



Department
for Education

National Tutoring Programme Year 3

Impact Report

September 2024

**Authors: Emma Moore, Chris Morton,
Gemma Schwendel and Stephen
Welbourne**



Government
Social Research

Acknowledgements

The authors would like to thank Rushda Khandker, Benjamin Cordwell, 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 for their comments on the draft report and to Vrinder Atwal for her administrative support.

This research report was written before the new UK Government took office on 5 July 2024. As a result, the content may not reflect current Government policy.

Contents

List of figures	5
List of tables	7
Executive Summary	9
Introduction	9
Evaluation aims	9
Limitations	11
Conclusions	12
Introduction	15
The National Tutoring Programme (NTP) and the socioeconomic attainment gap	16
Impact evaluation aims	17
Research questions	18
Methods	19
Participant sample	19
Outcome measures	19
Data sources	19
Defining intervention and comparison groups at school and pupil level	20
Statistical analysis	23
Limitations	29
Sample characteristics	32
School-level evaluation samples	32
Pupil-level evaluation samples	39
Main results maths: What was the impact of the NTP on maths outcomes in 2022-23?	42
What was the impact of the NTP on maths outcomes for all pupils, PP pupils and PLA pupils?	43
How does the impact of the NTP on maths outcomes vary according to tutoring dosage?	47
How does the impact of the NTP on maths outcomes vary according to geographic region?	49
Main results English: What was the impact of the NTP on English outcomes in 2022-23?	53
What was the impact of the NTP on English outcomes for all pupils, PP pupils and PLA pupils?	53

How does the impact of the NTP on English outcomes vary according to tutoring dosage?	56
How does the impact of the NTP on English outcomes vary according to geographic region?	58
Results by pupil characteristics: How does the school-level impact of the NTP on maths and English outcomes vary according to pupil characteristics?	62
How does the school-level impact of the NTP on maths outcomes vary according to pupil characteristics?	62
How does the school-level impact of the NTP on English outcomes vary according to pupil characteristics?	70
Long term impact: What is the longer-term impact of the NTP delivered in 2021-22 on maths and English attainment outcomes for pupils in Year 6 and Year 11 in 2022-23?	79
Discussion	81
Recommendations	85
References	86
Appendix A: complete results from the long-term impact analysis	89

List of figures

Figure 1: Intervention and comparison groups for the main pupil-level analysis.....	22
Figure 2: Intervention and comparison groups for the pupil-level sensitivity analysis	23
Figure 3: Proportion of Year 6 pupils receiving tutoring and average tutoring hours per NTP tutored pupil (ATH-NTP) amongst Year 6 pupils at schools taking part in the NTP. Cell counts indicate number of schools.	37
Figure 4: Proportion of Year 11 pupils receiving tutoring and average tutoring hours per NTP tutored pupil (ATH-NTP) amongst Year 11 pupils at schools taking part in the NTP. Cell counts indicate number of schools.	38
Figure 5: The impact of the NTP on KS2 and KS4 maths outcomes.	44
Figure 6: Pupil level sensitivity analysis for KS2 and KS4 maths outcomes.	47
Figure 7: Pupil KS2 maths scores determined by average tutoring hours delivered at their school amongst NTP pupils (ATH-NTP).	49
Figure 8: The impact of the NTP on maths outcomes at school level by region.	50
Figure 9: The impact of the NTP on English outcomes at KS2 and KS4.	54
Figure 10: Pupil-level sensitivity analysis for KS2 and KS4 English outcomes.	56
Figure 11: Pupil KS2 reading scores determined by average tutoring hours delivered at their school amongst NTP pupils (ATH-NTP).	57
Figure 12: The impact of the NTP on English outcomes at school level by region.	59
Figure 13: The impact of the NTP on maths outcomes at school level by pupil characteristics.....	63
Figure 14: The school-level impact of the NTP on maths outcomes for pupils from different ethnic backgrounds.....	65
Figure 15: The school-level impact of the NTP on maths outcomes for pupils from areas with different levels of deprivation.....	68
Figure 16: The impact of the NTP on English outcomes at school level by pupil characteristics.....	71
Figure 17: The school-level impact of the NTP on English outcomes for pupils from different ethnic backgrounds.....	74

Figure 18: The school-level impact of the NTP on English outcomes for pupils from areas with different levels of deprivation.....77

Figure 19: The impact of participation in the NTP in 2021-22 on English and maths outcomes at KS2 and KS4 in 2022-23.....80

List of tables

Table 1: Percentage of pupils selected for tutoring in maths or English in Year 2 of the NTP	30
Table 2: Number and percentage of Year 6 pupils included in the KS2 school-level evaluation sample split by pupil characteristics and NTP selection	33
Table 3: Number and percentage of Year 11 pupils included in the KS4 school-level evaluation sample split by pupil characteristics and NTP selection	34
Table 4: Number and percentage of pupils in the pupil-level evaluation sample that received NTP tutoring in 2021-22 split by pupil characteristics and 2022-23 NTP selection.....	39
Table 5: The impact of the NTP on KS2 and KS4 maths outcomes; complete results from the linear mixed effects models	45
Table 6: Summary of which pupils are in the intervention and comparison groups for the pupil-level main and sensitivity analysis.....	46
Table 7: The impact of the NTP on maths outcomes at school level by region; complete results from the linear mixed effects models.....	51
Table 8: The impact of the NTP on English outcomes at KS2 and KS4; complete results from the linear mixed effects models	55
Table 9: The impact of the NTP on English outcomes at school level by region; complete results from the linear mixed effects models.....	60
Table 10: The impact of the NTP on maths outcomes at school level by pupil characteristics; complete results from the linear mixed effects models	64
Table 11: The school-level impact of the NTP on maths outcomes for pupils from different ethnic backgrounds; complete results from the linear mixed effects models	66
Table 12: The school-level impact of the NTP on maths outcomes for pupils from areas with different levels of deprivation; complete results from the linear mixed effects models	69
Table 13: The impact of the NTP on English outcomes at school level by pupil characteristics; complete results from the linear mixed effects models	72
Table 14: The school-level impact of the NTP on English outcomes for pupils from different ethnic backgrounds; complete results from the linear mixed effects models	75

