Department for Education

Independent Evaluation of the National Tutoring Programme Year 2: Impact Evaluation

Research report

October 2023

Authors: Megan Lucas, Emma Moore, Chris Morton, Ruth Staunton and Stephen Welbourne: NFER

Government Social Research

Acknowledgements

The authors would like to thank Rushda Khandker, Christine Gilliver, Aoife Bermingham, Michael Scott, Emily Severn and Clare Baker from DfE for their project management, and guidance. They would also like to thank the National Tutoring Programme policy team for their contribution.

The authors would also like to thank Sarah Lynch at NFER for her project management. They are also grateful to Ben Styles and Carole Willis for their comments on the draft report and to Emma Hawkins for her administrative support.

Contents

List of figures

List of tables

Executive Summary

Introduction

The Covid-19 pandemic and subsequent closures of schools to most pupils caused significant disruption to pupils' education and learning, with disadvantaged pupils most severely affected. The National Tutoring Programme (NTP) was introduced in the 2020-21 academic year to help disadvantaged pupils in England recover lost learning and reduce the attainment gap. In the second year of the programme (2021-22), which is evaluated here, the NTP offered three routes of support:

Tuition Partners (TP): this route offered subsided tuition (70%) to schools from approved tuition partners.

Academic Mentors (AM): this route supported the most disadvantaged schools. Schools qualified for AM if their percentage of pupil premium (PP) eligible pupils was 20% or more[1], or if they were located in areas where educational standards are considered low (Priority Area for Raising School Standards or Opportunity Area). Academic mentors were employed by the school, and in 2021-22, 95% of their core salary cost was subsidised by the DfE.

School-Led Tutoring (SLT): this was a new NTP route offered in 2021-22 and provided schools with a ring-fenced grant to fund locally sourced tutoring provision. The SLT route was subsidised at 75% in the 2021-22 academic year.

Evaluation Aims

Primary aim: to evaluate the impact of the second year (2021-22) of the NTP on educational attainment for a) all pupils, and b) for PP pupils and/or pupils with prior low attainment (PLA) defined as having achieved lower than the expected standard at KS1 or KS2 (as applicable).

In addition, this evaluation also sought to understand:

- whether the impact of the NTP varied according to tutoring route, geographic region, dosage (the number of hours of tutoring pupils received), and concentration (percentage of pupils in a school that were selected for tutoring)
- whether there were differences in the impact of the NTP for pupils with different characteristics (e.g., special educational needs (SEN) status, gender, English as an additional language (EAL))

[1] Note that this was initially 30% or more but was later reduced to 20%.

- the longer-term impact of the TP and AM routes (introduced in the first year of the NTP in 2020-21).

Evaluation methods

For most analyses, we evaluated the impact of SLT on its own and AM/TP routes combined (because fewer schools used these routes compared with SLT). The exception to this was the analysis designed to explore the effects of different tutoring routes, where SLT, AM and TP were considered separately. For AM/TP, our analysis also considered the subject in which tuition was received, but this was not possible for SLT as subject data was not available.

Outcome measures

To evaluate the impact of the NTP on English and maths outcomes we used KS2 (reading and maths) and KS4 (English language and maths) scores obtained from the National Pupil Database. This meant that our analysis consisted of all mainstream primary and secondary schools in England and evaluated the impact of the NTP on English and maths outcomes for a) all Year 6 and Year 11 pupils and b) Year 6 and Year 11 PP and/or PLA pupils.

School and pupil inclusion definitions

Intervention schools were defined as having at least one pupil (or one PP and/or PLA pupil for the PP and/or PLA analysis) selected to receive tutoring in the relevant route and, where available, subject in 2021-22. This meant we evaluated the effects of three different interventions: SLT (any subject), AM/TP literacy and AM/TP maths.

For each intervention the pool of potential comparison schools was all schools not taking part in that specific intervention i.e., for intervention schools participating in AM/TP English, comparison schools were all schools not taking part in AM/TP English. Both intervention and comparison schools could also have been involved in other NTP interventions[2] (e.g., some intervention and comparison schools for the evaluation of AM/TP English might also have taken part in AM/TP maths) and this involvement in 'other routes' was included as part of the matching criteria and controlled for in the statistical analysis.

Two pupil inclusion definitions also applied to the analysis:

[2] Note that both intervention and comparison schools could also have been involved in other catch-up interventions alongside the NTP, but we were unable to account for this in any of our analysis.

- **School-level:** this approach compared the progress made by a) all pupils or b) PP and/or PLA pupils who attended[3] schools that participated in the relevant NTP route and subject with a) all pupils or b) PP and/or PLA pupils who attended comparison schools as applicable.
- **Pupil-level:** as the NTP is a pupil-level intervention, it should ideally be analysed at a pupil-level. To do this, our pupil-level analysis was restricted to pupils who had previously been selected for tutoring in the first year of the NTP (i.e., in 2020-21). Among these pupils, we compared the progress made by pupils who in the second year of the NTP (2021-22) attended intervention schools and were themselves selected for tutoring, with pupils who in the second year of the NTP attended comparison schools and therefore did not receive tutoring[4]. As with the school-level analysis, we undertook the pupil-level analysis for a) all pupils and b) PP and/or PLA pupils.

  This approach was designed to minimise the potential for selection bias which is often an issue in the evaluation of pupil-level interventions. In this case, the focus of the intervention was on supporting pupils to recover lost learning so it was expected that teachers would select pupils who were falling behind to receive tutoring. We would therefore expect to see potential negative selection bias within the evaluation, as without the intervention the pupils selected for tutoring would be expected to do 'less well' than their peers who were not selected for tutoring. Furthermore, if selection for tutoring is based on variables that are not available to the evaluation (e.g. recent performance in class tests of English or maths) then it would not be possible to identify an appropriate comparison group to control for this negative bias. We aimed to mitigate this potential bias by only analysing pupils who had been selected for tutoring in the previous year. Our approach means that both control and intervention pupils were recently selected for tutoring (in the first year of the NTP) so they are likely to be similar regarding any unobserved confounding variables, without the need to explicitly measure and model those confounders. This increases the likelihood that impact estimates observed in year 2 of the NTP at the pupil-level are due to the actual benefit of tutoring, and not the result of selection bias. However, we must acknowledge that this approach is not infallible: selection for a further year of tutoring may itself entail some selection bias (e.g., pupils that struggled in year 1 of the NTP may have been more likely to have further tutoring). In summary, this approach is likely to reduce the selection bias, but probably not completely remove it. As with the school-level analysis we

[3] Note this approach means that pupils who attended a school participating in the NTP are included within the intervention group even if they were not selected for tutoring themselves.
[4] Pupils who were selected for tutoring in the first year of the NTP and in the second year of the NTP attended intervention schools but were not themselves selected for tutoring were not included in this analysis.

undertook statistical matching using entropy balancing (described below) to ensure the groups were well matched on key observable characteristics.

Statistical analysis

Impact of the NTP: mixed effects linear regression models (which included weights to balance the groups on key characteristics) were used to analyse the impact of SLT and AM/TP on English and maths outcomes at school level and pupil level. This resulted in a total of eight mixed effects linear regression models.

Differences in school-level impact by implementation factors: this analysis investigated if the impact of the NTP at school level varied according to dosage (average number of hours of tutoring pupils received at school level), concentration (percentage of pupils within a school that received tutoring), tutoring route (SLT, TP or AM) and geographic region. To do this, we used a similar analytical approach to that used to investigate the impact of the NTP at school level (described above), adding each of these variables in turn to the regression model.

Differences in school-level impact by pupil characteristics: to investigate any differences in the impact of the NTP at school level based on pupil characteristics (PP/PLA status, gender, SEN status, EAL status and ethnicity), we repeated the linear regression models used for the impact analysis (described above) but replaced the intervention indicator with a variable indicating both intervention status and the relevant characteristic. This effectively allowed a subgroup analysis to be performed for each subgroup defined by the characteristics (e.g., males).

Analysis of the longer-term impact of the NTP: this analysis used the same population of pupils (i.e., pupils were in Year 6 or Year 11 in the 2021-22 academic year) and analysis approach as the impact analysis described above. However, this analysis considered whether their school had participated in the NTP in the previous academic year (2020-21), when the pupils were in Year 5 or Year 10. This means that at least one year will have passed between their school participating in the NTP and measurement of the KS2 or KS4 outcomes, allowing us to explore any longer-term benefits.

Limitations

Without randomisation, any evaluation of the NTP will have several limitations. The key limitations described below provide important context in which to consider the evaluation results and the conclusions we can draw.

- Schools chose which pupils received tutoring which makes it very likely that there is a high level of selection bias among tutored pupils, as we would generally expect schools to select pupils with the most acute needs for tutoring. It is also likely that there are unobserved factors which influenced schools' decisions to

select or not select pupils for tutoring. For example, some pupils who could potentially have benefited from tutoring may not have been selected due to concerns about non-attendance. In addition as statutory attainment data is only available at the end of Key Stages, we do not have a full understanding of pupils' attainment immediately prior to the introduction of the NTP. We anticipate that recent test performance is likely to have formed a major part of the decision as to which pupils should receive tutoring and it is likely that matching pupils on the basis of the most recent statutory test will only partially control for this. This makes it almost impossible to identify an appropriate comparison group to undertake robust analysis at a pupil level.

- The school-level analyses help to reduce issues related to selection bias (described above) but not all pupils (or PP and/or PLA pupils where applicable) in intervention schools were themselves selected for tutoring, meaning that the results of the school-level analyses are subject to dilution. For the school-level analyses for all pupils, the proportion of intervention pupils who were themselves selected for tutoring in the relevant route and for AM/TP subject ranged from 21% to 35%5 at KS2 and from 7% to 39%5 at KS4. In addition, no subject data was available for SLT, which means that the pupil-level analysis for SLT will also be affected by dilution.
- Our statistical models accounted for the presence of alternative NTP routes in both intervention and comparison schools by looking at the additional impact of a route compared to schools who were not using this route, controlling for the amount of tutoring that both intervention and comparison schools were providing via other routes. This means that the analysis explored the additional contribution of each NTP route, and therefore assumes that additionality is possible. In addition, we were not able to account for the presence of any other types of support (including private tutoring) pupils in either the intervention or comparison groups may have received.
- Even though our pupil-level analysis is restricted to pupils who were selected for tutoring in the first year of the NTP, there may still be some selection bias as schools chose whether to participate in the second year of the NTP, but we expect this to be reduced relative to the selection bias at the aggregate pupil level. Selection bias may also be present in the school-level analyses although in the light of recent work by Weidmann and Miratrix (2020) to try to quantify the magnitude of selection bias within non-randomised school-level analyses we consider this less of a risk.
- The analyses used multiple simultaneous tests but are not corrected for multiple comparisons. We note the population we investigated here represents the entire

[5] Note that these figures are for SLT where no data about the tutored subject was available. This means that the dilution rate for pupils who received SLT in a specific subject is likely to be higher than this figure suggests.

population of interest, so we do not need to rely on statistical inference to generalise our findings.

Key Findings

We found a consistent pattern of evidence at both school and pupil level to suggest that participation in SLT was associated with small improvements in KS2 and KS4 maths outcomes. There was also some more limited evidence at school level only that participation in SLT was associated with small improvements in KS2 and KS4 English outcomes. For both English and maths some of these results reached statistical significance, but in all cases the effect sizes were very small and equated to one months' additional progress or less. Although we note that due to dilution effects (see limitations) these effect sizes will be underestimates of the actual impact of SLT tutoring. In addition, our pupil-level analysis compared the impact of having tutoring in Year 1 of the NTP with having tutoring in both Year 1 and Year 2, meaning that the difference between these groups is likely to be smaller than the impact of having tutoring compared with not having tutoring.

We did not detect any evidence in the main impact evaluation that participation in AM/TP led to improvements in KS2 or KS4 English or maths outcomes at either school or pupil level. In some cases, participation in AM/TP was associated with negative effects on English (KS2 and KS4) and maths (KS4) outcomes. However, the very small effect sizes seen mean we cannot exclude the possibility that these results reflect selection bias.

The pattern of results observed indicates that SLT may be more successful for improving English and maths outcomes than AM/TP. However, this result needs to be considered cautiously as the effect sizes detected are small.

The effects of tutoring on KS2 and KS4 maths and English outcomes were similar for all pupils regardless of PP and/or PLA status.

A higher tutoring dosage and/or concentration was associated with better English and maths outcomes for SLT but this was not the case for AM/TP. However, it must be noted that these associations may not be causal given that fresh statistical matching of schools was not carried out for this analysis (i.e., matching was not carried out between schools with different tutoring dosages and/or concentrations).

PP pupils made up 29% of our cohort of Year 6 pupils but 46% of the Year 6 pupils who were selected for tutoring. In Year 11, PP pupils made up 26% of the pupil cohort overall, but 35% of the pupils selected for tutoring. This indicates that PP pupils were selected for the NTP at higher rates when compared with the pupil

cohort overall, but PP pupils still made up less than half of pupils selected for tutoring in both year groups.

The impact of the different NTP routes was broadly similar across pupils with different characteristics.

We did not detect any longer-term benefits of receiving tutoring in the first year of the NTP. This is consistent with the main analysis from the impact evaluation of the first year of the NTP and is also consistent with the AM/TP results reported here for the second year of the programme. It is, however, worth noting that the year 1 impact evaluation did detect some positive effects of tutoring for schools with higher concentrations of tutoring.

Conclusions

Overall, we found consistent evidence that SLT was effective for improving KS2 and KS4 English and maths outcomes, though the effect sizes were small. We did not find any evidence that participation in AM/TP tutoring resulted in improved outcomes at a school level, and we found some evidence to suggest that participation in the AM/TP route was associated with slightly worse outcomes in English (KS2 and KS4) and maths (KS4). However, given the potential for selection bias to remain despite our extensive attempts to remove it, the small effect sizes observed may be artifacts of this bias rather than genuine effects of the intervention on outcomes.

That said, it is important to consider the trends observed in this analysis in case they are genuine effects. The differences in the pattern of results between different tutoring routes may be due to differences in implementation. The SLT route was introduced in the second year of the NTP as schools called for more control over how they delivered tutoring. SLT allowed schools to use internal staff as tutors making it more likely the staff delivering the NTP had existing relationships with staff and pupils they could build on and that tutoring was delivered in-person rather than online. This may help explain why the implementation and process evaluation of the second year of the NTP showed that the highest levels of satisfaction amongst school leaders were for SLT (Lynch *et al.*, 2022). This extended to all aspects of tutoring including the ability of tutors to meet pupils' learning needs, how well tutoring sessions aligned with the school curriculum, tutors' relationships with pupils, and the quality of tuition, all of which may have contributed to better outcomes for pupils.

In general, the effect sizes observed in this analysis were smaller than might be anticipated based on previous evidence about the effectiveness of small group tuition (Ritter *et al.*, 2009; Dietrichson *et al.*, 2017; Nickow, Oreopoulos and Quan, 2020; EEF, 2021b). This is likely to be due to differences in the evaluation design and available data, and the subsequent limitations of the analysis approach which was feasible as well as

difficulties implementing the NTP at scale, rather than tutoring being ineffective as an approach. Here the school-level analyses were subject to dilution effects as not all pupils within the intervention group were themselves selected for tuition. In addition, no data about the subject in which tutoring was received was available for SLT. This means that for SLT there are also dilution effects present at the pupil level so our estimates of the SLT-related effect sizes will be underestimates of the true effect. It is also likely that challenges with delivering the NTP at scale including recruiting high-quality tutors, integrating tutoring with the school curriculum, and tutoring often being conducted during normal lesson times (rather than in addition to usual teaching and learning) (see Lynch *et al.*, 2022) as well as the relatively small number of tutoring hours may be impeding the effectiveness of the NTP, resulting in the NTP having lower-than-anticipated observed effects.

For SLT a higher tutoring dosage and/or concentration was consistently associated with better outcomes at a school level, but this was not the case for AM/TP. While this suggests that for SLT providing more hours of tutoring to pupils can lead to better outcomes at a school level, these associations may not be causal as fresh matching was not carried out for the analyses. To maximise the benefits of tutoring, it is important to increase the evidence base around best-practice for implementing tutoring in schools.

The key aim of the NTP was to help reduce the attainment gap for disadvantaged pupils. Here we found that SLT was associated with small positive impacts on English and maths outcomes. However, although PP pupils were selected for the NTP at higher rates than would be expected relative to the population, less than half of pupils selected to receive tutoring within our evaluation cohort were PP. This suggests that, although there is evidence that schools are prioritising PP pupils for the NTP, more could be done in terms of targeting this group especially in Year 11 where PP pupils only made 35% of pupils selected for tutoring.

Recommendations

- There is tentative evidence that the introduction of the SLT route has been successful. Further research into how to optimise the delivery of tuition is needed to be able to offer guidance to schools on which type of implementation is most effective. For example, this could include research using randomised controlled trial designs (RCTs) to build the evidence base around best practice in tutoring optimum tutoring dosage, session duration, frequency, mode of delivery (online versus in-person), how best to align sessions with the school curriculum and time of delivery (during the school day or outside of normal teaching hours).

- Future research on tutoring should collect data on tutoring routes and subject, to allow for continued monitoring of effectiveness by route and subject to further develop our understanding of ‘what works’.

- To help close the attainment gap for disadvantaged pupils, consider reintroducing targets for the delivery of tutoring to disadvantaged pupils, and using funding to incentivise the selection of these pupils for tutoring to ensure they are prioritised for additional support.

- We found that the impact of tutoring increased with the average number of tutoring hours pupils received, indicating that more hours of tuition can lead to greater benefits. Previous evidence from EEF (2021) also indicates that around 30 hours of tuition delivered over approximately 10 weeks has the greatest impact. We recommend that the NTP guidance reflects these findings on tutoring dosage and that this is communicated to schools.

- Undertake further research into the longer-term benefits of tutoring for pupils, exploring the extent to which tutoring can result in sustainable improvements in outcomes and help close the attainment gap.

1 Introduction

The Covid-19 pandemic and consequent restrictions on in-school attendance that applied to most pupils caused significant disruption to children's education and learning. The National Tutoring Programme (NTP) was introduced in the 2020-21 academic year and continues to be an important part of the UK Government's Covid-19 recovery strategy.

This report provides the results of the impact evaluation for the second year (2021-22) of the NTP. This explores the impact of the NTP on pupil attainment in English and maths at school level and pupil level and whether there are any differential effects by tutoring route, geographic region, dosage (number of hours of tuition pupils received), and concentration (percentage of children in each school who received tutoring). The report also investigates whether the impact on attainment outcomes differs by key pupil characteristics (such as pupil premium eligibility, prior lower attainment, gender, having a SEN, having EAL and ethnicity). It also considers the longer-term impact of NTP tuition for pupils who were selected for tutoring in the first year of the NTP (2020-21).

This report accompanies the implementation and process evaluation (IPE) for the second year of the NTP (Lynch *et al.*, 2022) which explored the implementation of the NTP, teacher perceptions of the NTP, reasons for non-engagement with the NTP, and perceptions of impact for pupils, staff, and schools. It also complements the independent review into the NTP being undertaken by Ofsted (Ofsted, 2022).

1.1 Context

In response to the Covid-19 pandemic, the UK Government asked all schools in England to restrict attendance for most of their pupils over three periods: March – May 2020; June – July 2020; and January – March 2021[6]. These restrictions caused significant disruption to pupils' learning and evidence indicates that this disruption has had a negative impact on pupil attainment (EEF, 2022). Further evidence indicates that pupils from disadvantaged backgrounds found it more challenging than their peers to keep up with learning during school closures (Major, Eyles and Machine, 2021), and that the attainment gap between disadvantaged pupils and their peers has increased since the Covid-19 pandemic began (DfE, 2022; EEF, 2022; Twist, Jones and Treleaven, 2022; Andrews, 2023) . Finding effective strategies to support pupils most affected by the disruption to their education is therefore key to helping them achieve their future potential.

[6] Note that during these periods schools were still open for children in vulnerable groups and children of keyworkers and there were also some exceptions where certain school year groups were able to attend.

Evidence for small group tuition

In the EEF review of the evidence on Covid-19 disruption and its impact on attainment, two key ways to support learning in these challenging times emerged:

1. to support effective remote learning to mitigate the extent to which the gap widens
2. 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 small-group tuition is effective, particularly where it is targeted at pupils' specific needs. Effect sizes vary across studies, with an average impact of two months additional progress for secondary schools and four months additional progress for primary schools (EEF, 2021b). A key finding is that the smaller the group and the more aligned it is to pupils' needs, the more effective the intervention (EEF, 2021b).

Meta-analyses have shown that tutoring programmes yield consistent and substantial positive impacts on learning outcomes, with average effect sizes ranging from 0.30 to 0.37 SD (Ritter *et al.*, 2009; Dietrichson *et al.*, 2017; Nickow, Oreopoulos and Quan, 2020; EEF, 2021b).

There is evidence to suggest that the advantages of small group tuition may be particularly relevant for disadvantaged pupils (Dietrichson *et al.*, 2017; Torgerson *et al.*, 2018). These pupils 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 pupils in homogenous small groups allows favourable comparisons between pupils and allows teachers to readily communicate pupil improvements (Mischo and Haag, 2002). These incentives, in turn, help maintain high levels of motivation (Pintrich and Schunk, 2002).

The research emphasises that tutoring needs to be high quality with sessions having the right duration and frequency to achieve optimal results. Tutor subject knowledge and pedagogic expertise are commonly identified as important delivery elements as well as the following structural characteristics; relationship with classroom learning, duration and frequency. Overall, it is recommended that tutors are knowledgeable in their subject area and trained in pedagogy, and that they deliver at least weekly sessions to pupils for a term or longer (Torgerson *et al.*, 2018; EEF, 2021a, 2021b).

1.2 Overview of the National Tutoring Programme

The NTP was set-up to provide additional support pupils and support was particularly targeted towards disadvantaged pupils, as these pupils were most affected by the disruption to education during the Covid-19 pandemic. The aim was to help them catch-up on missed learning and thus reduce the attainment gap (DfE, 2020). The NTP also

aims to embed tutoring as an intervention for disadvantaged pupils across the school system.

In 2021-22 schools were able to use NTP funding to provide tutoring via three distinct routes: Tuition Partners (TP), Academic Mentors (AM) and School-Led Tutoring (SLT).

Tuition Partners (TP)

The TP route could be accessed by all state-funded schools including special schools and Alternative Provision settings. It offered subsidised tuition (70% of total costs) for schools to access external tutors from approved Tuition Partners. Tutoring could be conducted face-to-face or online and was designed to be administered 1:1 or in small groups with a tutor/pupil ratio of up to 1:6 (with the aim that 80% of tuition was delivered in a ratio of 1:3). In February 2022, the criteria were updated to allow a tutor/pupil ratio of up to 1:6, although smaller groups were still recommended. For this route it was expected that 65% of the tuition would be delivered to PP pupils.

Academic Mentors (AM)

The AM route was designed to support schools with the most disadvantaged pupils. Schools were eligible to participate in the AM route if more than 20%[7] of their pupils qualified as PP, or if they were located in an area identified as being a priority area for raising educational attainment (Local Authority District or Opportunity Areas). Academic mentors were employed by the school as salaried staff members and 95% of the core salary cost was subsidised by DfE. AMs were required to have minimum qualifications of 3 A levels at grades at A* to C (or equivalent BTEC or T levels) and have obtained a grade 4 or C in both GCSE English and maths. They also completed a training programme prior to commencing tutoring within schools. This training lasted one week for applicants with qualified teacher status (QTS) and two weeks for those without QTS. Tutoring could be conducted face-to-face or online, ideally with a tutor/pupil ratio of 1:3 but ratios of up to 1:6 could be used.

School-Led Tutoring (SLT)

SLT was a new tutoring route introduced in 2021-22 to offer schools the flexibility to use tutors with whom they are familiar, including internal teachers or teaching assistants (DfE, 2020). It provided schools with a ring-fenced grant to subsidise (75% of costs were subsidised in 2021-22) locally sourced tutoring provision based on the number of pupils in the school eligible for PP. Schools were expected to source their own tutors for this route, who could be internal staff or external tutors. Internal staff without QTS or at least two years' experience in the subject and phase they planned to tutor in were required to attend an 11-hour training course prior to commencing tutoring. Tutors with QTS or the relevant experience (at least two years in the subject and phase) could complete an optional 2-hour training course. Tutoring could be conducted face-to-face or online,

[7] Initially this was 30 per cent but was later reduced to 20 per cent

ideally with a tutor/pupil ratio of 1:3 but ratios of up to 1:6 could be used if the school felt this was in the best interests of pupils. Local authorities also received an SLT grant to provide tutoring to Looked After Children (LAC) and any pupils they had placed in Independent Special Schools. Schools that did not use the SLT grant allocated to them were required to return the funding to the Education and Skills Funding Agency (ESFA).

For all three tutoring routes schools were advised that tutoring courses should be 12 to 15 hours long to have a meaningful impact on pupil attainment. Although we note that 15 hours is only around half of the amount of time found to be most effective for small group tuition by EEF (EEF, 2021b). Primary schools were able to offer tutoring in English, maths, and science, while secondary schools could provide tutoring in English, maths, science, humanities, and modern foreign languages. Schools were able to arrange tutoring times in discussion with their tutors/tuition providers but were advised that tutoring should take place at a time that encouraged high attendance and ensured that pupils did not miss core subjects.

