.
- No individual will be named in any report for this project. Pupils' personal data will be shared between the organisations mentioned in this privacy notice. The school will provide Teach First with information about your child. Teach First will share your child's data with the Evaluator. The Evaluator will be using a secure online portal to collect pupil data electronically. Your child's full name, date of birth and UPN will be shared with the NPD team to request their background characteristics.
- If data collected for the evaluation of the AM programme is to be used in other COVID-19 related research, it will be shared with the research organisations appointed to carry out that research.
- Three months after the publication of the evaluation report, all of the pseudonymised matched data (pupil data only) will be added to the EEF archive, which is managed by FFT on behalf of EEF and hosted by the ONS. This will enable the EEF and other research teams to use the pseudonymised data as part of subsequent research through the ONS Approved Researcher Scheme, including analysing long-term outcomes through the National Pupil Database. This data may also be linked to other research datasets for the purpose of Covid-19 related educational research. Further information about the EEF archive is available from: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluatingprojects/evaluator-resources/archiving-evaluation-data/>

### ***Rights and retention periods***

Parents (and KS4 pupils) can withdraw their child from the programme and/or from their data being processed, until it is added to the EEF archive. Should they withdraw from the programme or evaluation (i.e. decide not to engage with Academic Mentors or the evaluation), the Evaluator will still use the evaluation data that the school has provided up to that point and link it to NPD unless the parent/KS4 pupil indicates otherwise. If at any time, parents/KS4 pupils wish to withdraw their data or have errors corrected in it, contact details are provided in the Privacy Notices for who to contact about this.

As noted above, three months after the publication of the evaluation report, all of the pseudonymised matched data (pupil data only) will be added to the EEF archive, which is managed by FFT on behalf of EEF and hosted by the ONS. This will enable the EEF and other research teams to use the pseudonymised data as part of subsequent research through the ONS Approved Researcher Scheme, including analysing long-term outcomes through the National Pupil Database. This data may also be linked to other research datasets for the purpose of Covid-19 related educational research.

The Evaluator will securely delete any personal data relating to the evaluation one year after the publication of the final report, currently expected to be April 2022.

Teach First will securely delete any personal data collected for the evaluation alone at the end of the AM programme, when final grants have been paid (expected to be August 2021).

Teach First may keep personal data collected as part of the delivery of their mentoring services for longer – this is covered in the privacy notice they provide. Once data has been archived, it is held in the EEF archive until it is no longer needed for research purposes.

### ***Data controller and processing roles***

The Department for Education (DfE), the EEF and the Evaluator are joint data controllers for the evaluation. They decide how and what data will be collected and used. The Evaluator is also a data processor, as are Teach First. (Note Teach First are also a joint data controller in regard to data associated with the programme. This study plan is concerned with the evaluation.)

## Personnel

Name	Institute	Roles and responsibilities
<b>Pippa Lord</b>	NFER	Project Director and Consortium Lead – responsible for directing the Consortium and quality of delivery.
<b>Kathryn Hurd</b>	NFER	Workstream lead – responsible for overseeing data management, evaluation and comparison school recruitment, school contacting and testing
<b>Kinnery Koria</b>	NFER	Project manager – responsible for overseeing the day-to-day running of the operations of the project
<b>Rachel Hayes</b>	NFER	Project Leader for Impact workstream
<b>Veruska Oppedisano</b>	University of Westminster	Statistician and impact evaluation design
<b>Richard Dorsett</b>	University of Westminster	Overseeing impact evaluation design
<b>Greta Morando</b>	University of Westminster	Analyst on the impact evaluation
<b>Alice Phillips</b>	Department of Education	Impact evaluation design
<b>Kim Williams</b>	Department of Education	Impact evaluation design
<b>Arnaud Vaganay</b>	Department of Education	Impact evaluation design
<b>Ben Styles</b>	NFER	Impact QA