Table 15: The school-level impact of the NTP on English outcomes for pupils from areas with different levels of deprivation; complete results from the linear mixed effects models78

Table 16: The impact of participation in the NTP in 2021-22 on English and Maths outcomes at KS2 and KS4 in 2022-23; complete results from the linear mixed effects models89

Executive Summary

Introduction

The closure of schools to most pupils during the Covid-19 pandemic caused significant disruption to pupils' education and learning, with evidence suggesting that socioeconomically disadvantaged pupils were most severely affected (Major, Eyles and Machine, 2021). The National Tutoring Programme (NTP) was introduced in the 2020-21 academic year and aimed to support socioeconomically disadvantaged pupils to recover lost learning and help close the attainment gap for these pupils. With the immediate aftermath of the pandemic over, the focus of the NTP in its third year shifted towards tackling the persistent socioeconomic attainment gap. As with the second year of the NTP there were three possible routes of support schools could use:

Academic mentors (AM): academic mentors are employed by the school. Initially this route was only available to schools with the highest numbers of socioeconomically disadvantaged pupils, but from the third year of the NTP was an option for all schools to use.

Tuition partners (TP): this route offered subsidised tuition to schools from approved tuition partners.

School-led tutoring (SLT): this route was introduced in the academic year 2021-22 and allowed schools to source local tutoring provision.

In the third year of the NTP all three routes were open to all schools and the Department for Education (DfE) subsidised 60% of the costs for all routes. This meant schools were able to choose which routes or combination of routes to use to best support their pupils.

Evaluation aims

Primary aim: to evaluate the impact of the third year of the NTP on English and maths outcomes for a) all pupils, b) pupils eligible for pupil premium (PP pupils) and c) pupils with prior low attainment (PLA pupils) at Key Stage 1 (KS1) or Key Stage 2 (KS2).

In addition, this evaluation also sought to understand:

- whether the impact of the NTP varied according to the number of tutoring hours pupils received (dosage).
- whether there were differences in the impact of the NTP according to pupil characteristics (gender, Special Educational Needs and Disabilities (SEND) status, English as an Additional Language (EAL), ethnicity) or school characteristics (region, Income Deprivation Affecting Children Index (IDACI)).
- the longer-term impact of receiving the NTP in the academic year 2021-22.

To achieve these aims, we evaluated the impact of the NTP (any route and subject) on pupils' English and maths outcomes. This means that we compared the impact of participating in the NTP with non-participation in the NTP. We did not split the comparison by tutoring route or tutored subject because data regarding these was not available in the third year. The one exception to this was the longer-term analysis. For the longer-term analysis, information about NTP participation was based on data from the second year of the NTP. This meant that route data was available, and tutored subject information was also available for the AM and TP routes. Therefore, our comparisons for the longer-term analysis took account of route and, where available, subject.

Key Findings

KS2 Main Impact Findings:

- We found a consistent pattern of evidence at both school and pupil level to suggest that participation in the NTP was associated with small improvements in English and maths outcomes at KS2.
- The improvements seen in maths were larger than those seen in English.
- Although many of these results reached statistical significance, the effect sizes were small and equated to one months' additional progress or less.
- Due to dilution and negative bias, we expect the true impact of the NTP at KS2 to be greater than these results suggest.

KS4 Main Impact Findings:

- There was more limited evidence at both school and pupil level to indicate that the NTP may be associated with very small improvements in both English and maths outcomes at KS4.
- The effect sizes seen at KS4 were consistently smaller than those seen at KS2 and in all cases equated to less than one month's additional progress, although some did still reach statistical significance.
- Due to dilution and negative bias, we expect the true impact of the NTP at KS4 to be greater than these results suggest.

Once dilution effects and negative selection bias are considered, the effect sizes may be consistent with the expected impact of tutoring based on previous research.

Additional Findings:

The impact of the NTP was broadly similar across pupils and schools with different characteristics. We found some evidence which may indicate that PLA pupils could benefit more from tutoring than all pupils and PP pupils in both English and maths at both KS2 and KS4. However, the confidence intervals for the effect sizes for each of these groups frequently overlapped, indicating that any differences in benefit are very small.

Our dosage analysis sought to understand how best to divide a fixed number of tutoring hours between pupils to achieve the best outcomes at a school level. At KS2, our best estimate for the number of tutoring hours per tutored pupil that maximised school level outcomes across all pupils was 31 for maths and 20 for reading. However, there is a large amount of uncertainty around these estimates, as we did not have tutoring subject information. To arrive at our estimates, we assumed the proportion of tutoring hours in the relevant subject, based on the split of tutoring provided via the AM/TP routes in year 2.

We were unable to detect the best estimate for the number of tutoring hours per tutored pupil that would maximise school level outcomes at KS4. This may be because there are fewer secondary schools meaning the KS4 analysis had lower statistical power.

We did detect a small longer-term benefit of having participated in the SLT/AM routes on KS2 maths outcomes, but this equated to less than one month's additional progress. We did not detect any other longer-term benefits of participation in SLT/AM on English outcomes for KS2 pupils or on English or maths outcomes for KS4 pupils. However, we did find that the small negative effects we found on English outcomes at KS2 and KS4 for the TP route in the impact evaluation of the second year of the NTP, persisted longer-term. This effect size was small, so as with the second year of the NTP we cannot exclude the possibility that these results reflect selection bias. We did not detect any longer-term effects of having received tuition via the TP route for maths.

Limitations

Randomisation of pupils was not possible for the NTP, which means that any evaluation of its impact will have several limitations. The key limitations are described below (and in more detail in the main report) and provide important context in which to consider the interpretation of the evaluation findings reported here.

Dilution: at the school-level there are two potential sources of dilution. Firstly, not all pupils in intervention schools themselves received tutoring. A second source of dilution comes from the lack of information about the subject pupils received tutoring in. The

latter source of dilution is also an issue at pupil level. These dilution effects mean that the effect sizes estimates detected in our analysis are likely to underestimate the true effect size.