1.3 Impact evaluation aims

Primary aim: to evaluate the impact of the second year (2021-22) of the NTP on educational attainment for a) all pupils, and b) for PP pupils and pupils with prior low attainment (PLA) compared to the expected standard at KS1 and KS2.

In addition, this evaluation also sought to understand:

- whether the impact of the NTP varied according to tutoring route (SLT, TP or AM), geographic region, dosage (the number of hours tutoring pupils received), and concentration (percentage of pupils in a school that were selected for tutoring)
- whether there were differences in the impact of the NTP for pupils with different characteristics (e.g., SEN, gender, English as an Additional Language (EAL)
- the longer-term impact of the TP and AM routes (introduced in the first year of the NTP in 2020-21).

1.4 Research questions

RQ1: What was the impact of the NTP on educational attainment outcomes in the 2021-22 academic year?

RQ2: What has been the longer-term impact of the NTP on educational attainment outcomes for pupils who were involved in Year 1 (2020-21 to 2021-22)?

RQ3: How does the impact of the NTP vary by route (TP and AM), geographic region, dosage, and concentration?

RQ4: What are the characteristics of pupils involved in the NTP evaluation sample?

RQ5: Is the impact of NTP participation different between groups within pupil characteristics (SEN, Ethnicity, Gender, PP/PLA)?

2 Method

2.1 Recruitment and sampling

We planned to evaluate impact of the NTP for two different samples of pupils which are detailed below.

2.1.1 Population analysis

Our population analysis used KS2 and KS4 attainment scores to explore the impact of the NTP on a) all Year 6 and Year 11 pupils and b) Year 6 and Year 11 PP and/or PLA pupils, attending mainstream schools in England.

The population for this analysis consisted of all mainstream primary and secondary schools in England with pupils in Year 6 or Year 11. Data for the population analysis consisted of KS2 standard scores (reading and maths) and KS4 attainment 8 scores (English language and maths) obtained from the National Pupil Database (NPD).

2.1.2 Research Champion (RC) schools

Secondly, we planned to investigate the impact of the NTP for a) all pupils and b) PP and/or PLA[8] pupils, in Years 1 to 6 who attended 'Research Champion' (RC) schools. All state primary schools were invited to participate in the RC impact analysis and were eligible for inclusion as an RC school if they undertook standardised English or maths assessments with pupils in any of Years 1 to 6 in 2021-22[9] (which would be used as baseline and outcome measures) and routinely uploaded this data to the relevant assessment provider's online repository. However, capacity issues within schools meant that many were unable to undertake these assessments, and many schools also lacked viable data (i.e., they did not have the relevant baseline and end-point assessment data). This meant that the sample size we were able to achieve was much smaller (103 schools for literacy and 70 schools for maths at analysis stage) than the 106 intervention and 106 comparison schools anticipated in the Study Plan based on our sample size calculations (Staunton *et al.*, 2022), which were informed by our 2020/21 evaluation (Poet *et al.*, 2022a). It was agreed with the DfE that despite these issues we would undertake this analysis as planned. However, the small sample sizes mean that the analyses are underpowered given that not all intervention pupils in the models received tuition. The consequence of this is that there are wide confidence intervals around the estimates of the effects of diluted tutoring. For example, for SLT maths the 95% confidence interval ranged from -0.115 to 0.158. Under normal circumstances, this would be a reasonable width but in this study is considered wide given we were trying to detect diluted effects. In the light of these issues, full details of the analysis and results are included in Appendix

[8] Note that prior attainment data is not available for Year groups 1 to 4 due to Covid-related disruption.
[9] Provided by Renaissance Learning, Rising Stars/Hodder, GL Assessment or NFER. Note that the English assessments used for this analysis included reading but were not exclusively reading assessments, hence we have used the term 'English' here.

2, rather than the main report, as the wide confidence intervals mean that this analysis is not able to contribute usefully to answering the research questions.

2.2 Defining intervention and comparison schools

This aim of the evaluation presented in this report was to assess the impact of the NTP on English and maths outcomes. Ideally, the intervention group would have contained all schools which participated in the NTP and schools that did not participate in the NTP would have made up the comparison group. However, with the addition of the SLT route in the 2021-22 academic year, it was anticipated that many more schools would participate. Consequently, a comparison group comprising only non-participating schools would be too small and non-comparable to the intervention group – preventing us from being able to draw reasonable conclusions[10]. Consequently, we decided to assess the impact of the TP/AM and SLT routes separately with the comparison groups being drawn from all schools not participating in the route of interest. We combined the AM and TP routes for most analyses because fewer schools used these routes compared with SLT (the exception to this was the analysis exploring the variation in the impact of the different tutoring routes where AM and TP were considered separately). This definition allows for both comparison and intervention schools to have been involved in other NTP intervention routes and this involvement in 'other routes' was accounted for by matching on 'concentration' of other route participation (the proportion of pupils at the school taking that route) during the statistical analysis[11] and controlling for it in the models. The subject in which tuition was received was only recorded in the datasets available to us for AM/TP not for SLT[12]. This meant that for the population analysis we investigated the impact of three different interventions: SLT (any subject), AM/TP (literacy) and AM/TP (maths).

Intervention schools were defined as schools with at least one pupil (or for the analysis of PP and/or PLA pupils one PP and/or PLA pupil) selected to receive tutoring in the relevant route and subject in 2021-22 (see Sections 2.4.2 and 2.4.3 for further details). For each intervention the pool of potential comparison schools was all schools not taking part in that intervention[13] (see Table 1). This means that the analysis aimed to evaluate the additive impact of each route of the NTP, over and above any other tutoring that may be being delivered in schools. This evaluation considers the impact of NTP as implemented at scale, rather than only evaluating the impact of tutoring when fully

[10] Note we did not know the exact numbers of schools participating in the NTP at the point decisions about the analysis were made.
[11] Note that both intervention and comparison schools could also have been involved in other catch-up interventions alongside the NTP, but we were unable to account for this in any of our analysis.
[12] We note that in addition to this limitation there is potential for error in the variables that classify a pupil as taking part in each NTP route. This is highlighted by the poor agreement between the NTP participation variables obtained from Research Champion schools and from the NPD/Randstad (see technical appendix). It is not clear which of these data sources is more reliable, but the poor agreement decreases confidence in the accuracy of both.
[13] Note that the assignment of schools to intervention and comparison groups was aided by monitoring information provided by DfE.

implemented as intended (e.g., when pupils received the full 15-hour tutoring course, in the specified tutor/pupil ratios etc).

Table 1: Schools in intervention and comparison groups according to NTP route and subject.

TP (in relevant subject)	AM (in relevant subject)	SLT (in relevant subject)	Group label	Group for TP/AM impact analysis	Group for SLT impact analysis
N	N	N	None	Comparison	Comparison
N	N	Y	SLT only	Comparison	Intervention
N	Y	N	AM only	Intervention	Comparison
N	Y	Y	AM & SLT	Intervention	Intervention
Y	N	N	TP only	Intervention	Comparison
Y	N	Y	TP & SLT	Intervention	Intervention
Y	Y	N	TP & AM	Intervention	Comparison
Y	Y	Y	All three	Intervention	Intervention

2.3 Pupil inclusion definitions

Two different pupil inclusion definitions were applied for the school-level and pupil-level analyses as detailed below (see also Table 3):

School-level analysis

All pupils in both intervention and comparison schools were included in this analysis. For the PP and/or PLA analysis, all pupils who were PP and/or PLA pupils were included.

Pupil-level analysis

The impact evaluation of Year 1 of the NTP in 2020-21 only explored the impact of the NTP at a school level. However, this suffered from the problem of dilution (as not all pupils attending intervention schools were themselves selected for tutoring – see sub-section 2.5). This means that it will inevitably underestimate the effect sizes for the pupil-level impact of tutoring. As the NTP is a pupil-level intervention, it should ideally be analysed at a pupil level and we were keen to explore the potential impact of the NTP at a pupil level in the evaluation of the second year of the NTP. However, pupil-level interventions are often subject to the problem of selection bias. In this case, the focus of

the intervention was on supporting pupils to recover lost learning, so it would be expected that teachers would select pupils who were falling furthest behind to receive tutoring. We would therefore expect to see potential negative selection bias within the evaluation, as without the intervention the pupils selected for tutoring would be expected to do 'less well' than their peers who were not selected for tutoring. Furthermore, if this selection is based on variables that are not available to the evaluation (e.g. recent performance in class tests of English or Maths) then it will not be possible to identify an appropriate comparison group to control for this negative bias.

Our initial planned approach was to use a statistical prediction model to create a comparison group of pupils who attended schools that did not participate in the NTP but were likely to have been selected for the NTP had their school chosen to participate in the programme (see technical appendix for details of the prediction model). This statistical model used the characteristics of pupils who participated in the NTP to predict selection for tutoring in non-participating schools. However, the accuracy of the pupil prediction model was poor for Year 6 pupils, both when all pupils and when only PP and/or PLA pupils were included in the model. As shown in Table 2 the false discovery rates[14] were high for the SLT intervention (ranging from 41% to 47%) and very high for the AM/TP interventions (ranging from 56% to 68%). In contrast, the false omission rates were typically higher when only PP and/or PLA pupils were included in the model (ranging from 25% to 39%) than when all pupils were included (ranging from 18% to 25%) (see technical appendix for further details).

Table 2: Pupil-level prediction model false discovery and omission rates

Intervention	Population	False Discovery Rate (%)	False omission rate (%)
SLT	PP and/or PLA	41	39
SLT	All pupils	47	25
AM/TP literacy	PP and/or PLA	65	25
AM/TP literacy	All pupils	68	18
AM/TP numeracy	PP and/or PLA	56	28
AM/TP numeracy	All pupils	59	19

Source: National Pupil Database 2021/2022 (School Census, KS2), Randstad tuition participation data

These results suggest that important explanatory variables for tutoring selection were not present in the data indicating a severe risk of selection bias.

[14] The false discovery rate describes the how often the model was wrong among all the cases where our model predicts a pupil would have received tutoring. Conversely, the false omission rate describes how often the model was wrong among all the cases where our model predicts a pupil would not have received tutoring. For this analysis rates of less than 10% would be considered good, though rates of less than 20% would be considered acceptable.

Since the pupil prediction model performed poorly, we adopted an alternative approach to attempt to minimise the danger of selection bias within the sample by restricting our pupil-level analysis to pupils who had already been selected for tutoring in the first year of the NTP (i.e., in 2020-21)[15]. Within this group, we compared the progress made by pupils who in the second year of the NTP (2021-22) attended intervention schools and were themselves selected for tutoring, with pupils who attended comparison schools and therefore did not receive tutoring. Pupils who in the second year of the NTP attended intervention schools but were not themselves selected for tutoring were excluded from this analysis.

The rationale for this was that as both control and intervention pupils were recently selected for tutoring (in year 1) they should be similar regarding any unobserved confounding variables, without the need to explicitly measure and model those confounders. This increases the likelihood that impact estimates observed in year 2 of the NTP at the pupil level are due to the actual benefit of tutoring, and not the result of selection bias. However, we must acknowledge that this approach is not infallible: selection for a further year of tutoring may itself entail some selection bias (e.g., pupils that struggled in year 1 of the NTP may have been more likely to have further tutoring). In summary, this approach is likely to reduce the selection bias, but probably not completely remove it. As with the school-level analysis we undertook statistical matching using entropy balancing (described below) to ensure the groups were well matched on key observable characteristics.

[15] Note that in 20202-21 only AM and TP tutoring routes were available.

Table 3: Summary of pupil inclusion definitions

	Intervention Group	Comparison Group
School level	All pupils in Year 6 and Year 11 in schools where at least one pupil was selected for the NTP route. All PP and/or PLA pupils in Year 6 and Year 11 in schools where at least one PP and/or PLA pupil was selected for the NTP route.	All pupils in corresponding year groups to the intervention pupils in comparison schools. All PP and/or PLA pupils in corresponding year groups to the intervention pupils in comparison schools
Pupil level	All pupils who were selected for tutoring in Year 1 of the NTP and were also selected for tutoring in Year 2 of the NTP. All PP and/or PLA pupils who were selected for tutoring in Year 1 of the NTP and were also selected for tutoring in Year 2 of the NTP.	All pupils who were selected for tutoring in Year 1 of the NTP but attended a school that did not participate in Year 2 of the NTP. All PP and/or PLA pupils who were selected for tutoring in Year 1 of the NTP but attended a school that did not participate in Year 2 of the NTP.

2.3.1 Matching

The first step in the process of building comparison groups of schools was to identify 'common support'. This ensured that no key school characteristics[16] which could determine eligibility for the NTP, or the likelihood of engagement, were present in only the intervention group or the control group. This was done separately for each tutoring route and, where applicable, subject.

To ensure that the intervention and comparison groups were as closely matched we conducted statistical matching using entropy balancing (Hainmueller, 2012). Entropy balancing is a method that assigns weights to comparison pupils to balance[17] observed variables between the groups; these weights are then included in subsequent regression modelling. Unlike many other data pre-processing methods, variables are balanced directly, rather than via propensity scores. This approach resulted in a well-balanced match between the intervention and comparison groups (see technical appendix for details of the entropy balancing and the degree of balance achieved between groups).

[16] Key school characteristics were school prior attainment, priority area for raising school standards, region, and proportion of pupils eligible for pupil premium.
[17] For this study only the means of the two groups were balanced, although in principle entropy balancing can achieve balance for other statistics (e.g., the variance).

For the school-level analyses the intervention and comparison groups were matched based on school characteristics. For the pupil-level analyses, the intervention and comparison groups were matched on both school and pupil characteristics.

2.4 Data analysis

2.4.1 Sample characteristics

Pupil and school characteristic information were explored using descriptive statistics. This information was obtained from data related to NTP tuition inputted into the School Census and Randstad Tuition Hub up to the 25 July 2022 (population analysis of Year 6 and Year 11 and some RC analysis). This means that the sample does not include any tuition received over the summer holidays, or tuition data for the 2021-22 academic year submitted during the summer holidays.

2.4.2 School-Level Population analysis

The impact of the AM/TP and SLT routes was assessed separately and for AM/TP considered the subject in which tuition was received. It was not possible for us to consider the subject in which tuition was received for SLT as this data was not available. This meant we investigated the impact of three different interventions for Year 6 and Year 11 pupils separately: SLT (any subject) participation; AM and/or TP (numeracy) participation; AM and/or TP (literacy) participation at both school level and pupil level.

For Year 6, this resulted in four mixed effects linear regression models (including the weights described in Section 2.3.1) to analyse the impact of:

- SLT on Key Stage 2 maths outcomes
- SLT on Key Stage 2 reading outcomes
- TP/AM (numeracy) on Key Stage 2 maths outcomes
- TP/AM (literacy) on Key Stage 2 reading outcomes.

Similarly for Year 11, this resulted in four mixed effects linear regression models (including the weights described in Section 2.3.1) to analyse the impact of:

- SLT on Key Stage 4 maths outcomes
- SLT on Key Stage 4 English Language outcomes
- TP/AM (maths) on Key Stage 4 maths outcomes
- TP/AM (English) on Key Stage 4 English Language outcomes.

Although the intervention indicator in each regression was defined at a school level, the units in the analysis were pupils and pupil outcomes were not aggregated to a school-level average. The pupil's school was included as random effect (all other variables were fixed effects) to account for clustering of pupil outcomes within schools. School- and

pupil-level characteristics were included as covariates, including characteristics used at the matching stage.

2.4.3 Analysis of route, geographic region, dosage, and concentration

Analysis exploring how the impact of the NTP varied according to route (SLT, TP or AM), geographic region, dosage, and concentration, broadly followed the regression analysis approach described in Section 2.4.2. The route, region and concentration analysis simply used weights calculated from the relevant school-level match, so rely on covariate adjustment to control for confounding[18]. Further details of how each analysis was conducted are given below.

Dosage analysis

The dosage analysis was conducted at a school level, with dosage defined as the average number of hours of tutoring pupils at that school received. Only schools participating in the relevant NTP route were included in the analysis. As this is a school-level analysis, all pupils at participating schools were included irrespective of whether they themselves participated in tutoring or not.

Note that for SLT the number of recorded tutoring hours was not subject specific therefore analysis of dosage was based on the total number of tutoring hours received, but these hours are likely to have been split across subjects.

No matching took place for this analysis, so confounding was controlled only through covariate adjustment.

Concentration analysis

The concentration analysis was also conducted at a school level and no new matches were conducted for this analysis. Concentration was defined as a categorical variable with five levels where the level was determined by the percentage pupils who were selected for tutoring within a school:

- comparison (no pupils selected for the relevant NTP route)
- low concentration ($0 < x \leq 25\%$ of pupils selected for the relevant NTP route)
- medium concentration ($25 < x \leq 50\%$ of pupils selected for the relevant NTP route)
- medium-high concentration ($50 < x \leq 75\%$ of pupils selected for the relevant NTP route)
- high concentration ($75 < x \leq 100\%$ of pupils selected for the relevant NTP route)

Route

For our RQ1 analysis AM and TP were combined. This analysis explored the impact of school participation in AM, school participation in TP, or school participation in both AM

[18] For example, the KS2 SLT concentration analysis used weights from the KS2 school-level SLT match. This means different concentration levels would not be expected to have the same average characteristics after weighting.

and TP. In each case the comparison group was schools that participated in neither AM nor TP. SLT concentration was accounted for using matching and covariate adjustment, as in other school-level models. We then compared these effects with those found for SLT.

Geographic region

This analysis investigated if the impact of the NTP varied according to geographic region. To do this we examined the impact of the NTP within each of the nine geographic regions in England (London, East Midlands, West Midlands, East of England, North East, North West, South East, South West, Yorkshire & Humber) by repeating the regression models described in Section 4.4.2, this time replacing the intervention indicator with a variable indicating both geographic region and intervention status (e.g. 'London and did participate in SLT'). Pair-wise comparisons between the relevant levels of this variable were then used to extract estimates of the impact of an NTP route for each region.

2.4.4 Analysis of the impact of the NTP between groups within pupil characteristics

This analysis investigated any differences in the impact of the NTP for pupils with the following characteristics:

- SEN status: here we compared the impact for the NTP for pupils with and without SEN. We did not do additional analysis to further breakdown any differences in impacts between different types of SEN, due to the small numbers of pupils in some of these groups.
- Ethnicity: to explore if the impact of the NTP varied according to ethnicity, we examined the impact of the NTP within six aggregated ethnicity categories (white, black, Chinese, other Asian, mixed ethnic background, any other ethnic background). These categories were defined using the first letter of the ethnicity codes available from the NPD variable 'Ethnicity'.
- Language: this analysis explored if there was any differential impact of the NTP for pupils with English as a first language compared with pupils with English as an Additional Language (EAL).
- Gender: here we investigated if the impact of the NTP was different for male compared with female pupils.
- PP/PLA status: our RQ1 analysis (described in Section 4.4.2) explored the impact of the NTP for PP and/or PLA pupils. This comparison therefore investigated if there were any different impacts between pupils who were PP only, PLA only or both PP and PLA.

Before conducting this analysis Pearson's chi-squared tests were used to explore whether any groups of pupils (SEN type, gender, Ethnicity, English as Additional Language (EAL), PLA and PP) were selected for tutoring at higher or lower levels than

would be expected compared with the relevant population (Year 6 and Year 11 pupils at schools in England). If they are selected more or less frequently than the general population this could induce bias, as the apparent differential effect of tuition in a subgroup may actually be due to a higher or lower tuition uptake in that subgroup.

The chi-squared tests were statistically significant for all the listed variables, for both the Year 6 and Year 11 populations ($p<0.001$). However, it is worth noting that due to the large sample sizes involved, even small differences in tuition uptake can obtain statistical significance. A better indicator of whether the difference in tuition uptake is large (and thus potentially impactful on results) can be obtained by comparing the proportions of pupils selected for tutoring with those of the general population shown in Table 5. This table indicates that PP pupils received tutoring substantially more often than the general population (as might be expected), but for other characteristics the difference was fairly small and so unlikely to substantially bias the analysis described below.

We explored whether the impact of the NTP was different within each subgroup defined by the pupil-level characteristics above using the analysis approach described in Section 2.4.2 but this time with an additional interaction between that characteristic and the intervention indicator. This meant that the models were run with intervention indicator replaced by a variable indicating both intervention status and the relevant characteristic. For example, in a model looking at the impact of SLT by gender, the level would be ‘male, SLT’, ‘male, no SLT’, ‘female, SLT’, ‘female, no SLT’. Pair-wise comparisons between levels of this variable were then made to estimate the impact of an NTP route for a specific subgroup. For example, the comparison ‘male, SLT’ minus ‘male, no SLT’ estimates the impact of SLT in the male subgroup. This is the same approach as described for region in Section 2.4.3.

Similarly to the analysis described in Section 2.4.3**,** the weights included in each regression model were taken from the relevant RQ1 school-level matches.

2.4.5 Analysis of the longer-term impact of the NTP

To assess the longer-term impact of the NTP on pupil attainment, further analysis was conducted looking at the effect of school-level participation in the first year (2020-21) of the NTP on KS2 and KS4 attainment at the end of 2021-2022. The analysis was conducted on the same populations as for the school-level primary analysis: pupils entering Year 6 and Year 11 in the 2021-2022 academic year. The difference in this case is that the intervention is participation in the NTP in 2020-21, not 2021-22. Pupils were therefore in Year 5 or Year 10 when their school took part in the NTP and a year or more will have passed between then and measurement of their KS2 or KS4 attainment. For this analysis there were four definitions of the intervention:

- School-level participation in AM and/or TP[19] numeracy amongst Year 5 pupils in 2020-21
- School-level participation in AM and/or TP literacy amongst Year 5 pupils in 2020-21
- School-level participation in AM and/or TP maths amongst Year 10 pupils in 2020-21
- School-level participation in AM and/or TP English amongst Year 10 pupils in 2020-21

For each definition all pupils attending a school participating in that tuition route are in the intervention group, while pupils attending the remaining schools are the comparison group.

Analysis proceeded similarly to the school-level models for the primary analysis: for each definition of the intervention statistically calculated weights were applied to balance the average characteristics of schools in the intervention and comparison groups. Four two-level (pupil, school) linear regression models were calculated, adjusted for relevant covariates and including the matching weights calculated at the previous stage. Matching variables and covariates were the same as the RQ1 school-level analysis, with one exception: participation in the second year of NTP was not included in the matching or regressions in any form.

2.5 Limitations

Without randomisation, any evaluation of the NTP will suffer from several limitations. We have described the key limitations below as they provide important context in which to consider the results presented in subsequent chapters and the conclusions which we are able to draw.

Not all pupils (or PP and/or PLA pupils where applicable) in intervention schools were themselves selected for tutoring, meaning that the results of the school-level analyses are subject to dilution.

The NTP is a pupil-level intervention meaning that not all pupils and not all PP and/or PLA pupils who attended schools that chose to offer NTP tuition were themselves selected for tutoring. However, the intervention group for the school-level analyses described in this report contained all pupils (or for the PP and/or PLA analysis all PP and/or PLA pupils) who attended a school offering NTP tuition in the relevant route and subject. As shown in Table 4, at KS2 the percentage of pupils in participating schools who were selected to take part in each route ranged from 21% for AM/TP English to 35%

[19] Note that SLT was not an offered in the first year of the NTP hence it is not included in this analysis.

for SLT[20]. At KS4, the percentage ranged from 7% for AM/TP English to 39% for SLT. This means that for the school-level analyses approximately two thirds of pupils (and in some cases more) in the intervention group did not actually receive any tutoring. We would therefore expect any effect sizes observed in these analyses to be diluted, and to underestimate the effect for those pupils who actually received tutoring. For SLT, no subject data was available, meaning dilution will be present in both the school- and pupil-level analysis.

Table 4: The percentage of pupils in participating schools who were selected to take part in each of the NTP routes

	KS2			KS4		
	SLT	**AM/TP maths**	**AM/TP English**	**SLT**	**AM/TP maths**	**AM/TP English**
Number of pupils in participating schools	438,783	96,646	56,768	451,255	141,520	127,035
Number of pupils selected for each route in participating schools	155,126	24,860	11,922	177,683	13,036	9,460
The percentage pupils selected for each route in participating schools	35.4%	25.7%	21.0%	39.4%	9.2%	7.4%

Source: National Pupil Database 2021/2022 (School Census, KS2), Randstad tuition participation data

Our analysis approach explored the additionality of each NTP route, and therefore assumes that additionality is possible.

The design in this study compares schools participating in the NTP route of interest with other schools not participating in that route but which may be participating in other NTP routes. To account for this the overall amount of tutoring in the other routes is matched across control and intervention groups and controlled for in the models. However, this means that we are assuming that if a school participates in more than one route the effects will be additive and consistent regardless of how much tutoring is being delivered by other routes. For example, it is assumed that the additional benefit of SLT will be the same whether or not the school is also taking part in AM and/or TP. In addition, we were

[20] Note that this figure is for SLT where no data about the tutored subject was available. This means that the dilution rate for pupils who received SLT in a specific subject is likely to be higher than this figure suggests.

not able to account for the presence of any other types of support pupils in either the intervention or comparison groups may have received.

Results of the pupil-level analysis are likely to be prone to selection bias, as schools chose which pupils received NTP tuition.