## Risks

Table 5: Evaluation issues and risks

Risk	Assessment	Controls, countermeasures and contingencies
<p>The number of schools who consent to RL data share will be too small to form a representative sample for meaningful analysis</p>	<p><b>Likelihood:</b> Moderate <b>Impact:</b> High</p>	<p>We have delayed the impact evaluation by one month to allow schools until the end of September 2021 to consent to the data share within the RL online portal. RL have informed us that early in the academic year is when they see the most activity from schools and this is likely to mean more schools will see the in-product notifications. NFER have agreed to support this process by using some resource to telephone schools mid-September to improve sign-up.</p>
<p>Evaluation data is not able to be matched or is matched incorrectly (across datasets)</p>	<p><b>Likelihood:</b> High <b>Impact:</b> High</p>	<p>Data sharing agreement between TF and RL does not include pupil names. As well as the RL test scores, the school name, postcode and URN will be provided along with the DOB for every child. RL has also requested UPN, year group, age and gender for each child, but these are not mandatory variables. Therefore there is a concern that matching this data into the NPD will be challenging. We intend to over-recruit schools to mitigate this risk.</p>

## Timeline

Date	Activity	Responsible/leading
<b>Oct 2020</b>	Project set up, logic model development, materials development, study plan development	Evaluator
<b>Oct 2020 – July 2021</b>	<i>Mentoring period (whole programme)</i>	AMs
<b>5<sup>th</sup> January – 8<sup>th</sup> March 2021</b>	National lockdown period – many pupils learning from home, schools only open to children of keyworkers and vulnerable children. AM provision moved to online only.	
<b>April 2021</b>	Submit NPD request	UoW
<b>Mid May to early July 2021</b>	GCSEs (year 11) <i>Also window for end-point testing (+ to end of summer term)</i>	Schools
<b>Summer 2021</b>	Study plan finalisation and publish	Evaluator
<b>Mid August to early Sept 2021</b>	Data cleaning (MI/pupil data)	NFER
<b>September 2021</b>	Assessment data from RL sent to Teach First and shared with DfE	Teach First
<b>November 2021</b>	Study plan publication, to match with revised TP study plan publication	Evaluator
<b>Mid October 2021</b>	Draw comparison sample and placebo check	UoW
<b>Mid October 2021</b>	Run placebo check again on the selected comparison sample (weighting will be applied if required)	UoW
<b>November 2021</b>	Impact analysis	UoW
<b>November 2021</b>	NPD (unamended) data available and matched into dataset	NPD team/ UoW
<b>December 2021 to January 2022</b>	Write first draft of report	All
<b>February 2022</b>	Final report/report revisions	All

## References

Allen, R., Jerrim, J., Parameshwaran, M. and Thompson, D., 2018. *Properties of commercial tests in the EEF database*. [online] Available at: <[https://educationendowmentfoundation.org.uk/public/files/Support/EEF\\_Research\\_Papers/Research\\_Paper\\_1\\_-\\_Properties\\_of\\_commercial\\_tests.pdf](https://educationendowmentfoundation.org.uk/public/files/Support/EEF_Research_Papers/Research_Paper_1_-_Properties_of_commercial_tests.pdf)> [Accessed 5 Aug. 2021].

Carpenter, J.R., Kenward, M.G. and White, I.R., 2007. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Statistical Methods in Medical Research*, 16(3), pp.259–275. <https://doi.org/10.1177/0962280206075303>.

Chi, M.T.H., Siler, S.A., Jeong, H., Yamauchi, T. and Hausmann, R.G., 2001. Learning from human tutoring. *Cognitive Science*, 25(4), pp.471–533. [https://doi.org/10.1207/s15516709cog2504\\_1](https://doi.org/10.1207/s15516709cog2504_1).

Clarke, D., Romano, J.P. and Wolf, M., 2020. The Romano–Wolf multiple-hypothesis correction in Stata. *The Stata Journal*, 20(4), pp.812–843. <https://doi.org/10.1177/1536867X20976314>.