Selection bias: schools chose which pupils received tutoring. This means that we would expect them to select pupils for tutoring who are most in need of additional support e.g., they have fallen furthest behind their peers and/or are not meeting age-related expectations. This is consistent with the 23/24 NTP guidance which states that in addition to PP pupils, support can be targeted at pupils performing below the expected standard (GOV.UK, 2024). However, from an analysis perspective, this means that we would anticipate a high level of negative selection bias amongst tutored pupils, especially in the pupil level analysis, as without support these pupils would be expected to perform worse than their peers. It is also possible that there is selection bias present in the school-level analysis but based on the work by (Weidmann and Miratrix, 2020), which aimed to quantify the magnitude of selection bias within non-randomised school-level analyses, we consider this less of a risk.

Conclusions

Overall, we found consistent evidence across all three subgroups of pupils that the NTP was effective for improving maths and English attainment at KS2 and some more limited evidence at KS4. All improvements seen were small. This is consistent with findings from previous research indicating that tutoring tends to be more effective for younger pupils; especially for English (Nickow, Oreopoulos and Quan, 2020; EEF, 2021a, 2021b). However, unlike previous research we found larger effect sizes for maths than reading for KS2 pupils (see Nickow, Oreopoulos and Quan, 2020).

The difference in tutoring effectiveness we detected between primary and secondary school pupils may be because tutoring is more able to help fill learning gaps for primary school pupils, whereas for older pupils, there may be more significant gaps in their learning because they have accumulated over a longer period, especially for disadvantaged pupils, thus making it harder to address them through catch-up interventions such as tutoring sessions (E-ACT, No Date). Alternatively, it is possible that differences in tutoring delivery (e.g., dosage, session timing, session frequency, curriculum alignment delivery mode – online or in person) between primary and secondary schools could account for differences in tutoring effectiveness.

Our results are broadly consistent with the results we saw for SLT in the impact evaluation for year 2 of the NTP in terms of both the direction and size of the effects (which equated to one months' additional progress of less). The consistency of this evidence between these two impact evaluations is somewhat promising, especially for KS2 pupils, and builds on previous research indicating that small group tutoring is effective (e.g., EEF, 2021b; EEF, 2021a). At KS4 and for KS2 English the small effect sizes and lack of consistent statistically significant results means that we cannot

completely exclude the possibility that these effects may be artefacts of selection bias despite our considerable attempts to remove it, rather than genuine effects of the NTP on pupil outcomes. However, for KS2 maths finding statistically significant effects for nearly all subgroups, each of which has a separate match, coupled with the statistically significant longer-term effects of KS2 maths tutoring and the consistency with the year 2 evaluation results mean that we can be more confident that this is a real effect and not an artifact of school-level selection bias.

We note that these effect sizes are still smaller than we might expect based on evidence about the effectiveness of small group tuition (Ritter et al., 2009; Dietrichson et al., 2017; Nickow, Oreopoulos and Quan, 2020; EEF, 2021a). This difference likely to be due to factors related to the evaluation design (e.g., specific target population, ability to conduct pre- and post-tests) and tutoring implementation. In addition, dilution remains an issue within the present analysis (see limitations). This means, as with the year 2 impact evaluation, the effect sizes reported here are likely to underestimate the true effect.

A key challenge for this and previous evaluations of the NTP, has been developing an appropriate comparison group for the pupil-level analysis as schools choose which pupils received tutoring. They (rightly) are likely to have selected pupils most in need of support, resulting in negative selection bias (i.e., without tutoring these pupils would be expected to have lower attainment scores than their peers). To reduce selection bias within the pupil-level analysis, we restricted the analysis to pupils who were selected for tutoring in the second year of the NTP and found small positive effects. However, the results of our sensitivity analysis provided a strong indication that selection bias is still present in the pupil-level analysis and therefore more strongly indicate the estimates are likely to underestimate the true effect sizes.

It is also likely that tutoring implementation factors such as differences in tutoring dosage could account for some of the differences in outcomes between the NTP impact evaluations and previous studies of tutoring. The dosage analysis explored how best to divide a fixed number of tutoring hours between pupils to achieve the optimum impact at school level. The results suggest that concentrating a higher number of tutoring hours on a smaller number of pupils is likely to be most beneficial overall, even though this means that some pupils will not receive tutoring. This suggests that headteachers should focus on identifying pupils that would benefit from longer-term support when considering how best to spend money on tutoring. However, because of the lack of subject information, more research would be needed to fully understand the optimum tutoring dosage which maximises the benefits for pupils and schools. In addition, we still do not yet have a full understanding of how to optimise other aspects of tutoring such as session duration, mode of delivery (online vs in person), aligning sessions with the school curriculum, and time of delivery, to get the best results for schools as pupils.

The shift in focus of the NTP to reducing the persistent attainment gap for socioeconomically disadvantaged pupils in year 3 means that understanding any longer-term benefit of tutoring is key. To close the attainment gap, progress made via tutoring

would need to be sustained over time. The analysis only detected a longer-term benefit of receiving SLT/AM in year 2 on maths outcomes for KS2 pupils (though we note no subject information was available for SLT). We have been unable to find evidence to suggest other longer-term positive effects, despite the positive effects seen for SLT in the year 2 impact evaluation.

In addition, the proportion of socioeconomically disadvantaged pupils (i.e., PP pupils) selected to receive tutoring this year was smaller than last year (42% compared with 46% in year 6 and 31% compared with 35% in year 11) (see Sample Characteristics and Lucas et al., 2023). This makes it more challenging to close the attainment gap for these pupils as they are not necessarily receiving the support. If the goal is to close the attainment gap for socioeconomically disadvantaged pupils, then support needs to be targeted at them.