The nature of the NTP implementation meant that schools chose whether they wished to participate in specific routes of the NTP, and if so, which of their pupils received NTP tuition. In addition as statutory attainment data is only available at the end of Key Stages, we do not have a full understanding of pupils' attainment immediately prior to the introduction of the NTP. We controlled for English as an Additional Language (EAL) status, Special Educational Needs (SEN) status, IDACI quintile, gender, ethnicity, year group, and individual participation in any other NTP routes by including pupil-level covariates in the within-school analyses. We also attempted to control for pupil-level selection bias by restricting the analysis to those pupils who had already been selected for tutoring in the previous academic year. However, we cannot rule out the possibility that unobserved characteristics (e.g., pupil motivation) may have biased the results. Selection bias may also be present in the school-level analyses although in the light of recent work by Weidmann and Miratrix (Weidmann and Miratrix, 2020) we consider this less of a risk.

The analyses used multiple simultaneous tests but are not corrected for multiple comparisons, so statistically significant effects could be due to chance.

The analyses described in this report use multiple simultaneous tests on the same dataset. We used an alpha level of 0.05 to determine statistical significance for each of the individual tests, meaning that for each test there is a 5% (1 in 20) chance that that the null hypothesis (i.e., that there is no difference between the groups) is rejected when it is in fact true. However, as the number of comparisons increases, so too does the likelihood that a statistically significant result is due to chance (i.e., if 20 tests are conducted then it is reasonably likely, $p = 0.64$, that at least one of them could be statistically significant due to chance). While it is possible to apply statistical corrections to account for multiple comparisons these corrections have not been applied here. This means that all the results presented in this report need to be considered in the context of multiple testing, and it is likely that some statistically significant effects would not survive correction for multiple comparisons. That said the main purpose of statistical testing is to allow for sampling error when making generalisations from a sample to the general population. In most cases in this analysis we are directly measuring the effect in the entire population of interest (Year 6 and Year 11) so there is no need for statistical inference. However, for the findings to be applied to another year group or future cohort, statistical inference would be needed.

3 Characteristics of the samples

This chapter shows the composition of pupils included in the school- and pupil-level analyses by key pupil characteristic. It also shows the proportion of pupils in each sample participating in each of the NTP routes.

3.1 School-level evaluation samples

In total, 668,323 Year 6 pupils were included in the KS2 impact evaluation and formed our sample of 'all pupils'. Among these pupils 194,991 (29%) qualified as PP and 135,737 (20%) were PLA (Table 5). A total of 290,154 pupils made up our sample of PP and/or PLA pupils[21].

At KS4, 591,889 Year 11 pupils in total were included in the KS4 impact evaluation and formed our sample of 'all pupils' for KS4. Among these pupils 152,667 (26%) qualified as PP and 146,726 (25%) were PLA (Table 5). A total of 241,715 Year 11 pupils made up our KS4 sample of PP and/or PLA pupils.

For both the KS2 and KS4 evaluation samples, there is evidence to suggest that schools did prioritise PP pupils for NTP tuition.

Among all 178,511 Year 6 pupils selected for the NTP, 46% were PP, compared to 29% in the sample overall. Similarly, across all 197,971 Year 11 pupils selected for the NTP, 35% were PP, compared with 26% in the sample overall. In addition, 51% of Year 6 pupils and 47% of Year 11 pupils selected for the NTP were living in the 40% most disadvantaged areas according to the Income Deprivation Affecting Children Index (IDACI), providing further evidence that schools were targeting tutoring at pupils living in disadvantaged areas.

The proportions of PLA pupils and pupils with EAL or SEN were similar among those selected for the NTP and all pupils included in the evaluation sample, indicating that these pupils were selected for tutoring at the rates that would be expected given the prevalence of these characteristics within the population sample. Similarly, pupils' ethnic backgrounds did not appear to affect their likelihood of being selected for tutoring.

[21] Note that the number of PP and/or PLA pupils differs from the sum of the PP and PLA pupils as some pupils were both PP and PLA.

Table 5: Number and percentage of pupils included in the school-level evaluation sample split by pupil characteristics and NTP selection

	KS2			KS4		
	Total in sample	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**
Total number of pupils	668,323 (100%)	178,511 (100%)	489,812 (100%)	591,889 (100%)	197,971 (100%)	393,918 (100%)
Male	342,494 (51%)	88,186 (49%)	254,299 (52%)	303,521 (51%)	99,723 (50%)	203,798 (52%)
Female	325,819* (49%)	90,315 (51%)	235,503* (48%)	288,358* (49)	98,238* (50%)	190,110* (48%)
PP pupils	194,991 (29%)	81,135 (46%)	113,856 (23%)	152,667 (26%)	69,429 (35%)	83,238 (21%)
Non-PP pupils	469,479 (70%)	97,151 (54%)	372,328 (76%)	437,738 (74%)	128,373 (65%)	309,365 (79%)
PLA pupils	135,737 (20%)	36,649 (21%)	99,088 (20%)	146,726 (25%)	53,866 (27%)	92,860 (24%)
Non-PLA pupils	532,586 (80%)	141,862 (80%)	390,724 (80%)	445,163 (75%)	144,105 (73%)	301,058 (76%)
PP and/or PLA	290,154 (43%)	100,940 (57%)	189,214 (39%)	241,715 (41%)	98,711 (50%)	143,004 (36%)
Pupils with SEN	135,391 (20%)	41,682 (23%)	93,709 (19%)	98,229 (17%)	33,804 (17%)	64,425 (16%)

	KS2			KS4		
	Total in sample	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**
Pupils without SEN	532,932 (80%)	136,829 (77%)	396,103 (81%)	493,660 (83%)	164,167 (83%)	329,493 (84%)
Pupils with EAL	145,954 (22%)	40,222 (23%)	105,732 (22%)	105,054 (18%)	38,855 (20%)	66,199 (17%)
Pupils without EAL	522,369 (78%)	138,289 (78%)	384,080 (78%)	486,835 (82%)	159,116 (80%)	327,719 (83%)
White	484,003 (72%)	126,745 (71%)	357,258 (73%)	428,618 (72%)	137,117 (69%)	291,501 (74%)
Black	38,191 (6%)	13,433 (8%)	24,758 (5%)	36,339 (6%)	16,485 (8%)	19,854 (5%)
Chinese	4,348 (1%)	645 (<1%)	3,703 (1%)	2,436 (<1%)	628 (<1%)	1,808 (<1%)
Other Asian	75,333 (11%)	19,333 (11%)	56,000 (11%)	66,995 (11%)	23,349 (12%)	43,646 (11%)
Mixed ethnicity	42,647 (6%)	12,068 (7%)	30,579 (6%)	34,468 (6%)	12,278 (6%)	21,953 (6%)
Other ethnicity	14,428 (2%)	4,579 (3%)	9,849 (2%)	12,032 (2%)	4,659 (2%)	7,373 (2%)
IDACI Quintile 1 (most deprived)	134,256 (20%)	49,519 (28%)	84,737 (17%)	115,645 (20%)	49,099 (25%)	66,546 (17%)

	KS2			KS4		
	Total in sample	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**
IDACI Quintile 2	133,158 (20%)	41,677 (23%)	91,481 (19%)	116,894 (20%)	44,116 (22%)	72,778 (19%)
IDACI Quintile 3	133,265 (20%)	34,806 (20%)	98,459 (20%)	117,138 (20%)	38,576 (20%)	78,562 (20%)
IDACI Quintile 4	132,816 (20%)	29,158 (16%)	103,658 (21%)	119,354 (20%)	35,351 (18%)	84,003 (21%)
IDACI Quintile 5 (least deprived)	133,700 (20%)	23,047 (13%)	110,653 (23%)	122,030 (21%)	30,573 (15%)	91,457 (23%)

* indicates an approximate value due to SRS suppression of missing data values of <10.

Note: Percentages may not sum to 100 due to missing data or due to rounding.

Source: National Pupil Database 2021/2022 (School Census, KS2), Randstad tuition participation data

Among all the Year 6 pupils selected for tutoring, the vast majority (90%) participated in SLT (Table 6), while only 3% were selected to participate in AM. Among the PP and/or PLA pupils selected to participate in the NTP, pupils participated in each of the three routes in almost exactly the same proportions – 91% took part in SLT, 3% in AM and 17% in TP.

Among all Year 11 pupils selected for tutoring, the vast majority (90%) participated in SLT (Table 6), while only 12% participated in TP and 2% participated in AM. Among the PP and/or PLA pupils selected to participate in the NTP, very similar proportions of pupils participated in each of the three NTP routes – 93% took part in SLT, 2% in AM and 15% in TP.

Table 6: Number and percentage of pupils selected to take part in each of the NTP routes

	SLT	AM	TP
Number of all NTP pupils in Year 6	161,428 (90%)[22]	4,843 (3%)	29,989 (17%)
Number of PP and/or PLA NTP pupils in Year 6	91,704 (91%)	2,891 (3%)	17,409 (17%)
Number of all NTP pupils in Year 11	184,832 (93%)[23]	3,681 (2%)	24,318 (12%)
Number of PP and/or PLA NTP pupils in Year 11.	91,377 (93%)	2073 (2%)	14,386 (15%)

Source: National Pupil Database 2021/2022 (School Census, KS2), Randstad tuition participation data

3.2 Pupil-level evaluation samples

Table 7 shows the composition of the samples included in the each of the pupil-level models exploring the impact of the NTP on maths and English outcomes in KS2 and KS4.

For the pupil-level KS2 reading model, a total of 13,735 Year 6 pupils who had participated in Year 1 of the NTP formed our sample of 'all pupils' for the pupil-level impact evaluation of KS2 reading outcomes in 2021-22. Among these pupils, 6,470 qualified as PP and 4,118 were PLA. A total of 8,355 pupils made up our sample of PP and/or PLA pupils[24]. For the pupil-level maths model at KS2, 16,737 Year 6 pupils who had participated in Year 1 of the NTP formed our sample of 'all pupils' for the pupil-level impact evaluation of KS2 maths outcomes in 2021-22. Among these pupils, 7,283 qualified as PP and 4,451 were PLA. A total of 9,487 pupils made up our sample of PP and/or PLA pupils.

For the KS4 models, 14,336 Year 11 pupils who had participated in year 1 of the NTP formed our sample of 'all pupils' for the pupil-level impact evaluation of KS4 English language outcomes in 2021-22. Among these pupils, 5,880 qualified as PP and 4,061 were PLA. A total of 7,989 pupils made up our sample of PP and/or PLA pupils. Similarly, 16,544 Year 11 pupils who had participated in year 1 of the NTP formed our

[22] Note that some schools (and pupils) took part in more than one tutoring route, hence percentages do not sum to 100%.
[23] Note that some schools (and pupils) took part in more than one tutoring route, hence percentages do not sum to 100%.
[24] Note that the number of PP and/or PLA pupils differs from the sum of the PP and PLA pupils as some pupils were both PP and PLA.

sample of ‘all pupils’ for the pupil-level impact evaluation of KS4 maths outcomes in 2021-22. Among these pupils, 7,033 qualified as PP and 4,608 were PLA. A total of 9,304 pupils made up our sample of PP and/or PLA pupils.

Across all four pupil-level models, there is evidence that schools continued to prioritise PP pupils for NTP tuition in year 2 of the programme. For example, among all 6,325 pupils in the KS2 reading model who then went on to be selected for the NTP in 2021-22, 56% were PP, compared to 47% in the sample overall, indicating that schools continued to prioritise PP pupils for NTP tuition in year 2 of the programme. In addition, 61% of pupils who were selected to continue with the NTP in year 2 for literacy tuition were living in the 40% most disadvantaged areas according to the IDACI (versus 52% of pupils not selected for NTP in year 2). This pattern is present across the pupils comprising each of the pupil-level models.

Across all of the pupil-level model samples, pupils’ gender, EAL status, SEN status and ethnic backgrounds were similar among those were selected for the second year of the NTP compared with all pupils who received tutoring in the first year of the programme, indicating that these characteristics did not appear to affect their likelihood for being select for further tutoring.

Table 7: Number and percentage of pupils included in each pupil-level model split by pupil characteristics and NTP selection

	KS2 reading			KS2 maths			KS4 English			KS4 maths		
	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)
Total number of pupils	13,735 (100%)	6,325 (100%)	7,410 (100%)	16,737 (100%)	8,422 (100%)	8,315 (100%)	14,336 (100%)	6,602 (100%)	7,734 (100%)	16,544 (100%)	8,171 (100%)	8,373 (100%)
Male	7,388 (54%)	3,345 (53%)	4,043 (55%)	7,517 (45%)	3,654 (43%)	3,863 (47%)	7,815 (55%)	3,509 (53%)	4,306 (56%)	7,471 (45%)	3,617 (44%)	3,854 (46%)
Female	6,347 (46%)	2,980 (47%)	3,367 (45%)	9,220 (55%)	4,768 (57%)	4,452 (54%)	6,521 (46%)	3,093 (47%)	3,428 (44%)	9,073 (55%)	4,554 (56%)	4,519 (54%)
PP pupils	6470 (47%)	3520 (56%)	2950 (40%)	7283 (44%)	4373 (52%)	2910 (35%)	5880 (41%)	3116 (47%)	2764 (36%)	7033 (43%)	4011 (49%)	3022 (36%)
Non-PP pupils	7245 (53%)	2803 (44%)	4442 (60%)	9440 (56%)	4046 (48%)	5394 (65%)	8446 (59%)	3486 (53%)	4960 (64%)	9499 (57%)	4158 (51%)	5341 (64%)
PLA pupils	4118 (30%)	1999 (32%)	2119 (29%)	4451 (27%)	2323 (28%)	2128 (26%)	4061 (28%)	1865 (28%)	2196 (28%)	4608 (28%)	2383 (29%)	2225 (27%)

	KS2 reading			KS2 maths			KS4 English			KS4 maths		
	Total in sample	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**
Non-PLA pupils	9607 (70%)	4322 (68%)	5285 (71%)	12281 (73%)	6099 (72%)	6182 (74%)	10258 (72%)	4733 (72%)	5525 (71%)	11915 (72%)	5778 (71%)	6137 (73%)
PP and/or PLA pupils	8355 (61%)	4313 (68%)	4042 (55%)	9487 (57%)	5370 (64%)	4117 (50%)	7989 (56%)	3990 (60%)	3999 (52%)	9304 (56%)	5046 (62%)	4258 (51%)
Non PP and/or PLA pupils	5351 (39%)	2007 (32%)	3344 (45%)	7233 (43%)	3049 (36%)	4184 (50%)	6322 (44%)	2608 (40%)	3714 (48%)	7209 (44%)	3113 (38%)	4096 (49%)
Pupils with SEN	3,533 (26%)	1,671 (26%)	1,862 (25%)	3,588 (21%)	1,842 (22%)	1,746 (21%)	2,728 (19%)	1,279 (19%)	1,449 (19%)	2,759 (17%)	1,455 (18%)	1,304 (16%)
Pupils without SEN	10,202 (74%)	4,654 (74%)	5,548 (75%)	13,149 (79%)	6,580 (78%)	6,569 (79%)	11,608 (81%)	5,323 (81%)	6,285 (81%)	13,785 (83%)	6,716 (82%)	7,069 (84%)
Pupils with EAL	3,450	1,596	1,854	3,522	1,843	1,679	2,668	1,334	1,334	3,024	1,558	1,466

	KS2 reading			KS2 maths			KS4 English			KS4 maths		
	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)	Total in sample	Selected for the NTP (%)	Not selected for the NTP (%)
	(25%)	(25%)	(25%)	(21%)	(22%)	(20%)	(19%)	(20%)	(17%)	(18%)	(19%)	(18%)
Pupils without EAL	10,285 (75%)	4,729 (75%)	5,556 (75%)	13,215 (79%)	6,579 (78%)	6,636 (80%)	11,668 (81%)	5,268 (80%)	6,400 (83%)	13,520 (82%)	6,613 (81%)	6,907 (83%)
White	9,547 (70%)	4,320 (68%)	5,227 (71%)	12,085 (72%)	5,947 (71%)	6,138 (74%)	9,967 (70%)	4,382 (66%)	5,585 (72%)	11,401 (69%)	5,444 (67%)	5,957 (71%)
Black	1,012 (7%)	571 (9%)	441 (6%)	1,308 (8%)	762 (9%)	546 (7%)	1,106 (8%)	628 (10%)	478 (6%)	1,381 (8%)	832 (10%)	549 (7%)
Chinese	64 (1%)	22 (<1%)	42 (1%)	39 (<1%)	21 (<1%)	18 (<1%)	37 (<1%)	-	29 (<1%)	34 (<1%)	12 (<1%)	22 (<1%)
Other Asian	1,695 (12%)	720 (11%)	975 (13%)	1,630 (10%)	812 (10%)	818 (10%)	1,745 (12%)	844 (13%)	901 (12%)	1,947 (12%)	960 (12%)	987 (12%)
Mixed ethnicity	919	452	467	1,147	584	563	862	425	437	1,073	557	516

	KS2 reading			KS2 maths			KS4 English			KS4 maths		
	Total in sample	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**
	(7%)	(7%)	(6%)	(7%)	(7%)	(7%)	(6%)	(6%)	(6%)	(7%)	(7%)	(6%)
Other ethnicity	379 (%)	193 (3%)	186 (3%)	385 (2%)	226 (3%)	159 (2%)	339 (2%)	177 (3%)	162 (2%)	405 (2%)	209 (3%)	196 (2%)
IDACI Quintile 1 (most deprived)	4,227 (31%)	2,139 (34%)	2,088 (28%)	4,609 (28%)	2,646 (31%)	1963 (24%)	3,660 (26%)	1,867 (28%)	1,793 (23%)	4,635 (28%)	2,611 (32%)	2,024 (24%)
IDACI Quintile 2	3,498 (26%)	1,702 (27%)	1,796 (24%)	3,951 (24%)	2,082 (25%)	1869 (23%)	3,407 (24%)	1,665 (25%)	1,742 (23%)	3,885 (24%)	2,023 (25%)	1,862 (22%)
IDACI Quintile 3	2,435 (18%)	1,105 (18%)	1,330 (18%)	3,179 (19%)	1,508 (18%)	1671 (20%)	2,770 (19%)	1,266 (19%)	1,504 (19%)	3,097 (19%)	1,458 (18%)	1,639 (20%)
IDACI Quintile 4	1,992 (15%)	801 (13%)	1,191 (16%)	2,784 (17%)	1,233 (15%)	1551 (19%)	2,502 (18%)	1,095 (17%)	1,407 (18%)	2,720 (16%)	1,235 (15%)	1,485 (18%)

	KS2 reading			KS2 maths			KS4 English			KS4 maths		
	Total in sample	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**	**Total in sample**	**Selected for the NTP (%)**	**Not selected for the NTP (%)**
IDACI Quintile 5 (least deprived)	1,560 (11%)	566 (9%)	994 (13%)	2,180 (13%)	928 (11%)	1,252 (15%)	1,986 (14%)	704 (11%)	1,282 (17%)	2,196 (13%)	840 (10%)	1,356 (16%)

X indicates value suppressed due to SRS suppression of values of <10.

Note: Percentages may not sum to 100 due to missing data or due to rounding.

Source: National Pupil Database 2021/2022 (School Census, KS2), Randstad tuition participation data

4 What was the impact of the NTP on maths outcomes in 2021-22?

Key Findings

- We found evidence that participation in SLT was associated with small improvements in KS2 and KS4 maths outcomes that were equivalent to roughly 1 months' additional progress.
- We found some evidence that participation in TP maths tuition was associated with small negative effects on KS4 maths outcomes.
- The effect of tutoring on maths outcomes was similar for all pupils regardless of PP status.
- A higher dosage and/or concentration of tutoring was associated with better maths outcomes for SLT, but not for AM/TP.

This chapter presents results from the analysis exploring the school and pupil-level impact of the NTP on KS2 and KS4 maths outcomes. It also discusses how the school-level impact of the NTP varies by dosage (the average number of hours of tutoring delivered across all pupils in school), concentration (the proportion of pupils selected for tutoring), route (SLT, AM, TP, or AM+TP) and region. We anticipate that the school and pupil-level analyses should be complimentary with the school-level analysis giving the most reliable estimate of the direction of the effect and the pupil-level analysis giving the most reliable indication of the effect size at the pupil level.

4.1 What was the impact of the NTP on maths outcomes?

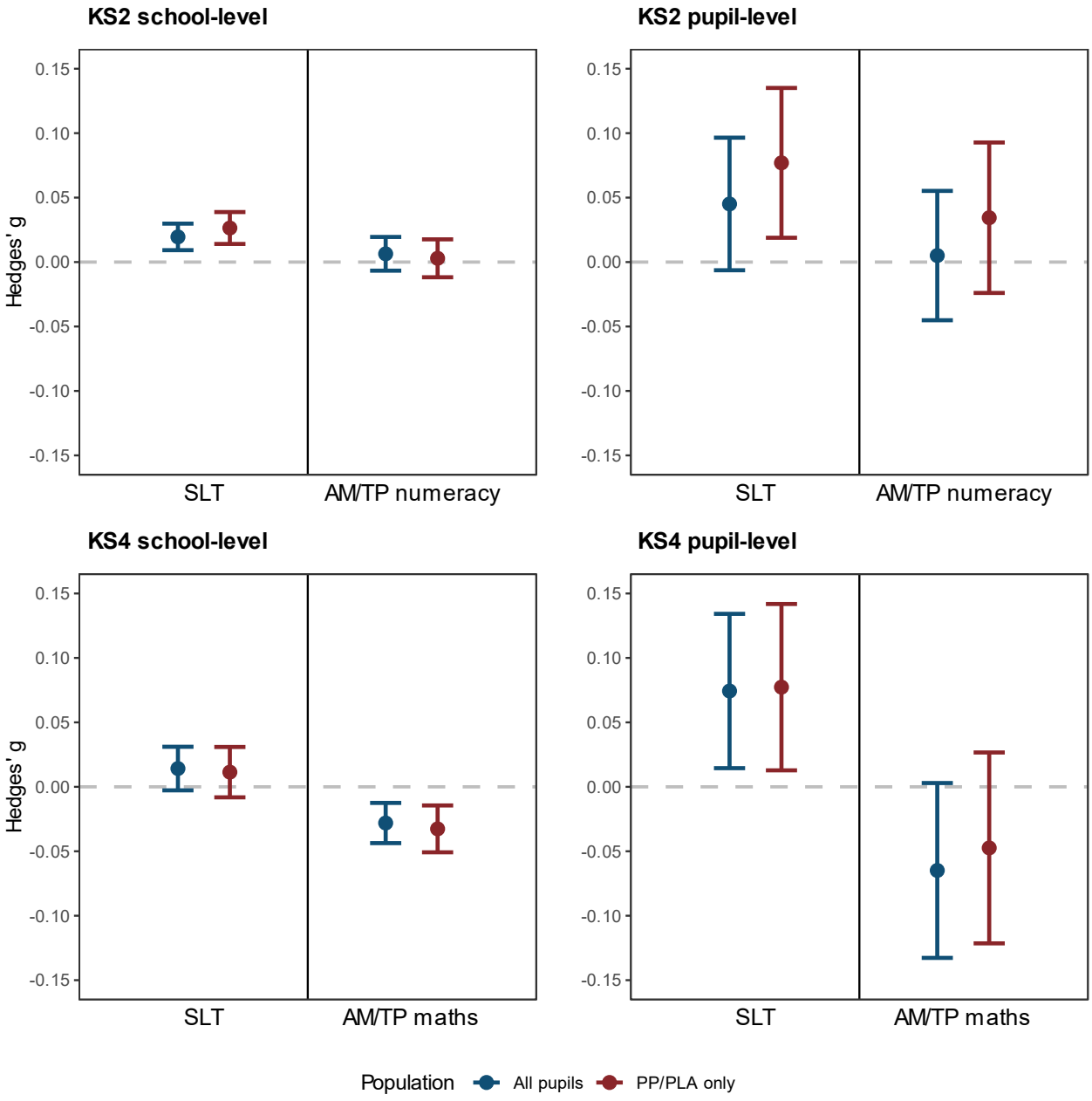
Our results indicate that participation in both SLT and AM/TP was associated with small improvements in maths outcomes (see Figure 1 and Table 8 for full statistical details including effect sizes and confidence intervals). The improvements to KS2 maths outcomes seen following participation in SLT were all small but some were statistically significant. For AM/TP the effects on KS2 outcomes were very small, and none came close to being statistically significant.

The picture for KS4 outcomes was similar to KS2 for the SLT route. We found positive effects of participation in SLT on maths outcomes that were of a similar magnitude to those seen for KS2, which reached statistical significance for the pupil-level analysis, but not for the school-level analysis. We also found evidence that participation in the AM/TP route was associated with small negative effects on KS4 maths outcomes that were statistically significant at the school level (and at pupil level for all pupils for KS4 English).

Overall, the effect sizes were all very small ranging from g = 0.077 for the SLT route (KS2 and KS4 outcomes for PP and/or PLA pupils at a pupil-level) to -0.065 for the AM/TP

route (KS4 outcomes for all pupils at a pupil-level). Using the EEF scale of effect sizes these effects would be equivalent of one months' additional progress or less (EEF, 2023). PP and/or PLA pupils made similar progress to other pupils across all routes and key stages.

Figure 1: The impact of the NTP year 2 on maths outcomes.