Coe, R., 2020. *Impact of school closures on the attainment gap: Rapid Evidence Assessment*. [online] London, UK: Education Endowment Foundation. Available at: <[https://dera.ioe.ac.uk/35707/1/EEF\\_%282020%29\\_-\\_Impact\\_of\\_School\\_Closures\\_on\\_the\\_Attainment\\_Gap.pdf](https://dera.ioe.ac.uk/35707/1/EEF_%282020%29_-_Impact_of_School_Closures_on_the_Attainment_Gap.pdf)>.

Collins, A. and Stevens, A., n.d. Goals and methods for inquiry teachers. In: R. Glaser, ed. *Advances in instructional psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cullinane, C. and Montacute, R., 2020. COVID-19 and Social Mobility Impact Brief #1: School Closures. [online] Available at: <<https://www.suttontrust.com/our-research/covid-19-and-social-mobility-impact-brief/>>.

Dietrichson, J., Bøg, M., Filges, T. and Jørgensen, A.-M.K., 2017. Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis: <https://doi.org/10.3102/0034654316687036>, 87(2), pp.243–282. <https://doi.org/10.3102/0034654316687036>.

Education Endowment Foundation, 2018a. *One to one tuition | Toolkit Strand | Education Endowment Foundation | EEF*. [online] Available at: <<https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/one-to-one-tuition/>> [Accessed 5 Aug. 2021].

Education Endowment Foundation, 2018b. *Small group tuition | Toolkit Strand | Education Endowment Foundation | EEF*. [online] Available at: <<https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/small-group-tuition/>> [Accessed 5 Aug. 2021].

Elbaum, B., Vaughn, S., Hughes, M.T. and Moody, S.W., 2000. How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92(4), pp.605–619. <https://doi.org/10.1037/0022-0663.92.4.605>.

Every Child a Chance Trust, 2009a. *Every Child a Reader. The first year of the national rollout*. London, UK: Every Child a Chance Trust.

Every Child a Chance Trust, 2009b. *Every Child Counts. The results of the first year 2008/09*. London, UK: Every Child a Chance Trust.

Fantuzzo, J.W., King, J.A. and Heller, L.R., 1992. Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology*, 84(3), pp.331–339. <https://doi.org/10.1037/0022-0663.84.3.331>.

Hodgen, J., Taylor, B., Jacques, L., Tereshchenko, A., Kwok, R. and Cockerill, M., 2020. *Remote mathematics teaching during COVID-19: intentions, practices and equity*. [Report] UCL Institute of Education: London, UK. London, UK: UCL Institute of Education. Available at: <<https://www.ucl.ac.uk/ioe/>> [Accessed 5 Aug. 2021].

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E. and Liu, J., 2020. Projecting the potential impacts of COVID-19 school closures on academic achievement. *Educational Research*, 49(8), pp.549–565. <https://doi.org/10.3102/0013189X20965918>.

Leinhardt, G. and Greeno, J.G., 1986. The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), pp.75–95. <https://doi.org/10.1037/0022-0663.78.2.75>.

Lepper, M.R., Drake, M.F. and O'Donnell-Johnson, T., 1997. Scaffolding techniques of expert human tutors. In: *Scaffolding student learning: Instructional approaches and issues*., Advances in learning & teaching. Cambridge, MA, US: Brookline Books.pp.108–144.

Lepper, M.R. and Woolverton, M., 2002. Chapter 7 - The Wisdom of Practice: Lessons Learned from the Study of Highly Effective Tutors. In: J. Aronson, ed. *Improving Academic Achievement: Impact of Psychological Factors on Education*, Educational Psychology. [online] San Diego: Academic Press.pp.135–158. <https://doi.org/10.1016/B978-012064455-1/50010-5>.

McArthur, D., Stasz, C. and Zmuidzinas, M., 1990. Tutoring Techniques in Algebra. *Cognition and Instruction*, 7(3), pp.197–244.

Merrill, D.C., Reiser, B.J., Ranney, M. and Trafton, J.G., 1992. Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *Journal of the Learning Sciences*, 2(3), pp.277–305. [https://doi.org/10.1207/s15327809jls0203\\_2](https://doi.org/10.1207/s15327809jls0203_2).