Recommendations

An objective of the NTP is to move towards a model of embedding tutoring within schools. With this in mind, we make the following recommendations:

- Senior leaders should consider which pupils are likely to benefit from regular tutoring and concentrate the tutoring hours they are able to offer on a smaller number of pupils. An optimum number of hours for greatest impact within a limited budget is likely to lie above 20 hours per pupil, closer to the EEF recommendation of 30 hours of tuition.
- DfE should seek to further build the evidence base around how to optimise the delivery of tutoring to maximise impact. This includes understanding the 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).
- If there is going to be continued funding for tutoring, consideration should be given to focussing it on maths and English in primary school. Evidence for the impact of primary maths tuition was strong in this evaluation (including a sustained impact), while previous evidence suggests strong benefits of English tuition in this age group.
- DfE should consider alternative approaches to closing the socioeconomic attainment gap. There is evidence that tutoring has had immediate benefits for pupils that participated, and that this may be slightly higher for PLA and PP pupils. However, it hasn't been effective as a mechanism for closing the socioeconomic performance gap because disadvantaged pupils have not been selectively targeted, and there is only limited evidence of sustained benefit (in KS2 maths).

Introduction

The Covid-19 pandemic led the UK Government to place restrictions on in-school attendance for most pupils over three periods: March – May 2020; June – July 2020; and January – March 2021¹. These restrictions caused significant disruption to pupils' education and learning and negatively impacted pupils' attainment (EEF, 2022). This was particularly the case for socioeconomically disadvantaged pupils for whom it was more difficult to keep up with learning during school closures compared with their more affluent peers (Major, Eyles and Machine, 2021). These pupils were also more likely to be withdrawn from school before closures and less likely to return immediately when schools were allowed to reopen (Nelson and Sharp, 2020). Additionally, they were less likely to return set work possibly because they typically had less contact with their teachers, poorer curriculum coverage and lower levels of parental support and IT access than their peers (Nelson and Sharp, 2020). This resulted in the socioeconomic attainment gap² widening during the pandemic (DfE, 2022; EEF, 2022; Twist, Jones and Treleaven, 2022; Andrews, 2023).

The National Tutoring Programme (NTP) was introduced in the academic year 2020-21 as a key part of the UK Government's Covid-19 recovery strategy. It focused on providing targeted support for socioeconomically disadvantaged pupils, as these pupils were most affected by missed learning during the Covid-19 pandemic. The original aim of the NTP was to help these pupils recover from missed learning during the Covid-19 pandemic and reduce the widening socioeconomic attainment gap (i.e., the difference in attainment between these pupils and their more affluent peers). It also aimed to establish and embed tutoring as an effective tool schools could use to help close the socioeconomic attainment gap by encouraging schools to use their pupil premium funding for tutoring.

This report provides the results of the impact evaluation for the third year (2022-23) of the NTP. It explores the impact of the NTP on pupil attainment in English and maths for all pupils, pupils eligible for pupil premium (PP pupils) and pupils with prior low attainment (PLA pupils) at both school level and pupil level. It also seeks to understand if there are any differential effects according to tutoring dosage, pupil characteristics (gender, having special educational needs or disabilities (SEND), having English as an Additional Language (EAL), and ethnicity), or school characteristics (region, Income Deprivation Affecting Children Index (IDACI)). It will also explore if there have been any longer-term benefits for pupils who received tutoring in the academic year 2021-22 (i.e., in the second year of the NTP).

¹ 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.

² The socioeconomic attainment gap refers to the differences in educational attainment between more and less socioeconomically affluent pupils.

This report follows the two previous impact evaluations for the first and second years of the NTP. The first of these evaluations related to the academic year 2020-21 and found positive effects of participating in TP on KS4 maths and English outcomes but only in schools where at least 70% of disadvantaged participated in tutoring (Poet et al., 2022a). There was also a positive relationship between the number of hours of tuition and English outcomes for KS2 pupils (Poet et al., 2022b).

Meanwhile, the results from the evaluation of the second year of the NTP (2021-22) found that participation in SLT was associated with small improvements in KS2 and KS4 maths outcomes. There was also some more limited evidence that participation in SLT was associated with small improvements in KS2 and KS4 English outcomes. However, we did not detect any evidence that participation in AM/TP led to improvements in either KS2 or KS4 English or maths outcomes. In addition, in some cases participation in AM/TP was associated with negative effects on English (KS2 and KS4) and maths (KS4) outcomes, although these effects were very small. This evaluation also found that a higher tutoring dosage (number of hours of tutoring pupils received) and/or concentration (the proportion of pupils within a school that received tutoring) was associated with better English and maths outcomes for SLT but not for AM/TP.

This report also accompanies the implementation and process evaluation (IPE) report for the third year of the NTP (Lynch et al., 2023) which focused on the implementation of third year of the NTP and the extent to which tutoring has been embedded within schools.

The National Tutoring Programme (NTP) and the socioeconomic attainment gap

The NTP was initially set-up in the academic year 2020-21 to provide additional targeted support for pupils who were socioeconomically disadvantaged, as these pupils were most affected by the disruption to learning during the Covid-19 pandemic. The aim of the NTP was to help these pupils catch-up on missed learning and reduce the socioeconomic attainment gap (DfE, 2020). The NTP also aimed to embed tutoring as an established intervention schools could use to help close the attainment gap.

The NTP consists of three tutoring routes:

Academic mentors (AM): academic mentors are employed by the school. Initially this route was only available to schools with the highest numbers of socioeconomically disadvantaged pupils, but in the third year of the NTP was an option for all schools to use.

Tuition partners (TP): this route offered subsidised tuition to schools from approved tuition partners.

School-led tutoring (SLT): this route was introduced in the academic year 2021-22 and allowed schools to source local tutoring provision.

In recognition schools needed the flexibility to tailor tutoring to the needs of their pupils (GOV.UK, 2023b), in the third year of the NTP schools were able to use any combination of the three routes described above. All routes were subsidised at 60% during the third year of the NTP (a reduction in the subsidy level from year 2) with schools making up the remaining costs from their core budget, typically pupil premium funding (GOV.UK, 2023b). This flexibility provided greater autonomy for schools to deliver tutoring but means that no data is available regarding combination of routes schools chose to use or the subject in which pupils received tutoring.

In the third year of the NTP (2022/23) the focus shifted from mitigating the impact of missed learning during the pandemic to tackling the persistent attainment gap between socioeconomically disadvantaged pupils and their peers. This gap in attainment is measured by the Department for Education (DfE) using the disadvantage gap index. This index compares the attainment of socioeconomically disadvantaged pupils (defined based on Free Schools Meals (FSM) eligibility) with all other pupils. It is measured on a scale of -10 to +10, where 0 indicates an equal distribution of scores between disadvantaged pupils and their non-disadvantaged peers. For primary school pupils, the index compares Key Stage 2 attainment in reading and maths. For secondary school pupils it compares GCSE grades in English and maths.