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 8: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on maths outcomes

Route	Outcome	Sample	Additional standard score points	Additional months progress[25]	Hedges' g (95% CI)	P value
SLT	School-level KS2	All pupils	0.149	0	0.020 (0.009, 0.030)	0.000
SLT	School-level KS2	PP and/or PLA pupils	0.209	0	0.026 (0.014, 0.039)	0.000
SLT	Pupil-level KS2	All pupils	0.277	0	0.045 (-0.006, 0.096)	0.086
SLT	Pupil-level KS2	PP and/or PLA pupils	0.486	1	0.077 (0.019, 0.135)	0.010
AM/TP numeracy	School-level KS2	All pupils	0.049	0	0.007 (-0.007, 0.019)	0.338
AM/TP numeracy	School-level KS2	PP and/or PLA pupils	0.023	0	0.003 (-0.012, 0.018)	0.698
AM/TP numeracy	Pupil-level KS2	All pupils	0.032	0	0.005 (-0.045, 0.055)	0.844
AM/TP numeracy	Pupil-level KS2	PP and/or PLA pupils	0.222	0	-0.034 (-0.024, 0.093)	0.248
SLT	School-level KS4	All pupils	0.029	0	0.014 (-0.003, 0.031)	0.101
SLT	School-level KS4	PP and/or PLA pupils	0.021	0	0.011 (-0.008; 0.031)	0.254
SLT	Pupil-level KS4	All pupils	0.126	1	0.074 (0.014, 0.134)	0.015

[25] Defined according to EEF (2023).

Route	Outcome	Sample	Additional standard score points	Additional months progress[25]	Hedges' g (95% CI)	P value
SLT	Pupil-level KS4	PP and/or PLA pupils	0.128	1	0.077 (0.013, 0.142)	0.019
AM/TP maths	School-level KS4	All pupils	-0.058	0	-0.028 (-0.044, -0.012)	$p < 0.001$
AM/TP maths	School-level KS4	PP and/or PLA pupils	-0.060	0	-0.033 (-0.051, -0.014)	$p < 0.001$
AM/TP maths	Pupil-level KS4	All pupils	-0.112	-	-0.065 (-0.133, 0.003)	0.061
AM/TP maths	Pupil-level KS4	PP and/or PLA pupils	-0.079	0	-0.047 (-0.121, 0.027)	0.210

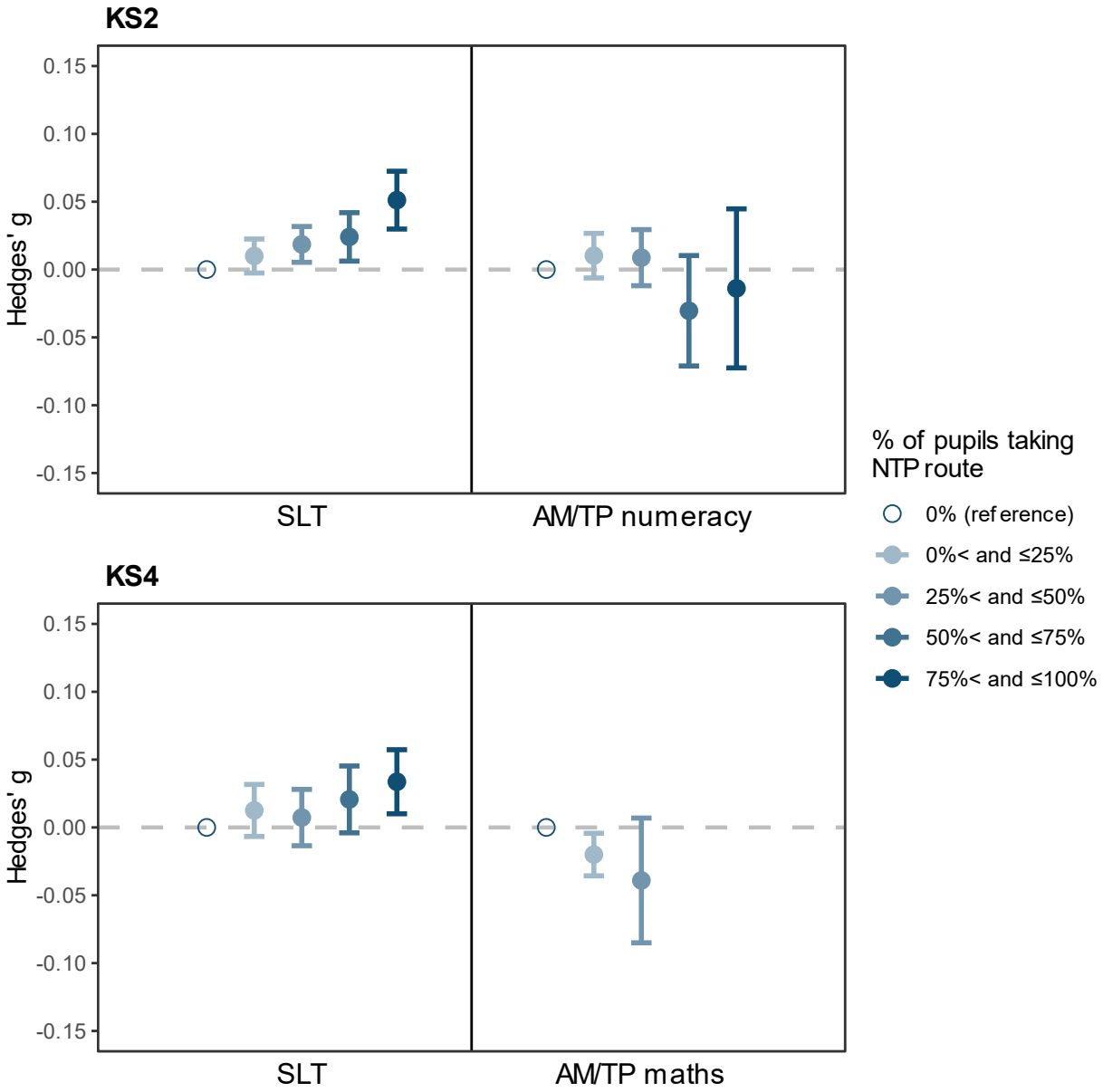
Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

4.2 How does the impact of the NTP on maths outcomes vary by school-level tutoring dosage and concentration?

At both KS2 and KS4, the dosage analysis detected very small statistically significant positive effects for increasing hours of SLT tuition (see Table 9 for full details of the dosage analysis). A similar effect can be seen for concentration where schools with higher concentrations of tutoring had progressively better overall maths outcomes (see Figure 2 for an illustration of the effects of concentration). Again, the effects sizes were small.

For schools who participated in the AM/TP route there was no relationship between tutoring dosage and maths outcomes, and this was born out in the concentration analysis for KS2 which showed a flat distribution. For KS4, the concentration analysis suggests that increasing concentration may be indicative of slightly poorer outcomes, but this effect is fragile as the sample size for high concentration AM/TP schools was very small.

Figure 2: The impact of the NTP on KS2 maths outcomes for schools with different overall concentrations of tutoring



Note that the bar for 50% < x ≤ 75% tutoring concentration is not included on the graph due to the small number of schools included in this analysis (n < 10).

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 9: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on maths outcomes by dosage

Route	Outcome	Additional standard score points (95% CI)	P value
SLT	School-level KS2 maths	0.007 per extra hour (0.004, 0.010)	0.000
SLT	School-level KS4 maths	0.001 per extra hour (0.000, 0.002)	0.003
AM/TP numeracy	School-level KS2 maths	0.008 per extra hour (-0.023, 0.039)	0.617
AM/TP maths	School-level KS4 maths	-0.004 per extra hour (-0.035, 0.025)	0.743

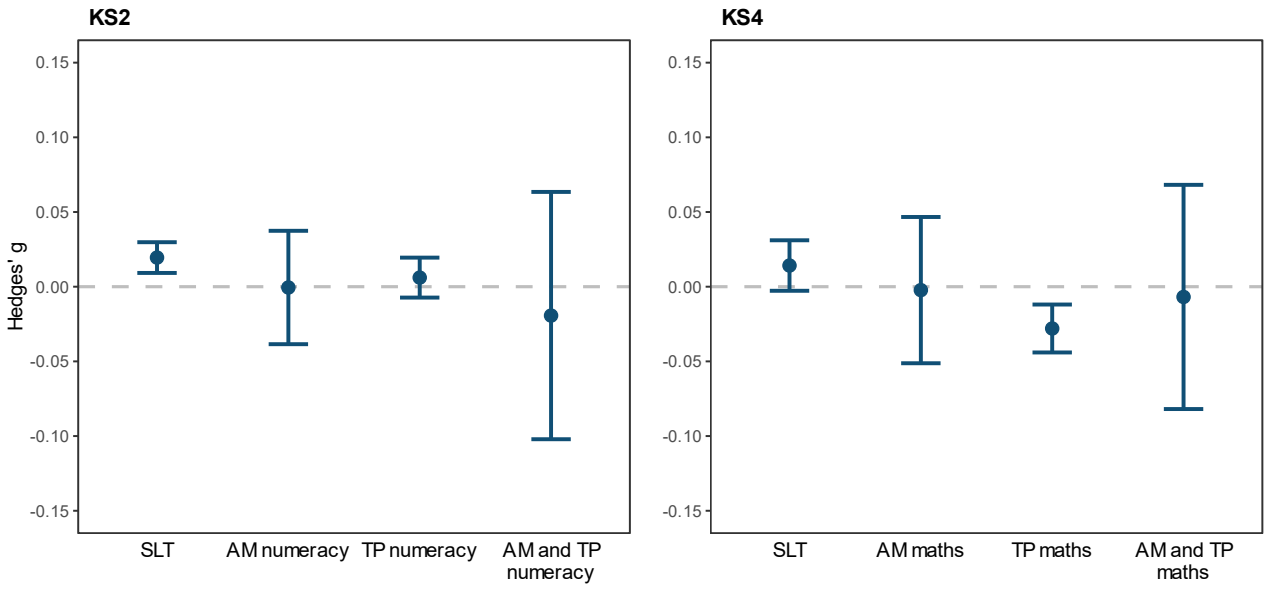
Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

4.3 How does the impact of the NTP on maths outcomes vary by tutoring route?

Our results indicate that participation in SLT was associated with small positive effects on both KS2 and KS4 maths outcomes while participation in TP was associated with negative outcomes at KS4 suggesting that TP performed worse than the other routes (see Figure 3 for an illustration of how school-level impact on maths outcomes differed depending on tutoring route and Table 10 for full statistical results). There were very few schools participating in the AM only or AM and TP routes so the confidence intervals around these estimates are very wide making it impossible to draw any reliable conclusions as to their effectiveness.

The differences in outcomes between the NTP routes, particularly between SLT and TP, may be related to differences in how these routes are implemented in schools. SLT allowed schools to use internal staff as tutors, making it more likely the staff delivering the NTP had existing relationships with staff and pupils that they could build on, and that tutoring was delivered in-person rather than online. Although the AM route saw tutors recruited through external providers, they were employed by the schools themselves becoming in-house members of staff. As such, it seems probable that in most cases the implementation of AM was more similar to SLT than to TP, where external providers were used.

Figure 3: The impact of the NTP on maths outcomes split by individual route



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 10: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on maths outcomes by route

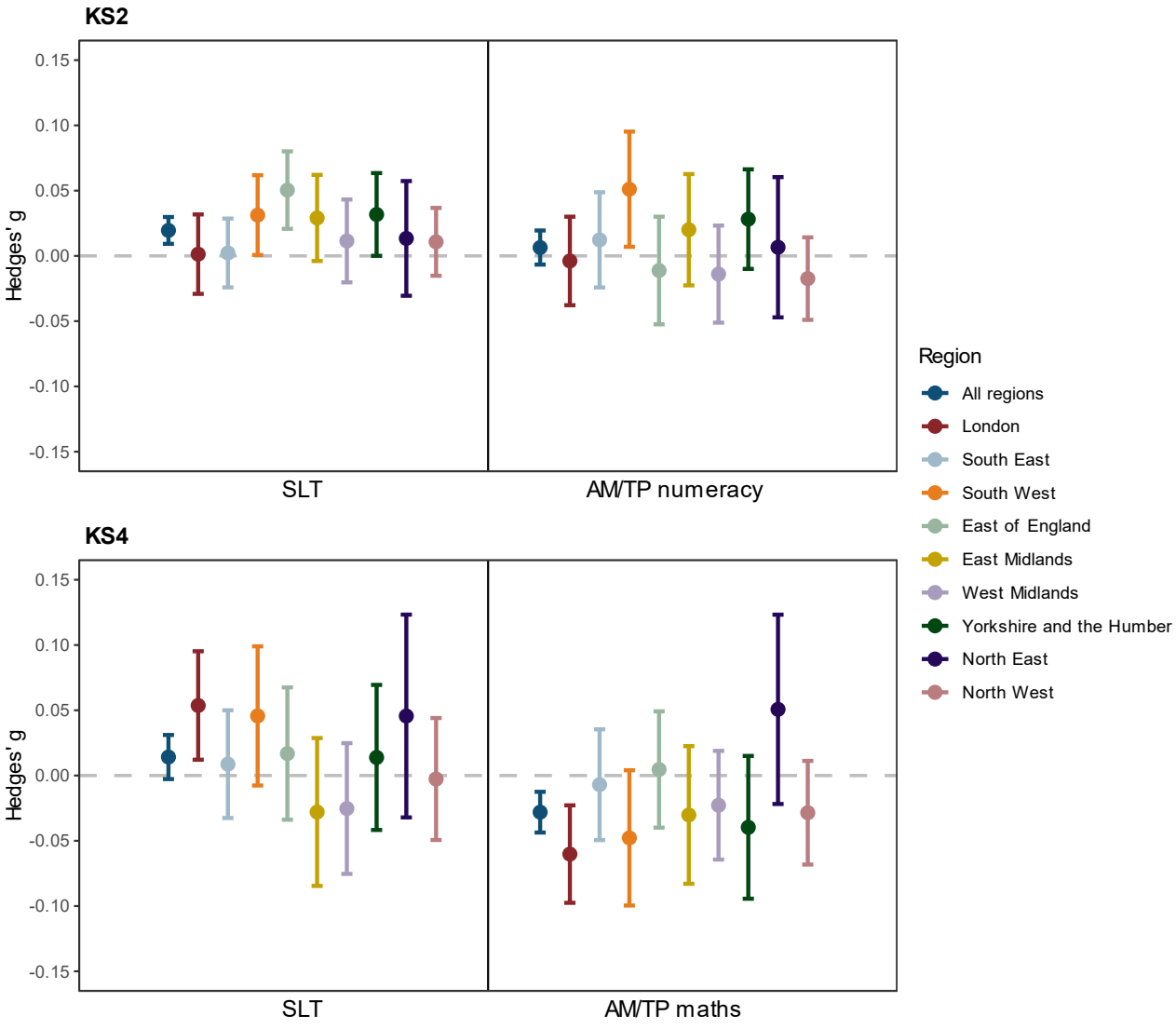
Route	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
SLT	KS2 maths	0.149	0.020 (0.009, 0.030)	<0.001
AM numeracy	KS2 maths	-0.004	-0.000 (-0.038, 0.037)	0.978
TP numeracy	KS2 maths	0.047	0.006 (-0.007, 0.019)	0.372
AM & TP numeracy	KS2 maths	-0.148	-0.019 (-0.102, 0.063)	0.647
SLT	KS4 maths	0.029	0.014 (-0.003, 0.031)	0.101
AM maths	KS4 maths	-0.005	-0.002 (-0.051, 0.047)	0.927
TP maths	KS4 maths	-0.058	-0.028 (-0.044, -0.012)	0.001
AM & TP maths	KS4 maths	-0.014	-0.007 (-0.082, 0.068)	0.858

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

4.4 How does the impact of the NTP on KS2 maths outcomes vary by geographic region?

Figure 4 illustrates the overall impact of the NTP within each of the nine geographic regions in England (see Table 11 for full statistical details). Because the number of schools in each region is much smaller than for the sample for England as a whole, the confidence intervals around these estimates are correspondingly large. Although there does appear to be some variation by region, all the confidence intervals overlap with the confidence intervals for England as a whole suggesting that these variations are likely to be the result of sample variability rather than meaningful regional differences.

Figure 4: The impact of the NTP on maths outcomes at school-level by region



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 11: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on maths outcomes by region

Route	Region	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
SLT	**London**	KS2 maths	0.010	0.001 (-0.029, 0.032)	0.932
	South East	KS2 maths	0.017	0.002 (-0.024, 0.029)	0.868
	South West	KS2 maths	0.239	0.031 (0.001, 0.062)	0.045
	East of England	KS2 maths	0.386	0.050 (0.021, 0.080)	0.001
	East Midlands	KS2 maths	0.223	0.029 (-0.004, 0.062)	0.084
	West Midlands	KS2 maths	0.088	0.011 (-0.020, 0.043)	0.479
	Yorkshire and the Humber	KS2 maths	0.243	0.031 (0.000, 0.063)	0.050
	North East	KS2 maths	0.103	0.013 (-0.031, 0.057)	0.550
	North West	KS2 maths	0.082	0.011 (-0.015, 0.037)	0.417
AM/TP Maths	**London**	KS2 maths	-0.030	-0.004 (-0.038, 0.030)	0.823
	South East	KS2 maths	0.094	0.012 (-0.024, 0.049)	0.509
	South West	KS2 maths	0.391	0.051 (0.007, 0.095)	0.023
	East of England	KS2 maths	-0.086	-0.011 (-0.052, 0.030)	0.594

Route	Region	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
	East Midlands	KS2 maths	0.153	0.020 (-0.023, 0.062)	0.357
	West Midlands	KS2 maths	-0.107	-0.014 (-0.051, 0.023)	0.462
	Yorkshire and the Humber	KS2 maths	0.216	0.028 (-0.010, 0.066)	0.148
	North East	KS2 maths	0.051	0.007 (-0.047, 0.060)	0.809
	North West	KS2 maths	-0.133	-0.017 (-0.049, 0.014)	0.280
SLT	**London**	KS4 maths	0.111	0.054 (0.012, 0.095)	0.011
	South East	KS4 maths	0.018	0.009 (-0.033, 0.050)	0.678
	South West	KS4 maths	0.095	0.046 (-0.008, 0.099)	0.093
	East of England	KS4 maths	0.035	0.017 (-0.034, 0.067)	0.515
	East Midlands	KS4 maths	-0.058	-0.028 (-0.085, 0.029)	0.334
	West Midlands	KS4 maths	-0.053	-0.025 (-0.075, 0.025)	0.323
	Yorkshire and the Humber	KS4 maths	0.029	0.014 (-0.042, 0.069)	0.625
	North East	KS4 maths	0.095	0.046 (-0.032, 0.123)	0.250

Route	Region	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
	North West	KS4 maths	-0.006	-0.003 (-0.049, 0.044)	0.911
AM/TP maths	**London**	KS4 maths	-0.125	-0.060 (-0.098, -0.023)	0.002
	South East	KS4 maths	-0.015	-0.007 (-0.049, 0.035)	0.746
	South West	KS4 maths	-0.099	-0.048 (-0.100, 0.004)	0.071
	East of England	KS4 maths	0.010	0.005 (-0.040, 0.049)	0.838
	East Midlands	KS4 maths	-0.063	-0.030 (-0.083, 0.023)	0.262
	West Midlands	KS4 maths	-0.047	-0.023 (-0.064, 0.019)	0.284
	Yorkshire and the Humber	KS4 maths	-0.082	-0.040 (-0.094, 0.015)	0.155
	North East	KS4 maths	0.105	0.051 (-0.022, 0.123)	0.170
	North West	KS4 maths	-0.059	-0.028 (-0.068, 0.011)	0.160

Note: bold denotes results that reached statistical significance at $p < 0.05$.

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

5 What was the impact of the NTP on English outcomes in 2021-22?

Key Findings

- We found evidence that participation in SLT was associated with very small positive effects on KS2 and KS4 English outcomes, but these effects equated to less than one months' additional progress so may not reflect meaningful changes for pupils.
- We found some evidence to indicate that participation in AM/TP was associated with small negative effects on KS2 and KS4 English outcomes, and TP performed less well than the other routes.
- The effect of tutoring on English outcomes similar across pupils with and without PP and/or PLA status.
- A higher tutoring dosage and/or concentration was associated with better English outcomes for SLT, but there was not a consistent pattern of results for AM/TP.

This chapter presents results from the analysis exploring the school- and pupil-level impacts of the NTP on KS2 and KS4 English outcomes. Similarly to Chapter 4, it also discusses how the school-level impact of the NTP varies by dosage (the average number of hours of tutoring delivered across all pupils in school), concentration (the proportion of pupils selected for tutoring), route (SLT, AM, TP or AM and TP) and region.

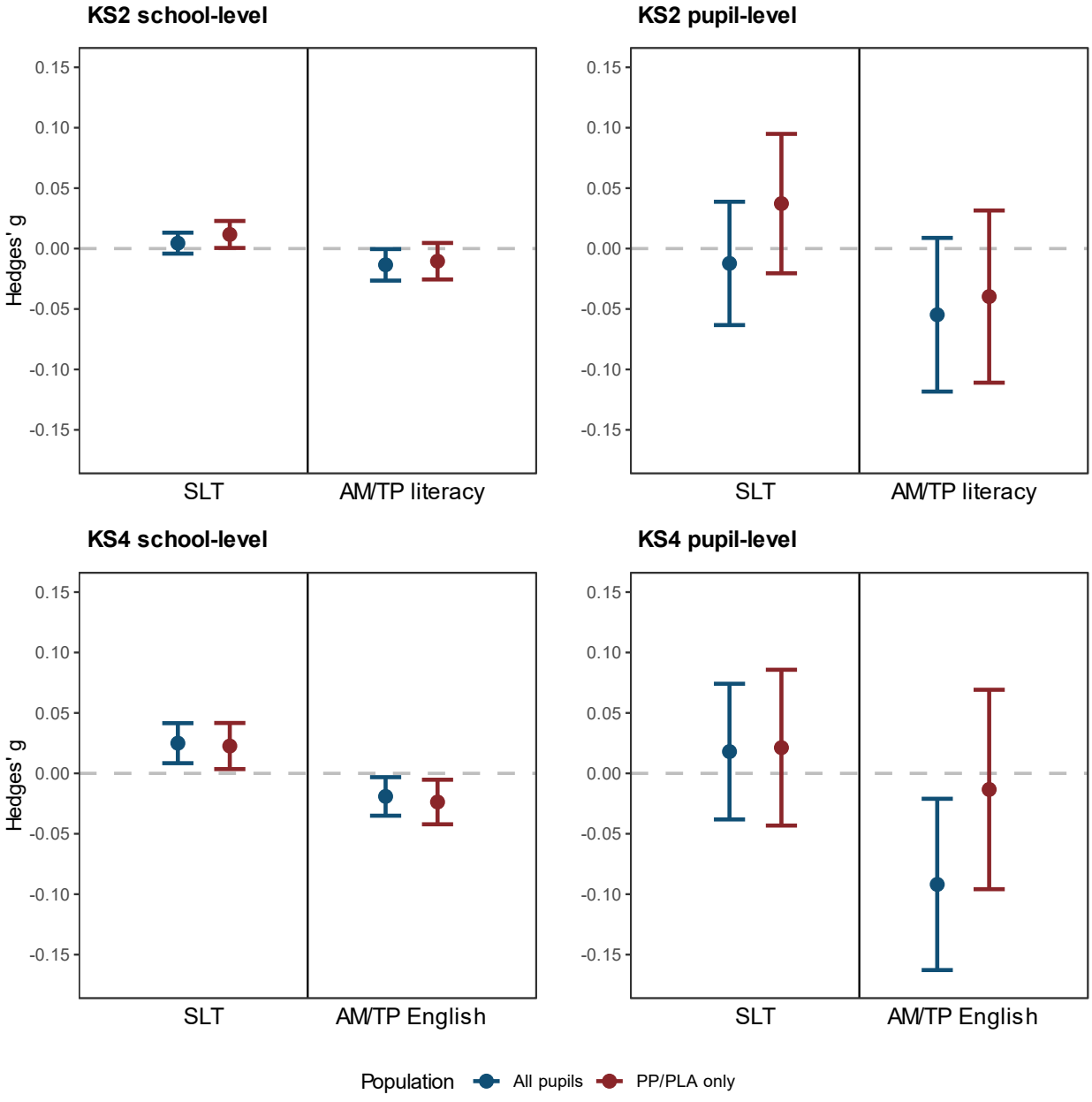
5.1 What was the impact of the NTP on English outcomes?

Our results showed that participation in SLT was generally associated with better English outcomes relative to the comparison group at both KS2 and KS4 (see Figure 5). At school level these effects were statistically significant for all pupils and PP and/or PLA pupils at KS4 and for PP and/or PLA pupils at KS2 (see Table 12 for full statistical details including effect sizes and confidence intervals). In general, we would have expected the magnitude of the effect sizes to be larger at pupil level than school level due to reduced dilution within the analysis. However, we found that the effect sizes at pupil-level were similar to those seen at school level, and that the pupil-level results did not generally reach statistical significance.

We found evidence to suggest that participation in AM/TP was consistently associated with negative effects on KS2 and KS4 English outcomes relative to the comparison group. These effects were generally small but were statistically significant at school level for KS4 and at a pupil-level for all pupils but not PP and/or PLA pupils for both KS2 and KS4.