Mischo, C. and Haag, L., 2002. Expansion and Effectiveness of Private Tutoring. *European Journal of Psychology of Education*, 17(3), pp.263–273.

Moss, G., 2020. Reasons to be cautious about lockdown learning loss estimates. *Schools Week*. [online] Available at: <<https://schoolsweek.co.uk/5-reasons-to-be-cautious-about-estimates-of-lockdown-learning-loss/>>.

Nickow, A., Oreopoulos, P. and Quan, V., 2020. The Impressive Effects of Tutoring on Prek-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. *National Bureau of Economic Research working paper no. w27476*, 20(20), pp.267–350. <https://doi.org/10.3386/w27476>.

Pintrich, P.R. and Schunk, D., 2002. *Motivation in education: Theory, research, and applications*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Ritter, G.W., Barnett, J.H., Denny, G.S. and Albin, G.R., 2009. The Effectiveness of Volunteer Tutoring Programs for Elementary and Middle School Students: A Meta-Analysis. *Review of Educational Research*, 79(1), pp.3–38. <https://doi.org/10.3102/0034654308325690>.

Rogoff, B., 1990. *Apprenticeship in Thinking. Cognitive Development in Social Context*. New York: Oxford University Press.

Schafer, J.L., 1999. Multiple imputation: a primer. *Statistical Methods in Medical Research*, 8(1), pp.3–15. <https://doi.org/10.1177/096228029900800102>.

Sharp, C., Nelson, J., Luca, M., Julius, J., McCrone, T. and Sims, D., 2020. *Schools responses to Covid-19: The challenges facing schools and pupils in September 2020 - NFER*. [online] Available at: <<https://www.nfer.ac.uk/schools-responses-to-covid-19-the-challenges-facing-schools-and-pupils-in-september-2020/>>.

Smyth, E., 2008. The More, the Better? Intensity of Involvement in Private Tuition and Examination Performance. *Educational Research and Evaluation*, 14(5), pp.465–476.

Torgerson, C.J., Bell, K., Coleman, E., Elliott, L., Fairhurst, C., Gascoine, L., Hewitt, C.E. and Torgerson, D.J., 2018. Tutor Trust: Affordable Primary Tuition Evaluation report and executive summary. [online] Available at: <[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Tutor\\_Trust.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Tutor_Trust.pdf)>.

Weidmann, B. and Miratrix, L., 2021. Lurking inferential monsters? Quantifying selection bias in evaluations of school programs. *Journal of Policy Analysis and Management*, 40(3), pp.964–986. <https://doi.org/10.1002/PAM.22236>.

Appendix A: Code of practice and ethics approval checklist

Section of Code of Practice	Consideration of Code of Practice (CoP)	Yes	No	N/A
<b>Ethics</b>	Level of consent required – does the project allow for the level of consent required?	✓		
	Will research participants be provided with all the required information to enable them to make an informed choice?	✓		
	Have you looked at and do you intend to follow the guidance on selecting children/young people for interview?	✓		
	Will you follow the protection and safety guidelines?	✓		
	If the project involves children/young people have all those involved undergone disclosures/child protection training?	✓		
<b>Data protection</b>	Will the project follow the 8 principles of the data protection act?	✓		
	Will the project follow the rules for the processing of sensitive personal data?	✓		
<b>Data security</b>	Will the project allow for safe transfer of data into and out of our systems?	✓		
	Will the project include a secure coding system for recording participants' names?	✓		
	Have data transfer issues / protocols been discussed / confirmed with the client?	✓		
<b>Caring for research participants</b>	Will the project take into account designing research questions that make sense to children/young people?	✓		
	Will the project follow the guiding principles for the development of assessment instruments, methods and systems? ( <u>Will only use standardised tests which we believe satisfy requirements</u> )	✓		
	Will the project involve taking, producing and using visual images? (Please refer to points to consider when taking photographs or video images, storing images, producing illustrations and using visual images)		✓	✓