In 2022, the disadvantage gap for primary school pupils had widened to 3.23 (GOV.UK, 2022) – the widest level since 2012, when it had previously been decreasing between 2012 and 2019. For secondary pupils, the disadvantage gap index was already widening before the pandemic, rising from 3.66 in 2017 to 3.70 in 2019. The gap widened further in the pandemic, to 3.84 in 2022, the highest level since 2012 (GOV.UK, 2022). It is important to note that the disadvantage gap prior to Covid was at least twice as large as the impact of Covid on attainment. However, the widening disadvantage gaps provides further evidence that disrupted learning during the Covid-19 pandemic had a disproportionate impact on socioeconomically disadvantaged pupils. Tackling this attainment gap was thus a priority for the Government in the third year of the NTP. With Government funding for tutoring due to continue reducing in subsequent years it was also a priority to embed tutoring as a permanent fixture in schools for supporting socioeconomically disadvantaged pupils to help improve outcomes long term.

Impact evaluation aims

Primary aim: to evaluate the impact of the third year of the NTP on English and maths outcomes for a) all pupils, b) pupils eligible for pupil premium (PP pupils) and c) pupils with prior low attainment (PLA pupils) at KS1 or KS2.

In addition, this evaluation also sought to understand:

- whether the impact of the NTP varied according to the number of tutoring hours pupils received (dosage).
- whether there were differences in the impact of the NTP according to pupil (gender, SEND status, EAL, ethnicity) or school (region, IDACI) characteristics.
- the longer-term impact of receiving the NTP in the academic year 2021-22.

Research questions

The research questions for this report relate specifically to the impact of the NTP in the third year of the programme, as listed below. Full research questions including those for the IPE can be found in the Study Plan (Welbourne et al., 2023).

RQ5: What is the impact of the NTP delivered in 2022/23 on the educational attainment outcomes (in maths and English) of pupils who are in Year 6 and Year 11 in 2022/23?

- RQ5.1: for all pupils? (school-level impact estimates)
- RQ5.2: for pupils eligible for pupil premium (PP)? (school-level impact estimates)
- RQ5.3: for pupils with lower prior attainment than the expected standard (PLA)? (school-level impact estimates)
- RQ5.4: for all tutored pupils? (pupil-level impact estimates)
- RQ5.5: for tutored PP pupils? (pupil-level impact estimates)
- RQ5.6: for tutored PLA pupils? (pupil-level impact estimates)
- RQ5.7: how do these impacts vary by pupil characteristics and region? (school-level impact estimates)
- RQ5.8: how does the impact of tutoring vary with dosage? (school-level impact estimate)

RQ6: What is the impact of the NTP delivered in prior years (2020/21 or 2021/22) on the educational attainment outcomes (in maths and English) of pupils who are in Year 6 and Year 11 in 2022/23?

- RQ6.1: for all pupils? (school-level impact estimates)
- RQ6.2: for pupils eligible for pupil premium (PP)? (school-level impact estimates)
- RQ6.3: for pupils with lower prior attainment than the expected standard (PLA)? (school-level impact estimates)

Methods

Participant sample

The participant sample for this analysis was all pupils attending state-funded schools in England who were in Year 6 (KS2 analysis) or Year 11 (KS4 analysis) in the 2022-23 academic year. Within these year groups, we explored the impact of the NTP for three groups of pupils: all pupils, PP pupils and PLA pupils.

Outcome measures

The baseline and outcome measures for each year group were as follows:

Year 6

- Baseline: KS1 (i.e., Year 2) reading and maths attainment (from 2018-19)
- Outcome: KS2 (i.e., Year 6) reading and maths attainment (from 2022-23)

Year 11

- Baseline: KS2 (i.e., Year 6) reading and maths attainment (from 2017-18)
- Outcome: KS4 (i.e., Year 11) English and maths attainment (from 2022-23)

Data sources

Data for this impact evaluation was obtained from the following sources:

National Pupil Database (NPD): this was used to identify which pupils received tutoring and therefore intervention and comparison schools. The pupil characteristic and prior attainment information held within the NTP was also used to identify PP and PLA pupils. The NPD also provided information about tutoring dosage, in terms of the numbers of hours of tutoring pupils received, which is based on school census data. However, the NPD does not contain any information about the tutoring route (i.e., AM, TP or SLT), or which subject pupils received tutoring in for 2022-23 academic year as this data was not collected by DfE. This means that no data about tutoring route or subject is available for this evaluation. The NPD was also used to access last year's school census data to understand pupils' previous participation in SLT in the academic year 2021-22 for the long-term analysis (RQ6).

DfE's NTP archive: this archive contains information about previous tutoring pupils have received from the AM or TP routes. It was used to provide data about participation in AM and TP for the second year of the programme in 2021-22 (i.e., when pupils were in Year 5 or Year 10) for the analysis of the longer-term impact of the NTP (RQ6).

NFER's register of schools: this database is derived from DfE's publicly available 'Get Information about Schools' database. It contains characteristic information about schools and was used in the analysis to understand the composition of schools contained within the sample and for school-level covariates.

Defining intervention and comparison groups at school and pupil level

For the reasons described in the introduction, no information about tutoring route or tutored subject was collected in the third year of the NTP, so this data was not available for the analysis. This means that we compared the impact of participating in the NTP with non-participation in the NTP rather than NTP route vs no NTP or NTP via other routes, as in the year 2 analysis. In addition, the fact that we do not know the tutored subject means that there is an extra level of dilution in the data. For example, when looking at the impact of the NTP on maths outcomes, we do not know how many pupils in the intervention group received maths tuition as opposed to tutoring in another subject.

School-level impact analysis

Our school-level analysis estimated the impact of the NTP on the educational attainment outcomes for a) all pupils, b) PP pupils and c) PLA pupils. It also explored variation in the impact of the NTP related to school and pupil characteristics, as well as investigating the longer-term impact of tutoring for pupils in Year 6 or Year 11 in 2022-23 who received tutoring the previous year (i.e., in the academic year 2021-22 when they were in Year 5 or 10).