Overall, the effect sizes we detected were very small, ranging from Hedges' g = 0.037 for the SLT route (KS2; PP and/or PLA pupils) to -0.092 for the AM/TP route (KS4; all pupils). Using the EEF scale of effect sizes these are equivalent to less than one months' additional progress (EEF, 2021c). PP and/or PLA pupils made similar progress to other pupils across all routes and both key stages, suggesting that outcomes did not differ for this subgroup of pupils.

Figure 5: The impact of the second year of the NTP on English outcomes



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 12: Results of the linear mixed effects models exploring the impact of SLT

Route	Outcome	Sample	Additional standard score points	Additional months progress[26]	Hedges' g (95% CI)	p value
SLT	School-level KS2	All pupils	0.036	0	0.004 (-0.004, 0.013)	0.310
SLT	School-level KS2	PP and/or PLA pupils	0.099	0	0.012 (0.001, 0.023)	0.040
SLT	Pupil-level KS2	All pupils	-0.091	0	-0.012 (-0.063, 0.039)	0.637
SLT	Pupil-level KS2	PP and/or PLA pupils	0.286	0	0.037 (-0.020, 0.095)	0.206
AM/TP English	School-level KS2	All pupils	-0.107	0	-0.013 (-0.026, 0.000)	0.043
AM/TP English	School-level KS2	PP and/or PLA pupils	-0.088	0	-0.010 (-0.026, 0.005)	0.175
AM/TP English	Pupil-level KS2	All pupils	-0.413	-	-0.055 (-0.118, 0.009)	0.092
AM/TP English	Pupil-level KS2	PP and/or PLA pupils	-0.311	0	-0.040 (-0.111, 0.032)	0.275
SLT	School-level KS4	All pupils	0.047	0	0.025 (0.008, 0.042)	0.003
SLT	School-level KS4	PP and/or PLA pupils	0.038	0	0.023 (0.004, 0.042)	0.020

[26] Defined according to EEF (2023).

Route	Outcome	Sample	Additional standard score points	Additional months progress[26]	Hedges' g (95% CI)	p value
SLT	Pupil-level KS4	All pupils	0.030	0	0.018 (-0.038, 0.074)	0.529
SLT	Pupil-level KS4	PP and/or PLA pupils	0.034	0	0.021 (-0.043, 0.086)	0.518
AM/TP English	School-level KS4	All pupils	-0.036	0	-0.019 (-0.035, -0.003)	0.019
AM/TP English	School-level KS4	PP and/or PLA pupils	-0.040	0	-0.024 (-0.042, -0.005)	0.012
AM/TP English	Pupil-level KS4	All pupils	-0.155	-	-0.092 (-0.162, -0.021)	0.011
AM/TP English	Pupil-level KS4	PP and/or PLA pupils	-0.022	0	-0.013 (-0.096, 0.069)	0.752

Note: bold denotes results that reached statistical significance at $p < 0.05$.

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

5.2 How does the impact of the NTP on English outcomes vary by school-level tutoring dosage and concentration?

The dosage analysis detected very small statistically significant positive effects of increasing hours of SLT tuition on KS2 and KS4 outcomes (see Table 13) for full details of the dosage analysis). A similar effect can be seen for concentration where schools with higher concentrations of tutoring appeared to have progressively better English outcomes (see Figure 6 for an illustration of the effects of concentration). Although this only appeared to be the case for schools that achieved above 50% concentration for KS2, and again the effect sizes were generally small (equating to less than one month's additional progress).

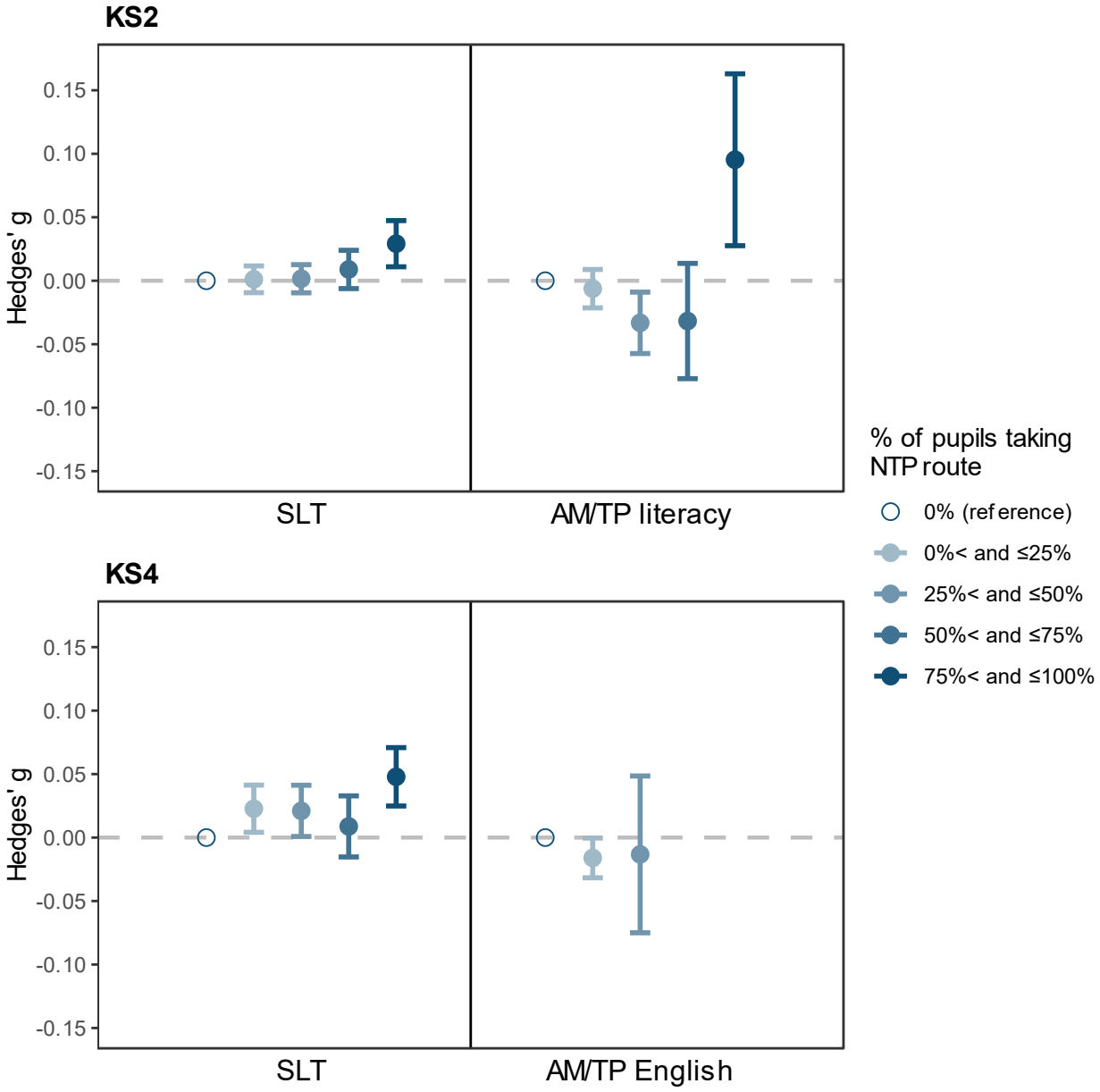
For schools using the AM/TP route we detected a small statistically significant positive effect of dosage on KS2 outcomes, but not on KS4 outcomes. We did not detect a consistent pattern of results for the effects on concentration on either KS2 or KS4 outcomes.

Table 13: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on English outcomes by dosage

Route	Outcome	Additional standard score point (95% CI)	p value
SLT	School-level KS2 English	0.005 per extra hour (0.003, 0.008)	0.000
SLT	School-level KS4 English	0.001 per extra hour (0.000, 0.002)	0.015
AM/TP English	School-level KS2 English	0.036 per extra hour (0.003, 0.070)	0.035
AM/TP English	School-level KS4 English	-0.011 per extra hour (-0.023, 0.046)	0.519

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Figure 6: The impact of the NTP on KS2 and KS4 English outcomes for schools with different overall concentrations of tutoring



Note: the bars for 50% < x ≤ 75% and 75% < x ≤ 100% tutoring concentration for AM/TP English in KS4 are not included on the graph due to the small number of schools included in these analyses (n < 10).

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

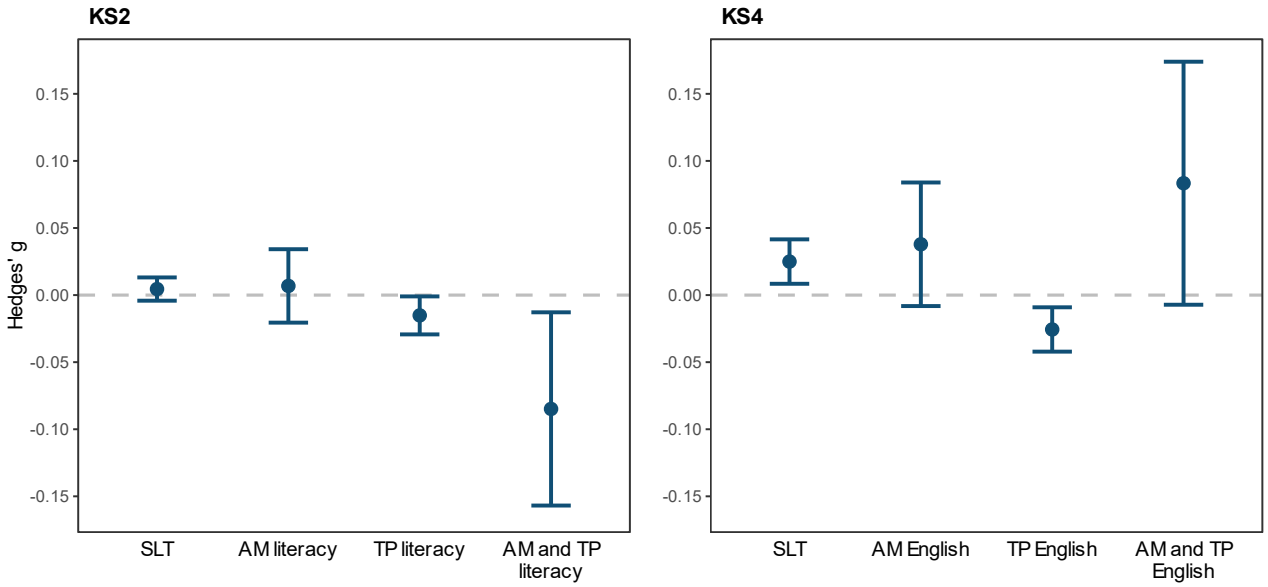
5.3 How does the impact of the NTP on English outcomes vary by tutoring route?

As described in section 5.1, the SLT route was associated with small positive effects on KS2 and KS4 English outcomes at a school level. Meanwhile, school-level participation in TP was associated with small negative effects on English outcomes (see Figure 7 for an illustration of how school-level impact on English outcomes differed depending on tutoring route and Table 14 for full statistical results).

Very few schools participated in the AM only or AM and TP routes. This means that the confidence intervals around these estimates are very wide, making it difficult to draw any reliable conclusions as to their effectiveness. However, given the direction and magnitude of the effect size point estimates for AM are similar to those for SLT, and the negative effects observed for TP, it is likely that the negative effects found for school- and pupil-level participation in TP and/or AM described in Section 5.1 are primarily being driven by the TP route.

As discussed above in Section 4.3, the difference in English outcomes by NTP route may be related to the differences in how SLT, AM and TP are delivered in schools, with the TP route delivered by external staff unfamiliar to pupils and tutoring often delivered online rather than in person.

Figure 7: The impact of the NTP on English outcomes split by individual route



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 14: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on English outcomes by route

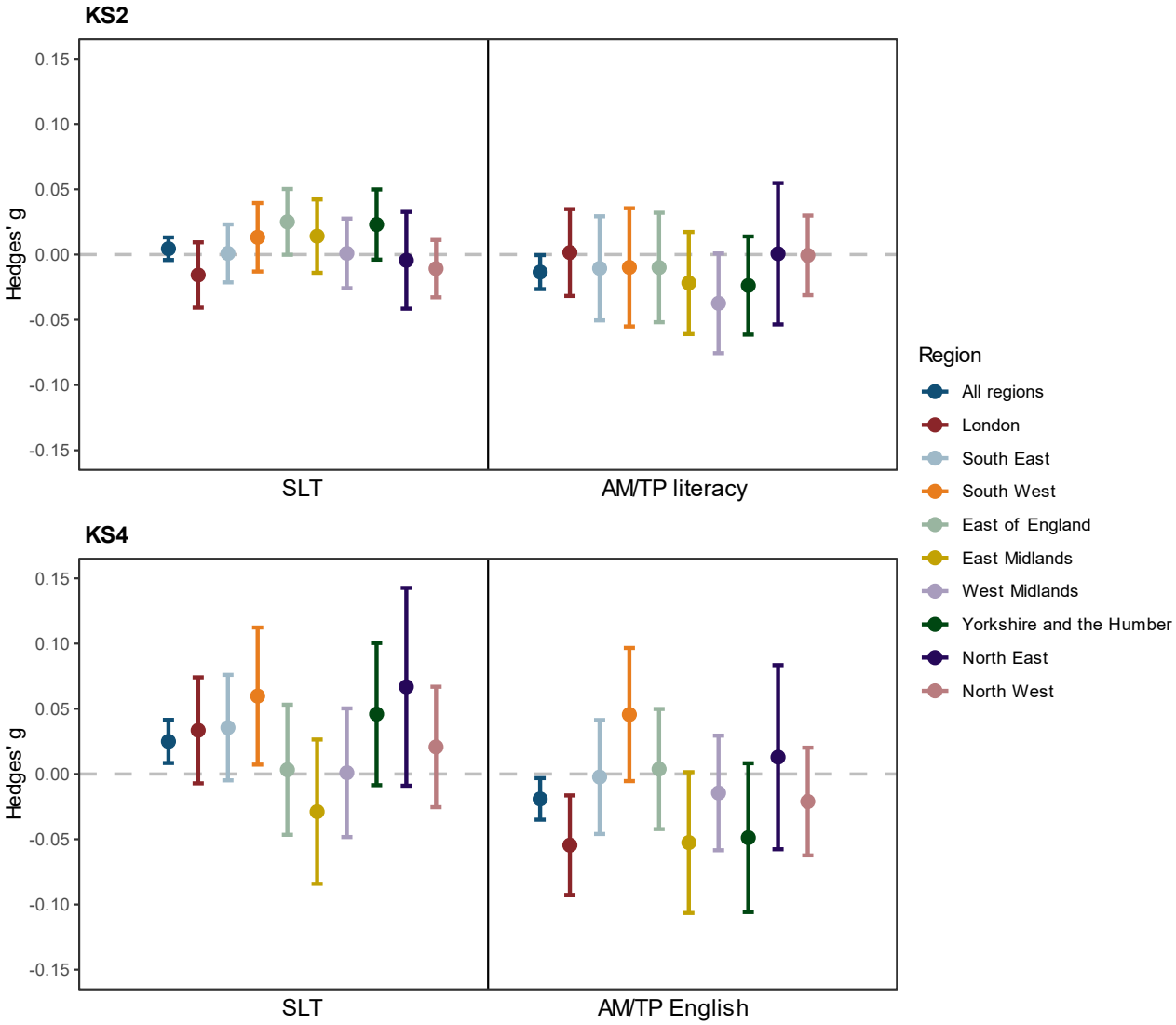
Route	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
SLT	KS2 reading	0.036	0.004 (-0.004, 0.013)	0.310
AM literacy	KS2 reading	0.054	0.007 (-0.021, -0.034)	0.624
TP literacy	KS2 reading	-0.120	-0.015 (-0.029, -0.001)	0.036
AM & TP literacy	KS2 reading	-0.675	-0.085 (-0.157, -0.013)	0.021
SLT	KS4 English Language	0.047	0.025 (0.008, 0.042)	0.003
AM English	KS4 English Language	0.071	0.038 (-0.008, 0.084)	0.107
TP English	KS4 English Language	-0.048	-0.026 (-0.042, -0.009)	0.002
AM & TP English	KS4 English Language	0.157	0.083 (-0.007, 0.174)	0.071

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

5.4 How does the impact of the NTP on English outcomes vary by geographic region?

Figure 8 illustrates the overall impact of the NTP on English outcomes within each of the nine geographic regions in England (see Table 15 for full statistical details). Because the number of schools in each region is much smaller than the sample for England as a whole, the confidence intervals around these estimates are correspondingly larger. Although there does appear to be some variation by region, particularly for KS4 outcomes, all the confidence intervals for the individual regions overlap with those for England as a whole. This suggests that the variations that can be seen are likely to be the result of sample variability rather than reflecting meaningful regional differences.

Figure 8: The impact of the NTP on English outcomes at school level by region



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 15: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on English outcomes by region

Route	Region	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
SLT	London	KS2 reading	-0.125	-0.016 (-0.041, 0.009)	0.220
	South East	KS2 reading	0.007	0.001 (-0.021, 0.023)	0.937
	South West	KS2 reading	0.106	0.013 (-0.013, 0.040)	0.322
	East of England	KS2 reading	0.199	0.025 (-0.000, 0.050)	0.051
	East Midlands	KS2 reading	0.112	0.014 (-0.014, 0.042)	0.324
	West Midlands	KS2 reading	0.007	0.001 (-0.026, 0.028)	0.947
	Yorkshire and the Humber	KS2 reading	0.183	0.023 (-0.004, 0.050)	0.093
	North East	KS2 reading	-0.035	-0.004 (-0.042, 0.033)	0.814
	North West	KS2 reading	-0.859	-0.011 (-0.033, 0.011)	0.335
AM/TP English	London	KS2 reading	0.012	0.002 (-0.032, 0.035)	0.928
	South East	KS2 reading	-0.084	-0.011 (-0.050, 0.029)	0.603
	South West	KS2 reading	-0.078	-0.010 (-0.055, 0.035)	0.672
	East of England	KS2 reading	-0.079	-0.010 (-0.052, 0.032)	0.643

Route	Region	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
	East Midlands	KS2 reading	-0.174	-0.022 (-0.061, 0.017)	0.275
	West Midlands	KS2 reading	-0.297	-0.037 (-0.076, 0.001)	0.055
	Yorkshire and the Humber	KS2 reading	-0.189	-0.024 (-0.061, 0.014)	0.216
	North East	KS2 reading	0.005	0.001 (-0.054, 0.055)	0.982
	North West	KS2 reading	-0.005	-0.001 (-0.031, 0.030)	0.967
SLT	London	KS4 English Language	0.063	0.033 (-0.007, 0.074)	0.106
	South East	KS4 English Language	0.067	0.036 (-0.005, 0.076)	0.085
	South West	KS4 English Language	0.113	0.060 (0.007, 0.112)	0.026
	East of England	KS4 English Language	0.006	0.003 (-0.047, 0.053)	0.898
	East Midlands	KS4 English Language	-0.054	-0.029 (-0.084, 0.026)	0.306
	West Midlands	KS4 English Language	0.002	0.001 (-0.048, 0.050)	0.969
	Yorkshire and the Humber	KS4 English Language	0.087	0.046 (-0.009, 0.100)	0.098
	North East	KS4 English Language	0.126	0.067 (-0.009, 0.143)	0.084
	North West	KS4 English Language	0.039	0.021 (-0.025, 0.067)	0.378

Route	Region	Outcome measure	Additional standard score points	Hedges' g (95% CI)	p value
AM/TP English	London	KS4 English Language	-0.103	-0.055 (-0.093, -0.016)	0.005
	South East	KS4 English Language	-0.004	-0.002 (-0.046, 0.041)	0.918
	South West	KS4 English Language	0.086	0.046 (-0.005, 0.097)	0.080
	East of England	KS4 English Language	0.007	0.004 (-0.042, 0.050)	0.872
	East Midlands	KS4 English Language	-0.099	-0.053 (-0.107, 0.001)	0.056
	West Midlands	KS4 English Language	-0.027	-0.015 (-0.058, 0.029)	0.517
	Yorkshire and the Humber	KS4 English Language	-0.092	-0.049 (-0.106, 0.008)	0.094
	North East	KS4 English Language	0.024	0.013 (-0.058, 0.084)	0.720
	North West	KS4 English Language	-0.040	-0.021 (-0.062, 0.020)	0.317

Note: bold denotes results that reached statistical significance at $p < 0.05$.

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

6 How does the school-level impact of the NTP on outcomes vary between groups of pupils with different characteristics?

Key Finding

- There is no evidence that the impact of the NTP varied significantly between any of the analysed subgroups.

This subsection presents the results of analysis exploring whether the school-level impact of the NTP on KS2 and KS4 English and maths attainment varies between groups of pupils with different characteristics in participating school versus comparison schools.

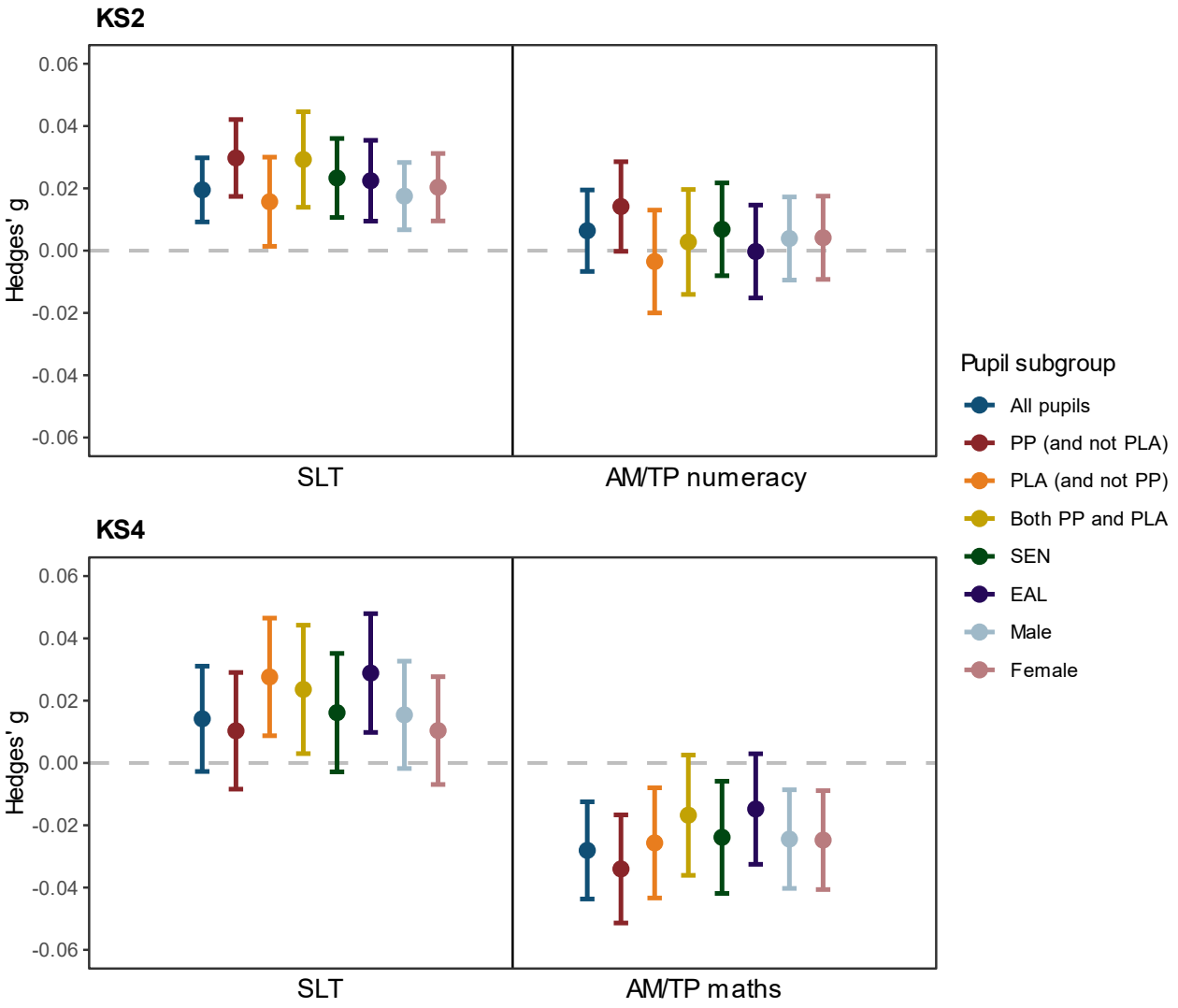
6.1 How does the school-level impact of the NTP on maths outcomes vary between groups of pupils with different characteristics?

For SLT, school participation was generally associated with better maths outcomes at both KS2 and KS4 relative to the comparison group. Figure 9 illustrates how the impact of the NTP on KS2 and KS4 maths outcomes vary by pupil characteristics. At KS2, the positive effects we detected across all of these groups were statistically significant. At KS4, the positive effects observed were statistically significant for pupils who were PLA only, pupils who were both PP and PLA and pupils with EAL, but the magnitude of the effects were similar across all groups (see Table 16 for full statistical details including effect sizes and confidence intervals).