The intervention and comparison groups for the school-level analysis were defined as follows:

Intervention group: all pupils (or subgroups of PP or PLA pupils) who attended schools that participated in the NTP in the 2022-23 academic year, regardless of whether not the pupil themselves received tutoring.

Comparison group: all pupils (or subgroups of PP or PLA pupils) who attended schools that did not participate in the NTP in the 2022-23 academic year.

Defining the groups in this way avoids the problem of pupil-level selection bias. However, as in previous years of the NTP evaluation, it creates the problem of dilution. This is because although all pupils in the intervention schools could be potential recipients of the NTP, not all pupils received tutoring.

To reduce the dilution in this analysis, we imposed a minimum threshold for the percentage of pupils within the relevant year group who received tutoring. The threshold was determined using a data driven approach to maximise statistical power (see technical appendix for details). This resulted in a minimum participation rate of 54% for

Year 6 and 64% for Year 11. This means that the KS2 intervention group only contained primary schools where at least 54% of Year 6 pupils received tutoring. For KS4 the intervention group only contained secondary schools where at least 64% of Year 11 pupils received tutoring.

Pupil-level impact analysis

Main analysis

Our pupil-level analysis aimed to estimate the impact of the NTP on the educational attainment outcomes of pupils who were in Year 6 or Year 11 in the 2022-23 academic year.

The advantage of the pupil-level analysis is that it reduces dilution, as all pupils in the intervention group received tutoring. It should therefore provide a better estimate of the impact of tutoring on the attainment of individual pupils. However, a key challenge with the pupil-level analysis was selection bias, as schools were able to choose which pupils they felt would benefit from tuition. This means that 1) schools were likely to choose pupils most in need of support, and 2) the pupil selection mechanism (or criteria) is partially unknown and unavailable for the evaluation. In addition, it is also likely that the selection mechanism varied across schools. The consequence of this is that in the previous two years of the NTP evaluation we were unable to create a valid comparison group of untutored pupils. We previously attempted to build a predictive model to identify pupils in schools that did not participate in the NTP who would likely have been selected for tutoring had their school taken part in the NTP (see Poet et al., 2022a, 2022b; Lucas et al., 2023). However, in both cases the model did not perform well enough to be considered a reliable basis for constructing a comparison group.

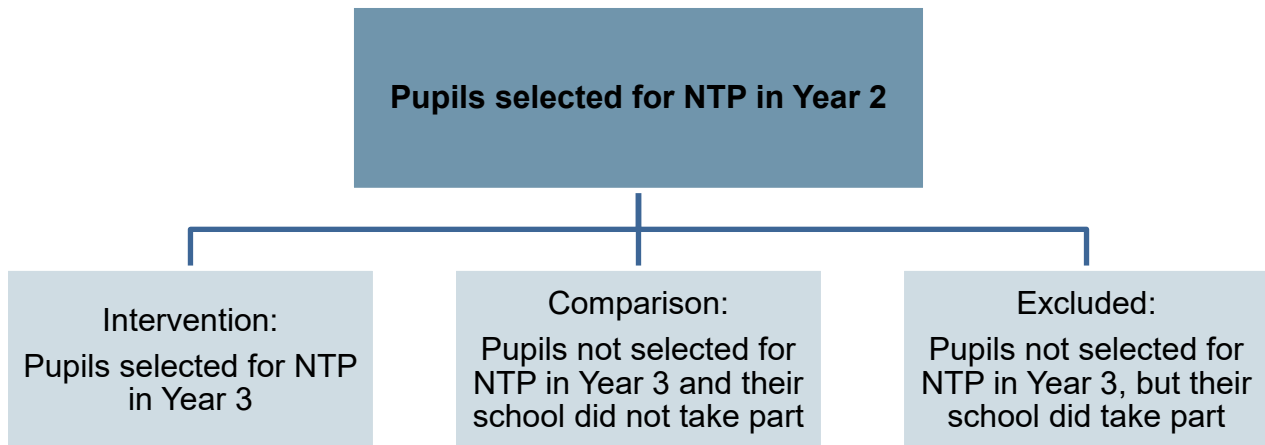
Therefore, for the year 2 impact evaluation, we reduced selection bias by limiting the analysis to pupils who had been tutored during the previous academic year (see (Lucas et al., 2023)). We took the same approach this year. We restricted the pupil-level analysis to pupils who received tutoring in the previous academic year (i.e., in the academic year 2021-22). We then compared the impact of the NTP for pupils who also received tutoring in the third year of the NTP with those who did not receive tutoring because their school did not take part in the NTP in 2022-23. This approach reduces the effects of selection bias (see limitations) but means that the difference between these groups is likely to be smaller than the impact of receiving tutoring compared with not receiving tutoring. This meant that our intervention and comparison groups for this analysis were as follows (see also Figure 1):

Intervention group: pupils who were selected for tutoring in the second and third years of the NTP.

Comparison group: pupils who were selected for tutoring in the second year of the NTP, but their school did not participate in the third year of the NTP.

Note that pupils who were selected for tutoring in the second year of the NTP but were not selected for tutoring in the third year of the NTP, even if their school took part, were excluded from this analysis.

Figure 1: Intervention and comparison groups for the main pupil-level analysis



Sensitivity analysis

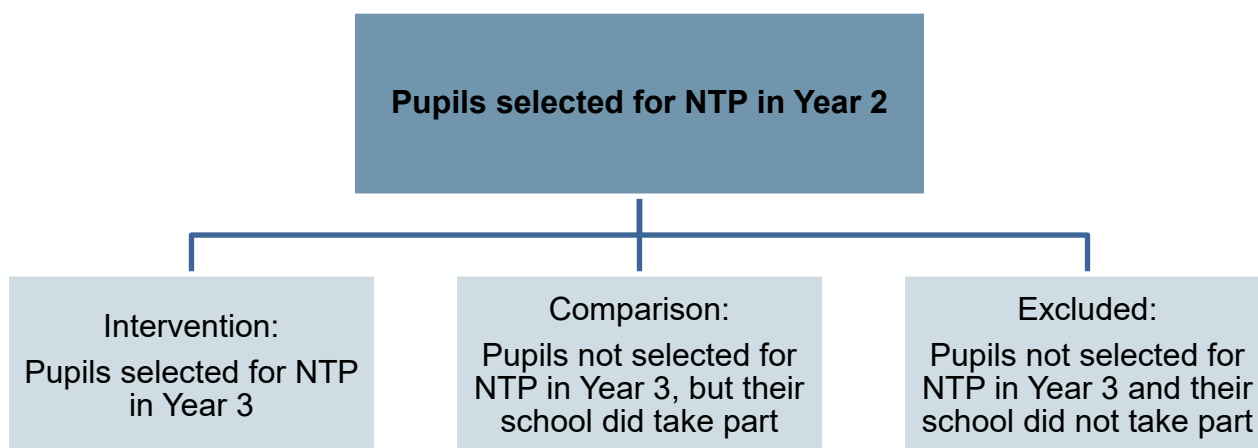
Our approach to the main pupil-level analysis aimed to reduce selection bias. However, it is possible that some selection bias remains within this analysis. This is because pupils who received tutoring in year 2 and then again in year 3 are likely to be pupils who were still at risk of falling behind after receiving the year 2 tutoring. To explore the possibility of any lingering selection bias within the pupil-level analysis, we also undertook a sensitivity analysis. For this analysis the intervention and comparison groups were defined as follows (see also Figure 2):

Intervention group: pupils who were selected for tutoring in the second and third years of the NTP. (note this is the same as for the main pupil-level analysis).

Comparison group: pupils in schools that participated in the NTP in both the second and third years, but who were selected for tutoring in the second year only.

Note that this time the comparison group for the main analysis (i.e., pupils who were selected for tutoring in the second year of the NTP but did not receive tutoring in the third year of the NTP because their school did not take part) were excluded from the analysis.

Figure 2: Intervention and comparison groups for the pupil-level sensitivity analysis



The sensitivity analysis compared pupils within the same schools (i.e., within those schools that continued to offer the NTP for a third year). Therefore, we would expect that the pupils schools chose to offer tutoring to in the third year of the NTP were more in need of support than those pupils who did not continue with tutoring after the second year of the NTP. This scenario would result in a negative selection bias – this bias (if it exists) will also be present but to a lesser extent in the main analysis. Therefore, if there is still lingering selection bias within the pupil-level analysis then we would expect that the effect sizes for the sensitivity analysis will be smaller (or more negative) than those for the main pupil-level analysis. This result would indicate that there is still likely to be negative selection bias within the main pupil-level analysis.

Statistical analysis

Matching

The matching process was done separately for each analysis. The first step in the process of building comparison groups was to identify 'common support'. This ensured that no key school characteristics³ which could determine eligibility for the NTP, or the likelihood of engagement, were present in only the intervention group or the comparison group.

To ensure that the intervention and comparison groups were closely matched we conducted statistical matching using entropy balancing (Hainmueller, 2012). Entropy

³ Key school characteristics were school prior attainment, priority area for raising school standards, region, and proportion of pupils eligible for pupil premium.

balancing is a method that assigns weights to comparison pupils to balance⁴ observed variables between the groups; these weights are included in subsequent regression modelling. Unlike many other data pre-processing methods, entropy balancing balances the variables directly, rather than via propensity scores.

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. For both the school and pupil-level analysis, fresh matching was conducted for each pupil group (all pupils, PP pupils, and PLA pupils). This approach resulted in a well-balanced match between the intervention and comparison groups (see technical appendix for more details of the entropy balancing and the degree of balance achieved between groups).

Sample characteristics

Pupil and school characteristics were analysed from the sources described in 'Data sources' using descriptive statistics.

School-level analysis

At school level, the impact of the NTP was assessed separately for Year 6 and Year 11 for the three different pupil groups (all pupils, PP pupils, PLA pupils). As noted above no information was available about tutoring route or subject. This meant that we investigated the impact of the NTP (any route, any subject) on English and maths outcomes at school level for Year 6 and Year 11 using linear mixed effects models, with scaled score or point score as the outcome, intervention group and any other appropriate covariates (see technical appendix for details) as fixed effects and school as the random effect.

For Year 6 this resulted in 6 linear mixed effects models (including the weights described in 'Matching') to analyse the impact of the NTP on:

- KS2 maths outcomes for all pupils (school level)
- KS2 maths outcomes for PP pupils (school level)
- KS2 maths outcomes PLA pupils (school level)
- KS2 English outcomes all pupils (school level)
- KS2 English outcomes for PP pupils (school level)
- KS2 English outcomes for PLA pupils (school level)

⁴ 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).

Similarly for Year 11 this resulted in 6 linear mixed effects models (including the weights described in ‘Matching’) to analyse the impact of the NTP on:

- KS4 maths outcomes for all pupils (school level)
- KS4 maths outcomes for PP pupils (school level)
- KS4 maths outcomes PLA pupils (school level)
- KS4 English language outcomes all pupils (school level)
- KS4 English language outcomes for PP pupils (school level)
- KS4 English language outcomes for PLA pupils (school level)

Dosage analysis

For this part of the analysis, we adopted a new approach as we recognised that a key question that school leaders need to know the answer to is:

How should I best distribute a fixed tutoring budget among disadvantaged pupils – is it better to provide fewer hours for more pupils, or to concentrate more tutoring hours on a smaller number of pupils?

This means understanding the optimum tutoring dosage at a school level is key rather than understanding the effect of more hours of tutoring on an individual pupil. The NTP guidance stipulates that tutoring courses should be 12-15 hours (GOV.UK, 2024). However, an evidence review conducted by EEF suggests that around 30 hours of tuition typically shows the greatest impact (EEF, 2021a). In addition, an Independent Review of tutoring conducted by Ofsted (GOV.UK, 2023a) found that some schools were dividing hours of tutoring among disadvantaged pupils so that they could all benefit, but this meant each pupil received fewer hours of tutoring. The dosage analysis therefore sought to understand how best to distribute a fixed number of tutoring hours among pupils to achieve the best overall school outcomes, on average, across all pupils (including both tutored and untutored pupils). This was done using two quantities:

ATH-NTP, the average tutoring hours amongst pupils that received NTP tutoring at a school.