For AM/TP, school participation was generally associated with better maths outcomes at KS2 and lower maths outcomes at KS4 relative to the comparison groups. However, none of the positive effects detected at KS2 were statistically significant. At KS4, the negative effects detected were statistically significant for pupils who were PP only, pupils who were PLA only, pupils with SEN, and both male and female pupils. Again, the magnitude of these effects was similar across all groups (see Table 16 for full statistical details including effect sizes and confidence intervals).

The effect sizes detected in this analysis are consistent with the effect sizes detected in relation to the impact of NTP on outcomes across all pupils in Year 6 and Year 11 discussed in Chapter 4. Furthermore, this analysis indicates the impact of the NTP on each of the pupil groups comprising the PP and/or PLA subgroup was similar across pupils who are PP only, PLA only and both PP and PLA only. The magnitude of these effect sizes is also broadly consistent with the effect detected on outcomes across the whole PP and/or PLA subgroup, for both Year 6 and Year 11.

Figure 9: The impact of the NTP on Maths outcomes at school level by pupil characteristic



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 16: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on Maths outcomes by pupil characteristic

Route	Outcome	Pupil characteristic	Additional standard score points	Additional months progress[27]	Hedges' g (95% CI)	P value
SLT	School-level KS2	PP only	0.228	0	0.030 (0.017, 0.042)	0.000
SLT	School-level KS2	PLA only	0.120	0	0.016 (0.001, 0.030)	0.032
SLT	School-level KS2	PP and PLA	0.224	0	0.029 (0.014, 0.045)	0.000
SLT	School-level KS2	SEN in intervention school	0.179	0	0.023 (0.011, 0.036)	0.000
SLT	School-level KS2	EAL	0.172	0	0.022 (0.009, 0.035)	0.001
SLT	School-level KS2	Male	0.134	0	0.018 (0.007, 0.028)	0.001
SLT	School-level KS2	Female	0.156	0	0.020 (0.010, 0.031)	0.000
AM/TP numeracy	School-level KS2	PP only	0.108	0	0.014 (-0.000, 0.029)	0.054
AM/TP numeracy	School-level KS2	PLA only	-0.027	0	-0.003 (-0.020, 0.013)	-0.027
AM/TP numeracy	School-level KS2	PP and PLA	0.021	0	0.003 (-0.014, 0.020)	0.745
AM/TP numeracy	School-level KS2	SEN in intervention school	0.052	0	0.007 (-0.008, 0.022)	0.368

[27] Defined according to EEF (2023).

Route	Outcome	Pupil characteristic	Additional standard score points	Additional months progress[27]	Hedges' g (95% CI)	P value
AM/TP numeracy	School-level KS2	EAL	-0.002	0	0.000 (-0.015, 0.015)	0.970
AM/TP numeracy	School-level KS2	Male	0.030	0	0.004 (-0.009, 0.017)	0.567
AM/TP numeracy	School-level KS2	Female	0.032	0	0.004 (-0.009, 0.018)	0.544
SLT	School-level KS4	PP only	0.021	0	0.010 (-0.008, 0.029)	0.280
SLT	School-level KS4	PLA only	0.057	0	0.028 (0.009, 0.047)	0.004
SLT	School-level KS4	PP and PLA	0.049	0	0.024 (0.003, 0.044)	0.025
SLT	School-level KS4	SEN in intervention school	0.034	0	0.016 (-0.003, 0.035)	0.096
SLT	School-level KS4	EAL	0.060	0	0.029 (0.010, 0.048)	0.003
SLT	School-level KS4	Male	0.032	0	0.015 (-0.002, 0.033)	0.079
SLT	School-level KS4	Female	0.022	0	0.010 (-0.007, 0.028)	0.239
AM/TP maths	School-level KS4	PP only	-0.071	0	-0.034 (-0.051, -0.017)	< 0.001
AM/TP maths	School-level KS4	PLA only	-0.053	0	-0.026 (-0.043, -0.008)	0.004

Route	Outcome	Pupil characteristic	Additional standard score points	Additional months progress[27]	Hedges' g (95% CI)	P value
AM/TP maths	School-level KS4	PP and PLA	-0.035	0	-0.017 (-0.036, 0.003)	0.089
AM/TP maths	School-level KS4	SEN in intervention school	-0.050	0	-0.024 (-0.042, -0.006)	0.009
AM/TP maths	School-level KS4	EAL	-0.031	0	-0.015 (-0.033, 0.003)	0.102
AM/TP maths	School-level KS4	Male	-0.051	0	-0.024 (-0.040, -0.009)	0.002
AM/TP maths	School-level KS4	Female	-0.051	0	-0.025 (-0.041, -0.009)	0.002

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

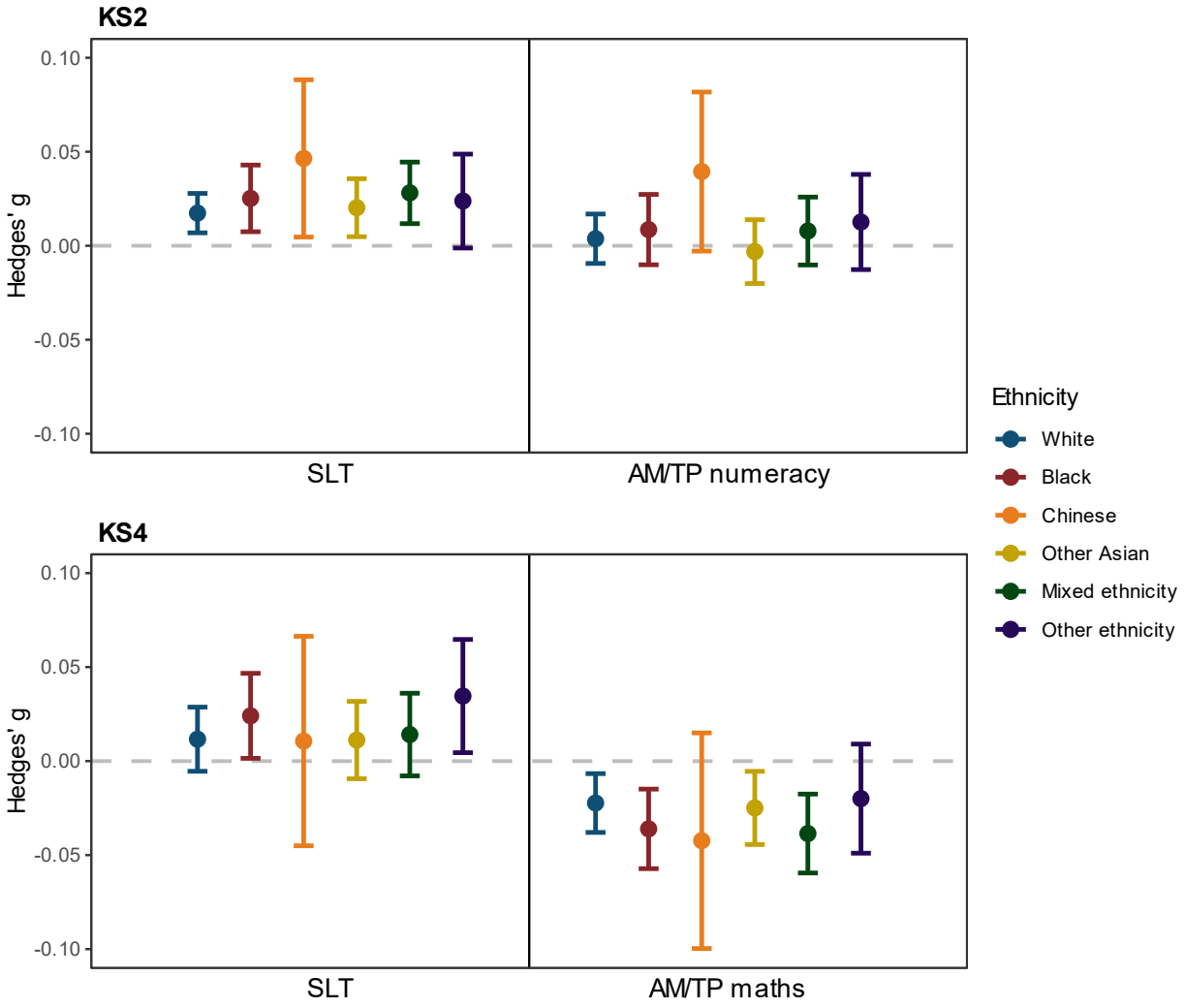
We also looked at the impact of the NTP on maths outcomes by ethnicity, as shown in Figure 10 and Table 17**,** and these findings were largely consistent with the findings presented above.

SLT participation was generally associated with better KS2 and KS4 maths outcomes relative to the comparison group. The effect sizes we detected were again small and equivalent to no additional months' progress (EEF, 2021c). At KS2, the effect sizes range from Hedge's g = 0.017 for pupils from white ethnic backgrounds to Hedge's g = 0.046 for pupils from Chinese ethnic backgrounds, with the majority of confidence intervals overlapping. Similarly, at KS4, the effect sizes range from Hedges' g = 0.011 for Chinese pupils to 0.035 for pupils from other ethnic backgrounds. Again, we found that the confidence intervals for these effect sizes overlapped.

For AM/TP, we again found that participation was generally associated with better maths outcomes in KS2 and lower maths outcomes in KS4. The effect sizes we detected were again small and equivalent to no additional months' progress (EEF, 2021c). At KS2, the effect sizes range from Hedge's g = -0.003 for pupils from other Asian backgrounds up to Hedge's g = 0.039 for pupils from Chinese backgrounds, with the majority of confidence intervals overlapping. For KS4, the effect sizes range from Hedge's g -0.042 for Chinese pupils to -0.020 for pupils from other ethnic backgrounds. Again, the confidence intervals of these effect sizes overlapped.

These effect sizes indicate that pupils from a Chinese ethnicity background may be more sensitive to the effects of tutoring on their maths outcomes (both positive and negative) than the other ethnic groups, particularly in KS4. However, it should be noted that sample size for this group is small and the confidence intervals for this effect size are very wide and so this needs to interpreted with caution. The impact of the SLT and AM/TP was similar across the rest of the ethnic backgrounds and are consistent with the impact of the SLT and AM/TP detected across all pupils in Year 6 and Year 11.

Figure 10: The school-level impact of the NTP on KS2 and KS4 maths outcomes for pupils from different ethnic backgrounds



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 17: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on Maths outcomes by pupil ethnic background

Route	Outcome	Ethnicity	Additional standard score points	Additional months progress[28]	Hedges' g (95% CI)	P value
SLT	School-level KS2	White	0.133	0	0.017 (0.007, 0.028)	0.001
SLT	School-level KS2	Black	0.193	0	0.025 (0.007, 0.043)	0.005
SLT	School-level KS2	Chinese	0.356	0	0.046 (0.005, 0.088)	0.029
SLT	School-level KS2	Other Asian	0.155	0	0.020 (0.005, 0.036)	0.010
SLT	School-level KS2	Mixed Ethnicity	0.216	0	0.028 (0.012, 0.045)	0.001
SLT	School-level KS2	Other Ethnicity	0.182	0	0.024 (-0.001, 0.049)	0.062
AM/TP numeracy	School-level KS2	White	0.029	0	0.004 (-0.009, 0.017)	0.578
AM/TP numeracy	School-level KS2	Black	0.066	0	0.009 (-0.010, 0.027)	0.368
AM/TP numeracy	School-level KS2	Chinese	0.302	0	0.039 (-0.003. 0.082)	0.068
AM/TP numeracy	School-level KS2	Other Asian	-0.024	0	-0.003 (-0.020, 0.014	0.719
AM/TP numeracy	School-level KS2	Mixed Ethnicity	0.060	0	0.008 (-0.010, 0.026)	0.395

[28] Defined according to EEF (2023).

Route	Outcome	Ethnicity	Additional standard score points	Additional months progress[28]	Hedges' g (95% CI)	P value
AM/TP numeracy	School-level KS2	Other Ethnicity	0.097	0	0.013 (-0.013, 0.038)	0.328
SLT	School-level KS4	White	0.024	0	0.012 (-0.005, 0.029)	0.180
SLT	School-level KS4	Black	0.050	0	0.024 (0.001, 0.047)	0.037
SLT	School-level KS4	Chinese	0.022	0	0.011 (-0.045, 0.066)	0.707
SLT	School-level KS4	Other Asian	0.023	0	0.011 (-0.009, 0.032)	0.287
SLT	School-level KS4	Mixed Ethnicity	0.029	0	0.014 (-0.008, 0.036)	0.208
SLT	School-level KS4	Other Ethnicity	0.072	0	0.035 (0.004, 0.065)	0.024
AM/TP maths	School-level KS4	White	-0.046	0	-0.022 (-0.038, -0.007)	0.005
AM/TP maths	School-level KS4	Black	-0.075	0	-0.036 (-0.057, -0.015)	0.001
AM/TP maths	School-level KS4	Chinese	-0.088	0	-0.042 (-0.100, 0.015)	0.148
AM/TP maths	School-level KS4	Other Asian	-0.052	0	-0.025 (-0.044, -0.005)	0.012
AM/TP maths	School-level KS4	Mixed Ethnicity	-0.080	0	-0.039 (-0.059, -0.018)	0.000

Route	Outcome	Ethnicity	Additional standard score points	Additional months progress[28]	Hedges' g (95% CI)	P value
AM/TP maths	School-level KS4	Other Ethnicity	-0.041	0	-0.020 (-0.049, 0.009)	0.178

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

6.2 How does the school-level impact of the NTP on English outcomes vary between groups of pupils with different characteristics?

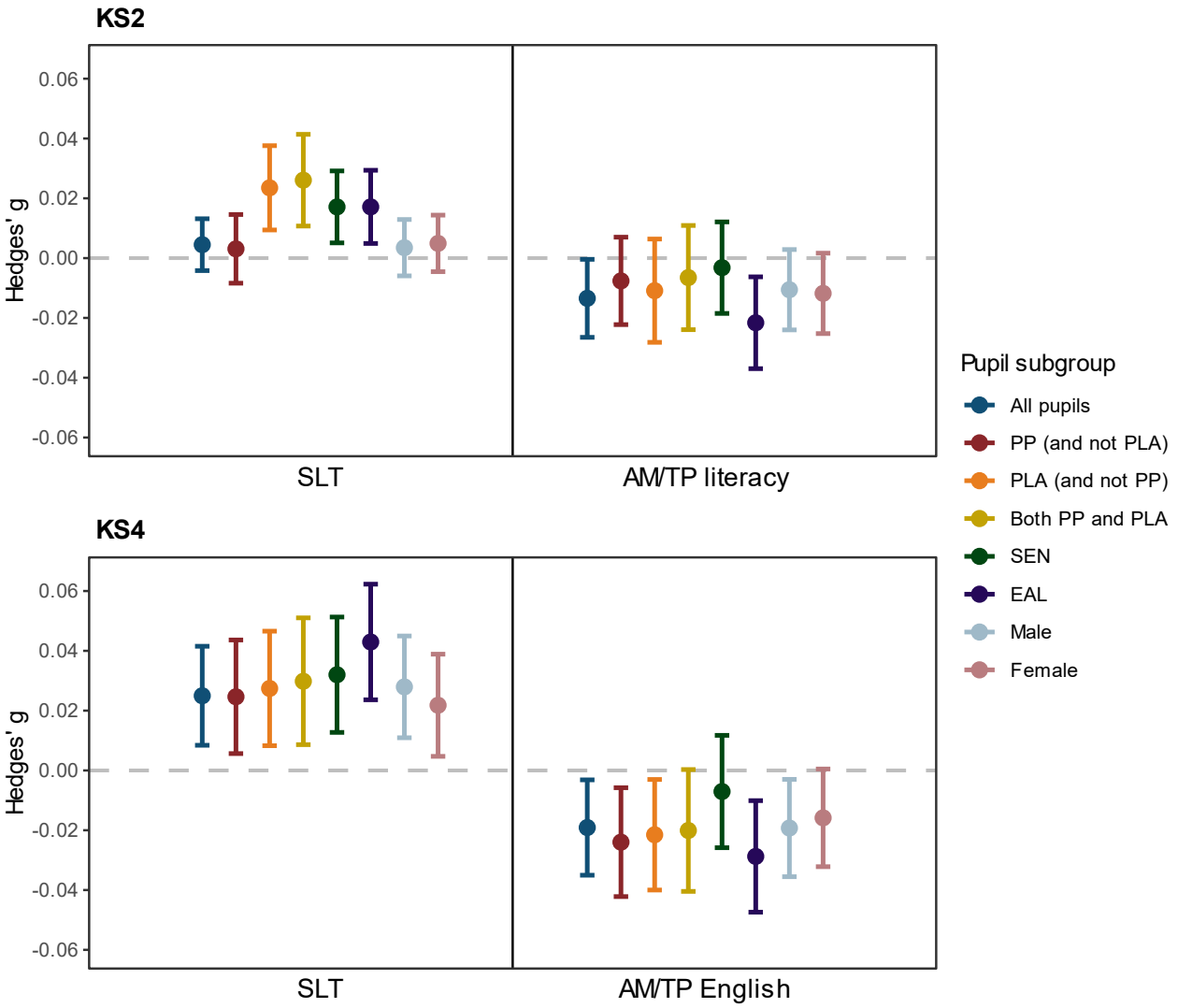
Figure 11 illustrates how the impact of the NTP on KS2 and KS4 English outcomes vary by pupil characteristics. The results are discussed in the following sub-sections.

For SLT, school participation was generally associated with better English outcomes at both KS2 and KS4 relative to the comparison group. Figure 11 illustrates how the impact of the NTP on KS2 and KS4 English outcomes vary by pupil characteristic. At KS2, the positive effects we detected for pupils in the following groups were statistically significant: PP only, PLA only, both PP and PLA, SEN and EAL. At KS4, the positive effects we observed were statistically significant across of these characteristic groups. (see Table 18 for full statistical details including effect sizes and confidence intervals).

For AM//TP, school participation was generally associated with lower English outcomes at both KS2 and KS4 relative the comparison group, as shown in Figure 11. At KS2, only the negative effect observed for pupils with EAL was statistically significant. At KS4, the negative effects we detected were statistically significant for PP only pupils, PLA only pupils, pupils with EAL and male pupils versus the comparison group. This is shown in Table 18**.**

The effect sizes we detected for both SLT and AM/TP and equivalent to no additional month's progress (EEF, 2023). For SLT, the effects range from Hedges' g = 0.003 (KS2; male) to 0.043 (KS4; EAL) and are equivalent to no additional month's progress (EEF, 2021c). For AM/TP, the effect sizes range from Hedges' g = -0.029 (KS4, EAL) to -0.003 (KS2, SEN). The effect sizes detected in this analysis are consistent with the effect sizes detected in relation to the impact of NTP on outcomes across all pupils in Year 6 and Year 11 discussed in Chapter 4. Furthermore, this analysis indicates the impact of the NTP on each of the pupil groups comprising the PP and/or PLA subgroup was similar across pupils who are PP only, PLA only and both PP and PLA only. The magnitude of these effect sizes is also broadly consistent with the effect detected on outcomes across the PP and/or PLA subgroup as a whole both for Year 6 and Year 11.

Figure 11: The impact of the NTP on English outcomes at school level by pupil characteristic



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 18: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on English outcomes

Route	Outcome	Pupil characteristic	Additional standard score points	Additional months progress[29]	Hedges' g (95% CI)	P value
SLT	School-level KS2	PP only	0.025	0	0.023 (0.009, 0.038)	0.001
SLT	School-level KS2	PLA only	0.187	0	0.016 (0.001, 0.030)	0.032
SLT	School-level KS2	PP and PLA	0.207	0	0.026 (0.011, 0.043)	0.001
SLT	School-level KS2	SEN in intervention school	0.136	0	0.017 (0.005, 0.029)	0.005
SLT	School-level KS2	EAL	0.136	0	0.017 (0.005, 0.029)	0.006
SLT	School-level KS2	Male	0.028	0	0.003 (-0.006, 0.013)	0.469
SLT	School-level KS2	Female	0.039	0	0.005 (-0.005, 0.014)	0.308
AM/TP literacy	School-level KS2	PP only	-0.061	0	-0.008 (-0.022, 0.007)	0.307
AM/TP literacy	School-level KS2	PLA only	-0.087	0	-0.011 (-0.028, 0.006)	0.216
AM/TP literacy	School-level KS2	PP and PLA	-0.052	0	-0.007 (-0.024, 0.012)	0.464
AM/TP literacy	School-level KS2	SEN in intervention school	-0.025	0	-0.003 (-0.018, 0.012)	0.682

[29] Defined according to EEF (2023).

Route	Outcome	Pupil characteristic	Additional standard score points	Additional months progress[29]	Hedges' g (95% CI)	P value
AM/TP literacy	School-level KS2	EAL	-0.172	0	-0.022 (-0.037, -0.006)	0.006
AM/TP literacy	School-level KS2	Male	-0.084	0	-0.011 (-0.024, 0.003)	0.123
AM/TP literacy	School-level KS2	Female	-0.094	0	-0.012 (-0.025, 0.002)	0.086
SLT	School-level KS4	PP only	0.046	0	0.025 (0.006, 0.044)	0.011
SLT	School-level KS4	PLA only	0.052	0	0.027 (0.008, 0.047)	0.005
SLT	School-level KS4	PP and PLA	0.056	0	0.030 (0.009, 0.051)	0.006
SLT	School-level KS4	SEN in intervention school	0.060	0	0.032 (0.013, 0.051)	0.001
SLT	School-level KS4	EAL	0.081	0	0.043 (0.024, 0.062)	0.000
SLT	School-level KS4	Male	0.053	0	0.028 (0.011, 0.045)	0.001
SLT	School-level KS4	Female	0.041	0	0.022 (0.005, 0.039)	0.012
AM/TP English	School-level KS4	PP only	-0.045	0	-0.024 (-0.042, -0.006)	0.010
AM/TP English	School-level KS4	PLA only	-0.041	0	-0.022 (-0.040, -0.003)	0.023

Route	Outcome	Pupil characteristic	Additional standard score points	Additional months progress[29]	Hedges' g (95% CI)	P value
AM/TP English	School-level KS4	PP and PLA	-0.038	0	-0.020 (-0.040, 0.000)	0.053
AM/TP English	School-level KS4	SEN in intervention school	-0.013	0	-0.007 (-0.026, 0.012)	0.462
AM/TP English	School-level KS4	EAL	-0.054	0	-0.029 (-0.047, -0.010)	0.003
AM/TP English	School-level KS4	Male	-0.036	0	-0.019 (-0.036, -0.003)	0.020
AM/TP English	School-level KS4	Female	-0.030	0	-0.016 (-0.032, 0.000)	0.057

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

We also looked at the impact of the NTP by ethnicity, as shown in Figure 12 and Table 19, and these findings were largely consistent with the findings presented above.

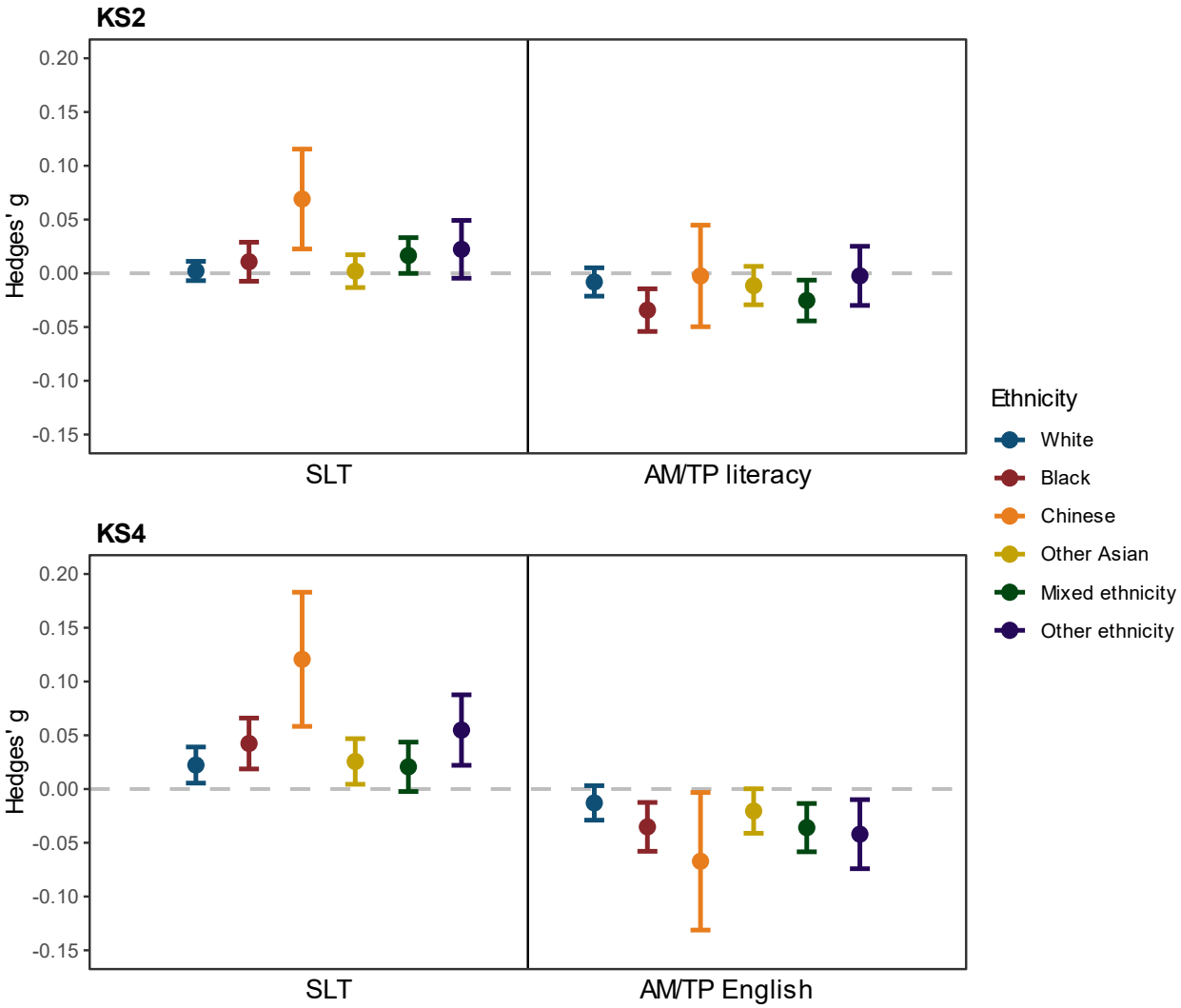
SLT participation was generally with better KS2 and KS4 English outcomes relative to the comparison group. The effect sizes we detected were again small and equivalent to no additional months' progress (EEF, 2021c). At KS2, the effect sizes range from Hedge's g = 0.002 for pupils from other Asian backgrounds up to Hedge's g = 0.069 for pupils from Chinese backgrounds, with the majority of confidence intervals overlapping. Similarly, at KS4, the effect sizes range from Hedges' g = 0.021 for pupils from mixed ethnic backgrounds to 0.121 for Chinese pupils (though the confidence intervals for this effect are particularly wide) and that the confidence intervals overlapped with one another.

For AM/TP, we again found that participation was generally associated with lower maths outcomes in KS2 and KS4. At KS2, the effect sizes range from Hedge's g = -0.002 for pupils from other ethnic backgrounds up to Hedge's g = -0.034 for pupils from black ethnic backgrounds, with the majority of confidence intervals overlapping once again. For KS4, the effect sizes range from -0.067 for Chinese pupils to -0.013 for white pupils and that the confidence intervals for these effect sizes again overlapped.

These effect sizes indicate that pupils from a Chinese ethnicity background may be more sensitive to the effects of tutoring on their English outcomes (both positive and negative) than the other ethnic groups, particularly in relation to SLT. However, it should be noted

that sample size for this group is small and the confidence intervals for this effect size are very wide and so this needs to be interpreted with caution. The impact of the SLT and AM/TP was similar across the rest of the ethnic backgrounds and are consistent with the impact of the SLT and AM/TP detected across all pupils in Year 6 and Year 11.

Figure 12: The school-level impact of the NTP on KS2 and KS4 reading outcomes for pupils from different ethnic backgrounds



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 19: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on English outcomes by pupil ethnic background

Route	Outcome	Ethnicity	Additional standard score points	Additional months progress[30]	Hedges' g (95% CI)	P value
SLT	School-level KS2	White	0.017	0	0.002 (-0.007, 0.011)	0.640
SLT	School-level KS2	Black	0.085	0	0.011 (-0.007, 0.029)	0.246
SLT	School-level KS2	Chinese	0.549	1	0.069 (0.023, 0.115)	0.004
SLT	School-level KS2	Other Asian	0.016	0	0.002 (-0.013, 0.017)	0.801
SLT	School-level KS2	Mixed Ethnicity	0.131	0	0.017 (0.000, 0.033)	0.052
SLT	School-level KS2	Other Ethnicity	0.177	0	0.022 (-0.005, 0.049)	0.104
AM/TP literacy	School-level KS2	White	-0.065	0	-0.008 (-0.021, 0.005)	0.228
AM/TP literacy	School-level KS2	Black	-0.273	0	-0.034 (-0.054, -0.014)	0.001
AM/TP literacy	School-level KS2	Chinese	-0.020	0	-0.002 (-0.050, 0.045)	0.917
AM/TP literacy	School-level KS2	Other Asian	-0.091	0	-0.011 (-0.029, 0.006)	0.210
AM/TP literacy	School-level KS2	Mixed Ethnicity	-0.202	0	-0.025 (-0.044, -0.006)	0.009

[30] Defined according to EEF (2023).

Route	Outcome	Ethnicity	Additional standard score points	Additional months progress[30]	Hedges' g (95% CI)	P value
AM/TP literacy	School-level KS2	Other Ethnicity	-0.019	0	-0.002 (-0.030, 0.025)	0.863
SLT	School-level KS4	White	0.042	0	0.022 (0.005, 0.039)	0.009
SLT	School-level KS4	Black	0.080	0	0.042 (0.019, 0.066)	0.000
SLT	School-level KS4	Chinese	0.227	2	0.121 (0.058, 0.183)	0.000
SLT	School-level KS4	Other Asian	0.048	0	0.026 (0.004, 0.047)	0.018
SLT	School-level KS4	Mixed Ethnicity	0.039	0	0.021 (-0.002, 0.044)	0.078
SLT	School-level KS4	Other Ethnicity	0.103	1	0.055 (0.022, 0.088)	0.001
AM/TP English	School-level KS4	White	-0.024	0	-0.013 (-0.029, 0.003)	0.114
AM/TP English	School-level KS4	Black	-0.066	0	-0.035 (-0.058, -0.012)	0.002
AM/TP English	School-level KS4	Chinese	-0.127	-	-0.067 (-0.131, -0.003)	0.040
AM/TP English	School-level KS4	Other Asian	-0.039	0	-0.020 (-0.041, 0.000)	0.052
AM/TP English	School-level KS4	Mixed Ethnicity	-0.068	0	-0.036 (-0.058, -0.013)	0.002

Route	Outcome	Ethnicity	Additional standard score points	Additional months progress[30]	Hedges' g (95% CI)	P value
AM/TP English	School-level KS4	Other Ethnicity	-0.079	0	-0.042 (-0.074, -0.010)	0.010

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

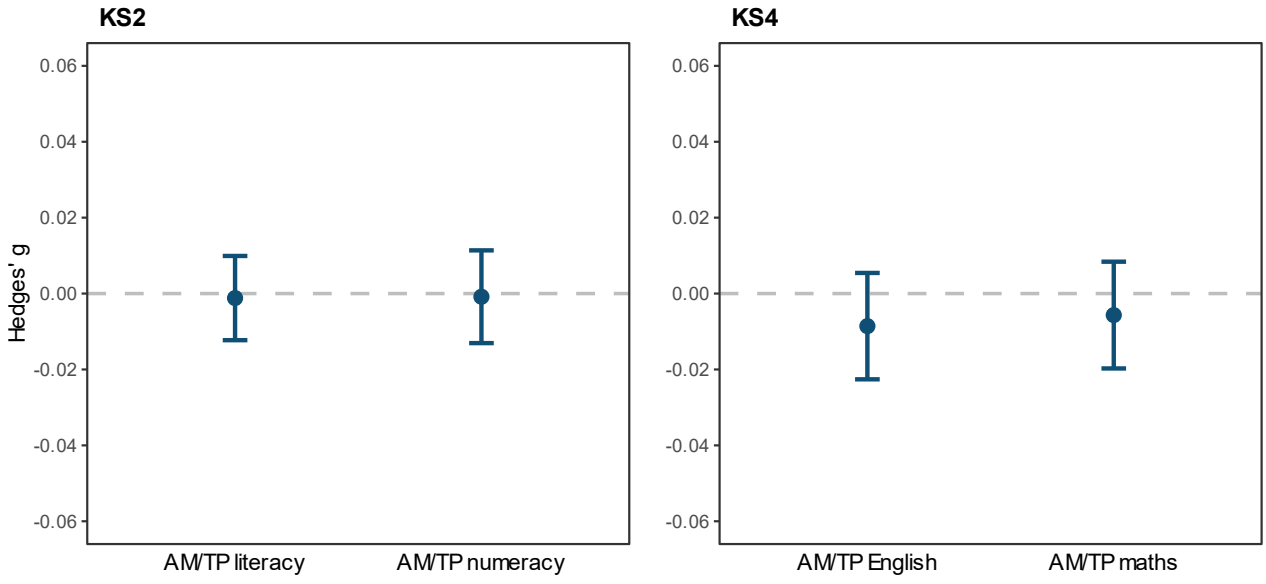
7 What was the longer-term impact of the NTP on school-level attainment for pupils who were involved in year 1 of the NTP (2020-21 to 2021-22)?

This section presents results of the analysis exploring the impact of taking part in the first year of the NTP in 2020-21 on school-level English/literacy and maths outcomes by the end of 2021-22 irrespective of whether schools went on to take part in the second year of the NTP in 2021-22. Note that in the first year of the NTP only the AM and TP routes were available and due to the cancellation of statutory tests there were significant methodological and data challenges. This means that this is the first time that the effect of participation in the first year of NTP has been evaluated using statutory tests.

Figure 13 illustrates the impact of participating in the first year of the NTP on KS2 and KS4 outcomes at school level in 2021-22. Participation in year 1 of the NTP was generally associated with slightly lower English and Maths outcomes at both KS2 and KS4, relative to the comparison group. None of these negative effects were statistically significant at school level and the effect sizes were all tiny, ranging from -0.009 to -0.001 (see Table 20 for full statistical details including effect sizes and confidence intervals).

These results suggest that school participation in the first year of the NTP did not benefit the longer-term English or maths outcomes at KS2 or KS4. These findings are consistent with the main analysis from the impact evaluation of the first year of the NTP (Poet *et al.*, 2022a). It is, however, worth noting that the year 1 impact evaluation did detect some positive effects of tutoring for schools with higher concentrations of tutoring.

Figure 13: The impact of participation in the NTP in 2020-21 on English and Maths outcomes at KS2 and KS4 in 2021-22



Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Table 20: Results of the linear mixed effects models exploring the impact of SLT and AM/TP on maths outcomes

Outcome	Additional standard score points	Additional months progress[31]	Hedges' g (95% CI)	P value
School-level KS2 Literacy	-0.009	0	-0.001 (-0.012, 0.010)	0.833
School-level KS2 Maths	-0.006	0	-0.001 (-0.013, -0.013)	0.893
School-level KS4 English	-0.016	0	-0.009 (-0.026, 0.005)	0.230
School-level KS4 Maths	-0.012	0	-0.006 (-0.020, 0.008)	0.429

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

[31] Defined according to EEF (2023).

8 Discussion

Overall, we found consistent evidence that participation in SLT led to small improvements in KS2 and KS4 maths outcomes. There was also some more limited evidence at school level only that participation in SLT was associated with small improvements in KS2 and KS4 English outcomes. For both English and maths some of these results reached statistical significance, but in all cases the effect sizes were very small and equated to one months' additional progress or less. We did not find any evidence that participation in AM/TP resulted in improved outcomes at a school level or pupil level, and we found some evidence which suggested participation AM/TP was associated with slightly worse outcomes in English (KS2 and KS4) and maths (KS4). However, given the potential for selection bias remaining despite our extensive attempts to remove it, the possibility remains that the small differences seen were artifacts of bias rather than genuine effects of the intervention on outcomes. Larger effects, which are reasonable to expect given this was a tutoring intervention (EEF, 2021b), would have allowed us to discount the possibility of lingering bias. Given these were not seen, we remain unable to confidently distinguish the effects from bias. For example, Weidmann and Miratrix (2020) show that although matching-adjusted effect estimates are consistent with there being no selection bias at the school level, the distribution of potential bias ranges from -0.1 to 0.1 standard deviations i.e. well above the effects seen in this study. Furthermore, population analyses have a tendency to generate spurious significant findings, particularly in a scenario of multiple testing. We can be confident that the effects of the different tutoring routes were not large.

That said, it is important to consider the trends observed in this analysis in case they are genuine. It is possible these differences in outcomes between NTP routes could be related to differences in implementation. DfE introduced the SLT route in the second year of the NTP in response to calls from schools for more autonomy over how they delivered tutoring. SLT allowed schools to use internal staff as tutors, making it more likely the staff delivering the NTP had existing relationships with staff and pupils they could build on and that tutoring was delivered in-person rather than online, especially when compared with TP where external providers were used. These factors may help explain why the implementation and process evaluation of the second year of the NTP showed that SLT had the highest levels of satisfaction amongst school leaders (Lynch *et al.*, 2022). This extended to all aspects of tutoring including the ability of tutors to meet pupils' learning needs, how well tutoring sessions aligned with the school curriculum, tutors' relationships with pupils, and the quality of tuition, all of which may have contributed to better outcomes for pupils. Interestingly, these results are similar to the results of meta-analysis exploring the effects of tutoring provision in the US. This study found stronger positive effects of tuition delivered by local school-district providers (similar to SLT) compared with tuition delivered by national commercial providers (similar to TP), especially in relation to maths outcomes (Chappell *et al.*, 2011). We note that although tutors for the AM route were recruited through external providers they became in-house members of staff at the school. This means they may have had more opportunities to build

relationships with staff and pupils and deliver tutoring in-person than tutors from tuition partners, which may help explain why the biggest differences in effects appear to be between SLT and TP.

In general, the results reported here for SLT are consistent with a large body of evidence indicating that small-group tuition is effective, especially when it is targeted at pupils' specific needs. However, the effect sizes observed in the present analysis, which equated to around one months' additional progress or less, were smaller than might be expected based on previous evidence about the effectiveness of small group tuition. The effect sizes previously detected vary across studies, but on average, small group tuition has been found to result in around two months additional progress for secondary school pupils and four months additional progress for primary school pupils (EEF, 2021b). Similarly, previous meta-analyses demonstrate that tutoring programmes consistently result in substantial positive impacts on pupils' learning outcomes, with average effect sizes ranging from 0.30 SD to 0.37 SD (Ritter *et al.*, 2009; Dietrichson *et al.*, 2017; Nickow, Oreopoulos and Quan, 2020). Although we note that the studies upon which these reviews were based were smaller-scale than the NTP evaluation, had very specific target populations, and generally had the ability to conduct pre- and post-tests for the evaluation. All of these factors make it more likely they would detect a larger effect size than we found here.

The smaller effect sizes observed in the present study are likely to be related to differences in the evaluation design and available data, and the subsequent limitations of the analysis approach which was feasible as well as difficulties implementing the NTP at scale, rather than tutoring being ineffective as an approach. Here the school-level analyses were subject to dilution effects as not all pupils within the intervention group were themselves selected for tuition, meaning that the effect sizes in these analyses will be under-estimates of the true effect (as discussed previously in sub-section 2.5). In addition, no data about the subject in which tutoring was received was available for SLT. This means that for SLT there were also dilution effects present at the pupil-level so our estimates of the SLT-related effect sizes will also be underestimates of the true effect.

Furthermore, as discussed earlier in Chapter 4, we would anticipate that the pupil-level effect sizes would be larger than those observed at school level as they do not suffer from dilution. The results of this analysis indicate that this was not consistently the case. It is likely that this relates at least in part to the interaction between the residual negative selection bias within the pupil-level analysis and the dilution effects in the SLT analysis due to the lack of subject information.

The wider literature on tutoring suggests that the smaller the group and the more aligned tuition is to pupils' needs the more effective it will be (EEF, 2021b). It is therefore also likely that challenges with delivering the NTP at scale including integrating tutoring with the school curriculum, tutoring often being conducted during normal lesson times (rather

than in addition to usual teaching and learning) and difficulties recruiting high-quality tutors, are contributing to the NTP having lower-than-anticipated observed effects. Previous research into the NTP indicates that not all schools have successfully integrated tutoring with the school curriculum. In these schools, the tutoring content delivered was often found to be generic and did not target the individual needs of pupils (Ofsted, 2022). Furthermore, some schools have not fully understood and/or implemented the delivery guidance for the NTP especially around the use of small groups and delivering sessions frequently, so are not implementing the NTP with fidelity to its intended delivery design (Ofsted, 2022). Ofsted (Ofsted, 2022) also highlighted that scheduling tutoring sessions was challenging for schools as they sought to identifying times that were convenient for teachers, tutors, and pupils, and maximise attendance. As a result, many schools were providing tutoring during the school day, including during lesson time. This is consistent with previous evidence that three-fifths of schools were delivering SLT sessions during lesson time in the school day (Lynch *et al.*, 2022), and with findings from a recent survey where 47 per cent of school leaders reported that their school only offers tutoring during normal lesson times (Moore and Lord, 2023). Tutoring is therefore often replacing and potentially disrupting the regular classroom-based teaching and learning participating pupils receive, rather than being strictly additional to normal lessons, which may be offsetting some of the potential benefit of participating in tutoring.

Finding high quality, suitably qualified tutors has been a challenge for some schools for various reasons, including candidates lacking suitable knowledge and behaviour management skills. In addition, although not all schools were eligible for AM, among schools that were eligible, the supply of tutors did not meet demand and some schools were dissatisfied with the time taken to match them with a tutor. Some schools have also reported negative experiences of engaging with TPs, for example difficulties engaging with the online booking system and being let down at the last minute by providers (Lynch *et al.*, 2022, forthcoming). Similarly, ongoing recruitment and retention issues in conjunction with high workloads among teachers (McLean, Worth and Faulkner-Ellis, 2023) meant that schools also reported challenges finding sufficient capacity among internal staff to provide high quality tutors for the SLT route (Lynch *et al.*, 2022). This appears to be an on-going issue, as a recent NFER survey found that only 36 per cent of school leaders planning to continue using the NTP in the academic year 2023-24 felt confident that they could continue to access suitable tutors in the future (Moore and Lord, 2023). There is therefore considerable evidence to indicate that issues with implementation may be impeding the potential effectiveness of the NTP.

For SLT a higher tutoring dosage and/or concentration was consistently associated with better outcomes at a school level, but this was not the case for AM/TP. However, these associations may not be causal given fresh matching of schools was not carried out for the analyses. This suggests that for SLT providing more hours of tutoring to pupils can lead to better outcomes at a school level. These results are somewhat inconsistent with the findings from the impact evaluation of the first year of the NTP. This evaluation found some evidence to indicate that among schools that participated in TP, more hours of tuition in the relevant subject was associated with better teacher-assessed grades in both

maths and English for Year 11 pupils and better reading outcomes for primary school pupils (Poet *et al.*, 2022b, 2022a). However, in both cases the evaluation methods used for the dosage analysis were different from the approach taken here, which used a school-averaged dosage measure and was therefore also partly contingent on tutoring concentration. Nevertheless, it is important to consider the potential impact of tutoring dosage on outcomes, as a previous review by EEF suggests that frequent tutoring sessions, three times a week lasting up to one hour for approximately 10 weeks is most effective for improving outcomes (EEF, 2021b). Meanwhile the NTP recommends that pupils receive around 12-15 hours of tuition, which only equates to around half of the total dosage suggested by EEF. To maximise the benefits of tutoring, it is important to increase the evidence base around best-practice for implementing tutoring in schools to determine for example, the optimum dosage, group size, delivery mode (online/in-person) and number of sessions per week, as well as how best to schedule the sessions and ensure that session content aligns with the school curriculum.

The key aim of the NTP was to help reduce the attainment gap for disadvantaged pupils. Here, we found that participation in SLT was associated with small positive impacts on English and maths outcomes for PP and/or PLA pupils as well as for all pupils. This is consistent with the results of a previous meta-analysis which suggests that disadvantaged pupils can benefit from small group tuition (Dietrichson *et al.*, 2017). Although we did not find any evidence to suggest that PP pupils benefited more from tuition than non-PP pupils, or that there were differences in the impact of the NTP related to any other pupil characteristics. In addition, although PP pupils were selected for the NTP at higher rates than would be expected relative to the population, PP pupils made up less than half of the pupils selected for the NTP. This suggests that although there is some evidence that schools are prioritising PP pupils for the NTP, more could be done to ensure disadvantaged pupils receive this additional support. Successfully closing the attainment gap for disadvantaged pupils is likely to be a gradual process that takes place over several years, therefore sufficient, sustained funding is needed to allow disadvantaged pupils to be consistently supported.

9 Recommendations

This report offers insights into the impact of the NTP on KS2 and KS4 maths and English outcomes. Based on the results described in this report, we have outlined a series of recommendations to help improve the effectiveness of the NTP and maximise outcomes for pupils. These recommendations are primarily for government but may also be of interest to schools continuing to implement the NTP.

- There is tentative evidence that the introduction of the SLT route has been successful. Further research into how to optimise the delivery of tuition is needed to be able to offer guidance to schools on which type of implementation is most effective. For example, this could include research using randomised controlled trial designs (RCTs) to build the evidence base around best practice in tutoring optimum tutoring dosage, session duration, frequency, mode of delivery (online versus in-person), how best to align sessions with the school curriculum and time of delivery (during the school day or outside of normal teaching hours).
- Future research on tutoring should collect data on tutoring routes and subject, to allow for continued monitoring of effectiveness by route and subject to further develop our understanding of 'what works'.
- To help close the attainment gap for disadvantaged pupils, consider reintroducing targets for the delivery of tutoring to disadvantaged pupils, and using funding to incentivise the selection of these pupils for tutoring to ensure they are prioritised for additional support.
- We found that the impact of tutoring increased with the average number of tutoring hours pupils received, indicating that more hours of tuition can lead to greater benefits. Previous evidence from EEF (2021) also indicates that around 30 hours of tuition delivered over approximately 10 weeks has the greatest impact. We recommend that the NTP guidance reflects these findings on tutoring dosage and that this is communicated to schools.
- Undertake further research into the longer-term benefits of tutoring for pupils, exploring the extent to which tutoring can result in sustainable improvements in outcomes and help close the attainment gap.

References

Andrews, J. (2023) *Recovering from the Covid-19 pandemic: analysis of Star Assessments*. Available at: https://renaissance.widen.net/view/pdf/sz0pq7dkzf/RecoveringFromTheCovid19PandemicMay2023.pdf?t.download=true&u=zceria&mkt_tok=ODI5LVBKRC02NDYAAAGMLuuieJa7680jbqdb1N799rdFUhi_bqK2TfcTDPd-A_nILC8FxMGXelEn7tbwF40uaKJiJxz0vy6Y6QBHtyxTi_TDeOGuaQk4itqplySKRjJxQoid4Q (Accessed: 23 August 2023).

Chappell, S. *et al.* (2011) 'A meta-analysis of Supplemental Educational Services (SES) provider effects on student achievement', *Journal of Education for Students Placed at Risk (JESPAR)*, 16(1), pp. 1–23. Available at: https://doi.org/10.1080/10824669.2011.554140.

DfE (2020) 'How the National Tutoring Programme can help students', *The Education Hub*, 8 December. Available at: https://educationhub.blog.gov.uk/2020/12/08/how-the-national-tutoring-programme-can-help-students/ (Accessed: 23 August 2023).

DfE (2022) 'Academic year 2021/22 key stage 2 attainment'. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment/2021-22#dataBlock-f072b9bb-f8a9-4119-bc67-08da8695004d-tables (Accessed: 23 August 2023).

Dietrichson, J. *et al.* (2017) 'Academic interventions for elementary and middle school students with low socioeconomic status: a systematic review and meta-analysis', *Review of Educational Research*, 87(2), pp. 243–282. Available at: https://doi.org/10.3102/0034654316687036.

EEF (2021a) *One to one tuition: high impact for moderate cost based on moderate evidence*, *EEF*. Available at: https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/one-to-one-tuition)Accessed: 23 August 2023).

EEF (2021b) *Small group tuition: moderate impact for low cost based on moderate evidence*. Available at: https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/small-group-tuition (Accessed: 23 August 2023).

EEF (2021c) *Teaching and learning/early years toolkit guide*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/toolkit/EEF-Toolkit-guide.pdf?v=1669051729 (Accessed: 23 August 2023).

EEF (2022) *The impact of COVID-19 on learning: a review of the evidence*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/guidance-for-teachers/covid-19/Impact_of_Covid_on_Learning.pdf?v=1674657586 (Accessed: 23 August 2023).

Hainmueller, J. (2012) 'Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies', *Political Analysis*, 20(1), pp. 25–46. Available at: https://doi.org/10.1093/pan/mpr025.

Lynch, S. *et al.* (2022) *National Tutoring Programme Year 2: implementation and process evaluation*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1113355/National_Tutoring_Programme_year_2_-_implementation_and_process_evaluation_-_research_report.pdf (Accessed: 23 August 2023).

Lynch, S. *et al.,* (forthcoming) *Evaluation of the National Tutoring Programme Year 3: Implementation and Process Evaluation.*

Major, L.E., Eyles, A. and Machine, S. (2021) *Learning loss since lockdown: variation across the home nations*. Available at: https://cep.lse.ac.uk/pubs/download/cepcovid-19-023.pdf (Accessed: 23 August 2023).

McLean, D., Worth, J. and Faulkner-Ellis, H. (2023) *Teacher labour market in England annual report 2023*. Available at: https://www.nfer.ac.uk/media/5286/teacher_labour_market_in_england_annual_report_2023.pdf (Accessed: 23 August 2023).

Mischo, C. and Haag, L. (2002) 'Expansion and effectiveness of private tutoring', *European Journal of Psychology of Education*, 17(3), pp. 263–73. Available at: https://doi.org/DOI:10.1007/BF03173536.

Moore, E. and Lord, P. (2023) *Tutoring sustainability: understanding the views of school leaders*. Available at: https://www.nfer.ac.uk/media/5395/tutoring_sustainability_understanding_the_views_of_school_leaders.pdf (Accessed: 28 June 2023).

Nickow, A.J., Oreopoulos, P. and Quan, V. (2020) 'The impressive effects of tutoring on PreK-12 learning: a systematic review and meta-analysis of the experimental evidene', *National Bureau Economic Research*, W27476, pp. 267–350. Available at: https://doi.org/10.3386/w27476.

Ofsted (2022) *Independent review of tutoring in schools: phase 1 findings*. Available at: https://www.gov.uk/government/publications/independent-review-of-tutoring-in-schools-and-16-to-19-providers/independent-review-of-tutoring-in-schools-phase-1-findings (Accessed: 23 August 2023).

Pintrich, P.R. and Schunk, D.H. (2002) *Motivation in education: theory research and applications*. London: Pearson.

Poet, H. *et al.* (2022a) *Evaluation of Year 1 of the Tuition Partners Programme: impact evaluation for primary schools*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/projects/Evaluation-Report-Primary-school-impact-evaluation.pdf?v=1666038048 (Accessed: 23 August 2023).

Poet, H. *et al.* (2022b) *Evaluation of year 1 of the Tuition Partners Programme: impact evaluation for year 11*. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/projects/Evaluation-Report-Y11-impact-evaluation-TP.pdf?v=1666084377 (Accessed: 23 August 2023).

Ritter, G.W. *et al.* (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. Available at: https://doi.org/10.3102/0034654308325690.

Staunton, R. *et al.* (2022) *Evaluation of the National Tutoring Programme Year 2 study plan*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1113456/NTP_Year_2_evaluation_Study_Plan.pdf (Accessed: 23 August 2023).

Torgerson, C., J. *et al.* (2018) *Tutor Trust: affordable primary tuition evaluation report and executive summary*. EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/projects/Tutor_Trust.pdf (Accessed: 23 August 2023).

Twist, L., Jones, E. and Treleaven, O. (2022) *The impact of Covid-19 on pupil attainment*. Available at: https://www.nfer.ac.uk/media/4876/the_impact_of_covid_19_on_pupil_attainment.pdf (Accessed: 23 August 2023).

Weidmann, B. and Miratrix, L. (2020) 'Lurking inferential monsters? Quantifying selection bias in evaluations of school programs', *Journal of Policy Analysis and Management*, 40(3), pp. 964–986. Available at: https://doi.org/10.1002/pam.22236.

Appendix A: Results of the Research Champion school analysis

The impact analysis for Research Champion (RC) schools investigated the impact of the NTP at a school level and at a pupil level within schools on standardised English and maths scores for PP and/or PLA pupils.

The approach used in the evaluation of Year 1 of the NTP allowed us to successfully recruit sufficient numbers of schools and use standardised classroom-based assessments to assess the impact of TP in primary schools (Poet *et al.*, 2022a). Based on this approach, we proposed to undertake the RC analysis in the hope of gaining a greater understanding of the impact of tutoring for all pupils who were selected to receive it, not just pupils in Year 6 and Year 11 who undertook national assessments. In addition, we were concerned that the on-going disruption caused by the Covid-19 pandemic might result in national exams being cancelled and replaced with teacher-assessed grades.

As described in Section 2.1.2 of the main report, our sample of Research Champion (RC) schools was smaller than anticipated in the Study Plan (Staunton *et al.*, 2022) meaning that the analyses were underpowered, and therefore uninformative. However, for interest and transparency we have included further details of the analysis approach and the RQ1 results for RC schools here.

A2.1 Additional methods information

A2.1.1 How was a Research Champion (RC) school defined?

All state primary schools were invited to participate in the RC impact analysis. Schools that participated in the NTP[32] formed the intervention group, while those not participating in the NTP formed the comparison group. Primary schools were eligible for inclusion in the impact evaluation as an RC school if they undertook standardised English or maths assessments with pupils in any of Years 1 to 6 in 2021-22[33] (which would be used as baseline and outcome measures) and routinely uploaded this data to the relevant assessment provider's online repository. RC schools were asked to provide pupil-level data for all pupils who sat these assessments in 2021-22. This data included whether or not a pupil received NTP tuition and if so, from which route and in which subject[34]. Only schools with viable data (i.e., enough pupils had both baseline and endpoint test data) were included in the RC impact analysis. Schools that were not eligible to participate as a

[32] Defined as having at least one pupil (for all pupils analysis) or at least one PP and/or PLA pupil (for PP and/or PLA analysis) participating in the relevant tutoring route (AM, TP or SLT) and subject (maths or literacy).
[33] Provided by Renaissance Learning, Rising Stars/Hodder, GL Assessment or NFER. Note that the English assessments used for this analysis included reading but were not exclusively reading assessments, hence we have used the term 'English' here.
[34] This means that unlike our population analysis (see Main Report), here we had information about which subject pupils received tutoring in for SLT.

RC school were invited to participate in the implementation and process evaluation (IPE) only.

A2.1.2 How were intervention and comparison schools defined?

The approach used to define intervention and comparison schools was the same as that used for the population analysis (see Section 2.2 of main report). Intervention schools were defined as having taken part in each intervention if at least one PP and/or PLA pupil at the school participated in the appropriate NTP route and subject in 2021-22. For each intervention, the pool of potential comparison schools was all RC schools not taking part in that intervention. This means that as with the population analysis, the RC analysis aimed to evaluate the additive impact of each NTP route, over and above any other tutoring that may be being delivered in schools. This evaluation considers the impact of NTP as implemented at scale, rather than only evaluating the impact of tutoring when fully implemented as intended (e.g., when pupils received the full 15-hour tutoring course, in the specified tutor/pupil ratios etc). This strategy of comparing across groups that had another NTP route in progress was necessary given the number of schools available, but it does assume that additive impact was possible.

A2.1.3 Pupil inclusion definitions

For RC schools we conducted two types of impact analysis: school-level and pupil-level within-school, this meant that two pupil inclusion definitions were applied to the analysis:

School-level: all PP and/or PLA pupils in both intervention[35] and comparison schools were included in this analysis. This approach compared the progress made by PP and/or PLA pupils who attended schools that participated in the relevant NTP route and subject with PP and/or PLA pupils who attended comparison schools. Therefore, this analysis investigated the impact of the school's choice to participate in a specific route of the NTP, rather than which pupils were selected to receive tutoring.

Pupil-level within-school: the NTP is a pupil level intervention, therefore it should ideally be analysed at a pupil level and the within school analysis is one way of doing this. The within school analysis only included intervention schools. The intervention group consisted of PP and/or PLA pupils who attended RC schools that participated in the NTP and who were themselves selected for tutoring in the relevant route and subject. The comparison group was made up of PP and/or PLA pupils who attended the same schools but were not selected for NTP tuition. A difference-in-differences approach was used for the within-school analysis. For this approach each pupil was treated as having two outcomes: one measured prior to the NTP intervention (this baseline measurement is instead included as a covariate in the school-level models) and one measured after it. An interaction term between time period (pre- or post-intervention) and intervention status is

[35] Defined as having at least one PP and/or PLA pupil participating in the relevant tutoring route (AM/TP or SLT) and subject (literacy or numeracy).

interpreted as the differential progress made by intervention pupils relative to comparison pupils.

A2.1.4 Assessing the impact of the NTP on English and maths attainment

For RC schools the impact of the TP/AM and SLT routes was assessed separately and considered the subject in which tuition was received. We therefore investigated the impact in RC schools of four different interventions (SLT literacy, SLT numeracy, AM/TP literacy and AM/TP numeracy), at a school level and pupil level within schools.

This meant that for both the school level and pupil level within school analyses, four linear mixed effects regression models were estimated (including weights as described in main report) to analyse the impact of:

- SLT (numeracy) on standardised maths outcomes
- SLT (literacy) on standardised English outcomes
- TP/AM (numeracy) on standardised maths outcomes
- TP/AM (literacy) on standardised English outcomes.

To reduce dilution within intervention schools for the school-level analyses, year groups in intervention schools with no PP and/or PLA pupils receiving NTP tuition were removed prior to regression modelling[36]. As intervention and comparison schools were not matched on a one-to-one basis, corresponding year groups in comparison schools were not removed unless there were no intervention pupils in that year group for any intervention schools. However, although year groups in intervention schools with no pupils participating in the relevant route and intervention were removed from the analysis, the results are still subject to dilution as not all PP and/or PLA pupils in each year group which received tutoring were themselves selected for the NTP.

In addition, both intervention and comparison schools (and potentially pupils) may also have taken part in other NTP routes. Variables indicating participation in NTP routes other than the route relevant to the specific intervention model were therefore included in the statistical matching and as regression covariates. This approach aimed to ensure that the difference between the intervention and comparison groups was purely the intervention route of interest. This meant the intervention group and the weighted comparison group were equal in the amount of tuition they received in the alternative routes at a school level. The dosage and concentration of other NTP routes was not well-matched between the groups (i.e., the amount of tuition from alternative routes was not balanced at a pupil level). However, pupil level participation in 'other' NTP routes apart from the route of interest was included as a covariate in the regression models to help

[36] This means that a RC intervention school had no pupils in Year 2 receiving NTP tuition in the relevant route and subject, then Year 2 pupils for that school were removed from the regression model.

account for this. We therefore implicitly assume in our interpretation that, since there was no difference in the amount of tuition through alternative routes between the intervention group and the weighted comparison group, the difference between the two groups isolates the difference due to the intervention route of interest (please refer to the technical appendix for more information about the matching and the degree of balance achieved between the groups), and tells us about the additive impact of that specific NTP route.

A2.2 How many Research Champion (RC) schools were included in the impact evaluation?

As defined above, RC schools were mainstream primary schools in England willing to take part in the research that were already using standardised assessments and routinely uploading this data to the relevant assessment provider's online repository. RC schools also had to have at least one year group following the required baseline and endpoint assessment schedule[37] in at least one subject (either English or maths). All mainstream state primary schools in England were approached during recruitment and this included schools that had participated as RC schools in the evaluation of Year 1 of the NTP evaluation (Poet *et al.*, 2022a). Eligible RC schools were then asked to provide pupil-level data for all pupils who sat the relevant assessments in 2021-22.

Recruiting RC schools was much more challenging for the Year 2 evaluation of the NTP than it was for the Year 1 evaluation of the NTP. There were several reasons for this. Firstly, in Year 1 we were provided with a key contact for schools directly from the relevant tuition partner, whereas in Year 2, we contacted all primary schools and asked them if they were participating in the NTP. This meant we only had a generic email address for the school, rather than a key contact. Secondly, in the Year 1 evaluation we were able to offer schools an incentive to participate, whereas this was not the case in Year 2. Thirdly, in the Year 2 evaluation we only recruited schools who were uploading their assessment data to the relevant assessment provider's online repository. We also found that sign-up to the NTP was slower than expected, and that the total number of schools signing up to the AM and TP routes especially was lower than expected due to the addition of the SLT route. However, the main difficulty we encountered was that the schools which were willing to sign-up did not have the firm plans in place to conduct the assessments necessary to be involved in the evaluation. Around 540 schools responded positively to our invitation to participate, but 65% of these schools were not eligible to become RC schools.

Further attrition to the number of recruited RC schools occurred due to (i) 48 schools withdrawing from the evaluation and (ii) 32 schools not providing the necessary pupil

[37] For Years 1 to 6 the baseline standardised assessment had to take place in the autumn term or the previous academic year's summer term. The endpoint assessment had to take place in the summer term for Years 1 to 5. For Year 6 an endpoint assessment was not required as KS2 outcome scores could be used instead.

data to be included in the analysis. At this stage we were aware that the number of recruited schools was going to be smaller than the sample size of 106 schools in each intervention and comparison group, which was deemed necessary in the Study Plan (Staunton *et al.*, 2022), for a Minimum Detectable Effect Size (MDES) of 0.13 SD for this intervention. Consequently, we updated the MDES calculations based on the actual numbers of schools recruited (Table 21**,** rows 2 and 3). These updated calculations were discussed with DfE in April 2022, and it was agreed that we would continue with the analysis, despite the smaller number of schools.

Table 21: Updated MDES calculations at the end of recruitment and at the analysis stage of the project.

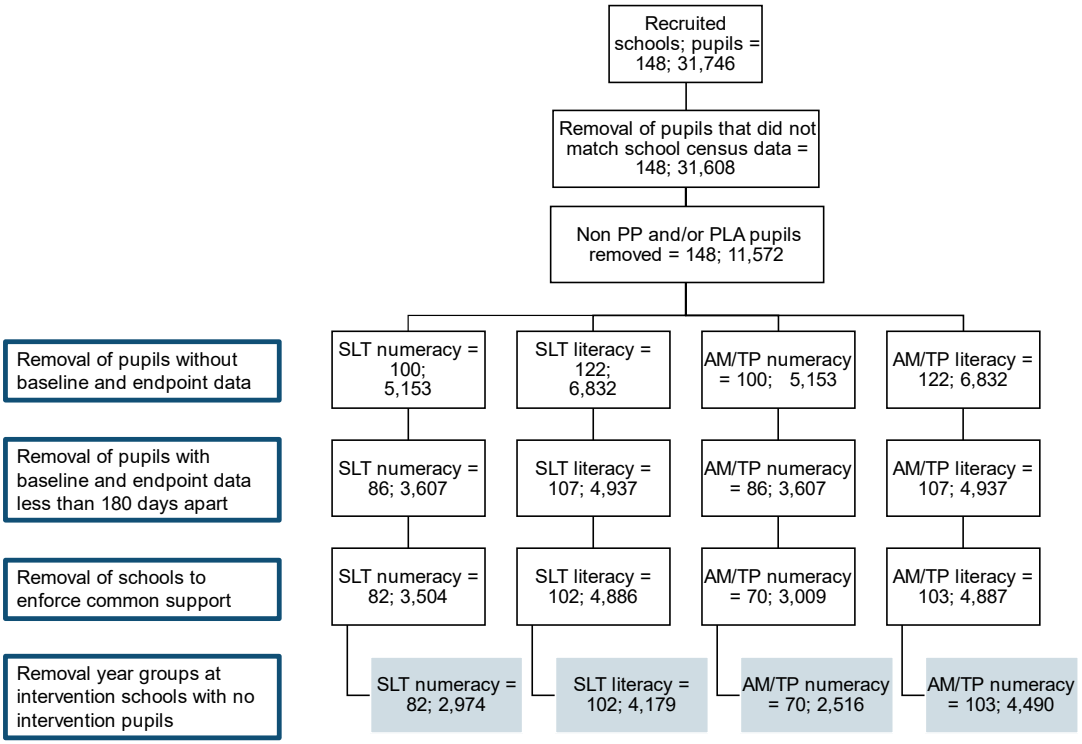
Stage	Subject	Number of schools (total)	Allowing for 20% dropout at data cleaning	Average number of PP/PLA pupils per school	ICC	Pre-post correlation	MDES
Proposal/ Study Plan	Both	212	Y	24	0.15	0.7	0.13
End of Recruitment	Literacy	140	Y	24	0.15	0.7	0.16
	Numeracy	119	Y	24	0.15	0.7	0.18
Analysis	Literacy	103	N	44	0.12	0.77	0.14
	Numeracy	70	N	36	0.14	0.76	0.18

These MDES relate to the TP/AM school-level RQ1 analysis, as fewer schools participated in these routes.

Following recruitment, data cleaning and quality checks took place and only RC schools with viable data (see Section A2.1.1) were included in each analysis. This resulted in further data attrition for the following reasons (also see Table 21**)**. Firstly, the analysis started with 11,572 PP and/or PLA pupils (out of a total of 31,608 pupils overall). However, of these PP and/or PLA pupils only 3,607 pupils eligible for inclusion in the maths analysis and 4,937 pupils eligible for inclusion in the literacy analysis had the baseline and endpoint data required. Pupils were not included in the analysis if their baseline and endpoint data was collected less than 180 days apart. This was to ensure that pupils had sufficient time between baseline and endpoint assessment to receive NTP tuition. As shown in Figure 14, the remaining data attrition occurred due to: (i) schools being removed at the matching stage due to what are called 'common support'

restrictions, and (ii) enforcement of the pupil inclusion definitions in Table 4 of the Study Plan (Staunton *et al.*, 2022), described in Section A2.1.3.

Figure 14: Data attrition for a) schools and b) pupils during enforcement of the analysis criteria.



This meant that the number of intervention schools included in the school-level analyses varied from 27 (AM and/or TP literacy) to 46 (SLT literacy) and the number of comparison schools varied from 43 (AM and/or TP numeracy) to 76 (AM and/or TP literacy) (see Table 22). The number of schools included in the pupil-level within-school analysis was similarly variable (see Table 23), ranging from 27 schools (AM/TP literacy) to 46 schools (SLT literacy). As these numbers were smaller than anticipated, we updated the MDES calculations again based on the data we were able to analyse (see Table 21, rows 4 and 5). Compared with the MDES outlined in the Study Plan (Staunton *et al.*, 2022), we had a much smaller sample of schools than anticipated, but a larger average number of PP and/or PLA pupils per school, lower intra-cluster correlation (ICC), and higher pre-post correlation helped to balance out the reduction in school numbers. Although the analysis is not as underpowered as we might have anticipated given the smaller number of schools, with an MDES of 0.14 for literacy and 0.18 for numeracy, the dilution issue (see below) still means it was underpowered.

Table 22: Number of RC schools and PP and/or PLA pupils included in the school-level comparisons by tutoring route and subject.

	Number of intervention schools in analysis	Number of comparison schools in analysis	Number of intervention PP and/or PLA pupils in analysis	Number of comparison PP and/or PLA pupils in analysis	Percentage of PP and/or PLA pupils in intervention schools who received tutoring
SLT numeracy	34	48	1098	1876	30%
SLT literacy	46	56	1871	2308	33%
AM/TP numeracy	27	43	1004	1512	41%
AM/TP literacy	27	76	1499	2991	44%

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from RC schools.

It is important to acknowledge that any effect of the NTP on individual pupils is diluted in the school-level comparisons. As shown in Table 22 only 30% of PP and/or PLA pupils in schools involved in SLT numeracy and 33% of PP and/or PLA pupils in schools involved in SLT literacy were selected to receive tutoring. Similarly, 41% of PP and/or PLA pupils in schools involved in AM/TP numeracy and 44% of PP and/or PLA pupils in schools involved in AM/TP literacy were selected to receive tutoring. The above MDES calculations have to be interpreted in this context, as it means that to achieve a school level MDES of 0.18, with 30% dilution we would need to see an effect size of around 0.6 amongst tutored pupils. This effect size is larger than we might expect to see based on previous evidence about the effectiveness of small group tutoring (EEF, 2021b) and given the NTP is being implemented at scale, quite unrealistic.

This issue with dilution does not apply to the pupil-level within-school analysis. This is because although the proportion of pupils were selected for tutoring was the same, the within-school analysis compared PP and/or PLA pupils who were selected for tutoring with all PP and/or PLA pupils in the same school who were not selected for tutoring. Therefore, pupils who did not receive tutoring are not included in the intervention group. However, the pupil-level within-school analysis is subject to the problem of selection bias as schools chose which pupils received tutoring. In this case, we anticipated that

teachers would select pupils who were falling behind to receive tutoring, resulting in a potential negative bias. If this selection is based on variables that are not available to the evaluation (e.g. recent performance in class tests of English or Maths) then it will not be possible to identify an appropriate comparison group to control for this negative bias.

Table 23: Number of schools and pupils included in the pupil-level within-school analyses.

	Number of schools	Number of intervention pupils	Number of comparison pupils
SLT numeracy	34	330	768
SLT literacy	46	609	1262
AM/TP numeracy	27	409	595
AM/TP literacy	27	663	836

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools.

A2.3 What was the impact of the NTP on English outcomes among RC schools?

A2.3.1 School-Led Tutoring

The impact analysis exploring the association between SLT literacy tuition and English outcomes did not detect any statistically significant differences between PP and/or PLA pupils whose schools chose to participate in SLT literacy and PP and/or PLA pupils whose schools did not participate in SLT (p = 0.952; Hedges' g = 0.002 (-0.078, 0.083)) (Table 24).

Similarly, within RC schools, we did not detect a statistically significant difference in progress between PP and/or PLA pupils who were or were not selected for SLT literacy tuition (p = 0.508; Hedges' g = -0.040 (-0.160, 0.079)).

Table 24: Results of the linear mixed effects models exploring the impact of SLT literacy tuition on English outcomes

	N pupils (Intervention; Comparison)	Additional standard score points	Additional months progress[38]	Hedges' *g* (95% CI)	*p*-value
School-level English score	1871; 2308	0.038	0	0.002 (-0.078, 0.083)	0.952
Within-school English Score	609; 1262	-0.628	0	-0.040 (-0.160, 0.079)	0.508

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

A2.3.2 Academic Mentors/Tuition Partners

Results of the impact analysis comparing English outcomes for PP and/or PLA pupils in RC schools who chose to participate in AM/TP literacy with PP and/or PLA pupils attending comparison schools did not detect a statistically significant difference in pupils' English outcomes between the groups (p = 0.105; Hedges' g = -0.078 (-0.171, 0.015)).

Likewise, analysis comparing the outcomes of PP and/or PLA pupils who were selected for AM/TP literacy tuition with PP and/or PLA pupils within the same school who did not participate in the intervention did not detect a statistically significant difference in progress between the groups (p = 0.113; Hedges' g = 0.102 (-0.024, 0.229)) (Table 25). We acknowledge that the point estimate for the within-school analysis indicated that on average PP and/or PLA pupils selected for tutoring made the equivalent of an additional two months' progress in English (EEF, 2021c) compared with PP and/or PLA pupils who attended the same schools but were not selected for tutoring. This may indicate that AM/TP tuition had a positive impact on English outcomes for PP and/or PLA pupils, but this needs to be considered in the context of wide confidence intervals and that the difference between the groups was not statistically significant, so may not reflect meaningful change.

[38] Defined according to EEF (2023).

Table 25: Results of the linear mixed effects models exploring the impact of AM and/or TP literacy tuition on English outcomes.

	N pupils (Intervention; Comparison)	Additional standard score points	Additional Months Progress	Hedges' *g* (95% CI)	*p*-value
School-level English Score	1499; 2291	-1.190	-1	-0.078 (-0.171, 0.015)	0.105
Within-school English Score	663; 836	1.558	+2	0.102 (-0.024, 0.229)	0.113

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

A2.4 What was the impact of the NTP on maths outcomes among RC schools?

A2.4.1 School-Led Tutoring

The impact analysis exploring the effect of SLT numeracy tuition on maths outcomes did not detect any significant differences in maths outcomes between PP and/or PLA pupils who attended schools that chose to participate in SLT numeracy and PP and/or PLA pupils who attended schools that did not participate in SLT numeracy (p = 0.759; Hedges' g = 0.022 (-0.115, 0.158)) (Table 26).

Similarly, within RC schools, we did not detect a significant difference in progress between PP and/or PLA pupils who were selected for SLT numeracy tuition and PP and/or PLA pupils at the same schools who were not selected for SLT (p = 0.308; Hedges' g = 0.082 (-0.076, 0.240)). These results therefore offer no evidence that SLT was associated with better maths outcomes for PP and/or PLA pupils. We note that the difference between the groups did equate to an additional month's progress (EEF, 2021c). However, given this result was not statistically significant and was only equivalent to a change of approximately one standard score point, it is unlikely to reflect a meaningful progress in maths.

Table 26: Results of the linear mixed effects models exploring the impact of SLT numeracy tuition on maths outcomes

	N pupils (Intervention; Comparison)	Additional Standard Score Points	Additional Months Progress	Hedges' *g* (95% CI)	*p*-value
School-level Maths Score	1098; 1876	0.331	0	0.022 (-0.115, 0.158)	0.759
Within schools Maths Score	330; 768	1.255	+1	0.082 (-0.076, 0.240)	0.308

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

2.4.2 Academic Mentors/Tuition Partners

Results of the impact analysis comparing maths outcomes for PP and/or PLA pupils in RC schools who chose to participate in AM/TP numeracy with PP and/or PLA pupils in comparison schools did not detect a significant difference in pupils' maths outcomes (p = 0.293; Hedges' g = 0.073 (-0.061, 0.206)) (Table 27). Although the difference between the groups was not statistically significant, it did equate to an additional month of progress (EEF, 2021c). However, this difference only equated to an additional scaled score point on the standardised maths assessments, so is unlikely to reflect a meaningful change.

Results of the pupil-level within school analysis also indicated that PP and/or PLA pupils who were selected for AM/TP numeracy tuition did not make more progress in maths outcomes compared to comparison pupils (p = 0.834; Hedges' g = 0.016 (-0.135, 0.167)).

Table 27: Results of the linear mixed effects models exploring the impact of AM/TP numeracy tuition on maths outcomes

	N pupils (Intervention; Comparison)	Additional standard score points	Additional Months Progress	Hedges' *g* (95% CI)	*p*-value
School-level maths score	1004; 1512	1.111	+1	0.073 (-0.061, 0.206)	0.293
Within school maths score	409; 595	0.244	0	0.016 (-0.135, 0.167)	0.834

Source: National Pupil Database 2021/2022 (School Census), tuition participation data collected directly from schools, assessment provider data.

Reference: RR1370

ISBN: 978-1-83870-500-8

For any enquiries regarding this publication, contact us at www.education.gov.uk/contactus

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