ATH-all, the average tutoring hours amongst all pupils at a school (i.e. both those selected for tutoring and those that weren't).

To represent a school's fixed tutoring budget, we 'fixed' ATH-all by conditioning on it as a model covariate. We then explored the impact of ATH-NTP on attainment outcomes for all Year 6 or Year 11 pupils in the school. Importantly, because we fixed ATH-all in this analysis, higher ATH-NTP means focusing more tutoring hours on a smaller number of pupils. This is because with a fixed budget of tutoring hours, hours for tutored pupils can only be increased (i.e. higher ATH-NTP) by tutoring fewer pupils.

We suspected that the relationship between ATH-NTP and attainment would be non-linear and that there would be an 'optimum' value of ATH-NTP for attainment, represented by the maximum of an inverted 'U'-shaped curve (i.e. a quadratic relationship). We used linear and polynomial models to investigate whether this was the case, using statistical testing to decide between polynomial models⁵.

We excluded schools with an ATH-all of greater than 40 (3% of schools at KS2 and 4% at KS4) as we suspected the few schools reporting very high levels of tutoring might have made errors in data recording. As a sensitivity check we repeated the model but only included schools that offered 1 to 10 hours of tutoring on average across all pupils (excluding 38% of schools at KS2 and 43% at KS4). Results from this sensitivity check are not included in the main report (see technical appendix).

Pupil characteristics

This analysis investigated the impact of the NTP for pupils with the characteristics listed below. In all cases, the main impact analysis was repeated for each subgroup of pupils. Intervention and control schools were defined, using the same minimum NTP participation thresholds as the main analysis.

- SEND status: we investigated the impact for the NTP for pupils with and without SEND. Due to the small numbers of pupils with some types of SEND we did not do any additional analysis to explore the impact of the NTP for pupils with different types of SEND.
- 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 the impact of the NTP for pupils with English as a first language and pupils with English as an Additional Language (EAL).
- Sex: we investigated the impact of the NTP for male and female pupils.

School characteristics

Geographic region: This analysis investigated if the impact of the NTP varied according to geographic region. To do this we restricted the analysis to each of the nine geographic regions in England in turn (London, East Midlands, West Midlands, East of England,

⁵ This approach does not consider all possible relationships between ATH_{NTP} and attainment (e.g. logarithmic, leading to a plateau after a certain point). However, we consider it reasonable to expect there is a maximum a priori: neither very low ATH_{NTP} (e.g. an hour of tutoring for all pupils) nor very high ATH_{NTP} (e.g. hundreds of hours for a single pupil) seem likely to be the best overall strategy.

North East, North West, South East, South West, Yorkshire & Humber) and explored the impact of the NTP at school level within each region.

Income deprivation affecting children index (IDACI): This analysis explored if the impact of the NTP varied according to IDACI quintile. Every postcode in England is ranked according to IDACI with lower ranks corresponding to more socioeconomically disadvantaged postcodes. This means that pupils living in disadvantaged areas will have a lower IDACI rank. These ranked postcodes were then divided into quintiles with the most socioeconomically disadvantaged postcodes ranked as 1 and the most socioeconomically advantaged postcodes ranked as 5. The attainment of pupils who attended schools that participated in the NTP was then compared with that of pupils who attended schools that did not participate in the NTP within each quintile.

Longer-term impact of the NTP

To assess the longer-term impact of the NTP on pupil attainment, we explored the impact of school-level participation in the second year of the NTP (i.e., 2021-22) when pupils were in Year 5 or Year 10 on KS2 and KS4 outcomes at the end of the 2022-23 (i.e., when pupils were in Year 6 or Year 11). The intervention group was defined based on school participation in the NTP in 2021-22 rather than 2022-23. This meant route data (i.e., SLT, AM and TP) was available as well as subject information for AM and TP for year 2. Consequently, as with the analysis for the impact evaluation of the second year of the NTP (Lucas et al., 2023), the intervention group consisted of schools participating in the NTP in the relevant route and, where applicable, subject. Meanwhile the comparison group consisted of schools not participating in the relevant route or subject, but they could potentially have taken part in other tutoring routes. Participation in other tutoring routes was therefore matched between the intervention and comparison groups as with the analysis for the year 2 impact evaluation (see Lucas et al., 2023). Due to the small number of schools that participated in AM in the 2021-22 academic year, AM (in the relevant subject) was combined with SLT (any route) for the purposes of the analysis to increase the statistical power. This is because AM and SLT were similar in that school staff (or staff known to the school) delivered tutoring for these routes rather than an external provider as is the case in TP⁶. This means that for this analysis there were eight definitions of the intervention:

- School-level participation in SLT (any subject) or AM (English) amongst pupils in Year 5 in the academic year 2021-22.
- School-level participation in TP (English) amongst pupils in Year 5 in the academic year 2021-22.

⁶ Note this differs from the impact analysis of Year 2 of the NTP where the AM and TP routes were combined.

- School-level participation in SLT (any subject) or AM (maths) amongst pupils in Year 5 in the academic year 2021-22.
- School-level participation in TP (maths) amongst pupils in Year 5 in the academic year 2021-22.
- School-level participation in SLT (any subject) or AM (English) amongst pupils in Year 11 in the academic year 2021-22.
- School-level participation in TP (English) amongst pupils in Year 11 in the academic year 2021-22.
- School-level participation in SLT (any subject) or AM (maths) amongst pupils in Year 11 in the academic year 2021-22.
- School-level participation in TP (maths) amongst pupils in Year 11 in the academic year 2021-22.

For each definition, new matches were performed and all pupils attending a school that participated in the relevant intervention route(s), and where applicable subject, were included in the intervention group. Pupils attending the remaining schools formed the comparison group (see above). These definitions were applied to the analysis which used linear mixed effects models to look at the impact of each intervention for all pupils, PP pupils and PLA pupils. We also set a minimum participation threshold for the interventions groups as in the main analysis. This meant that for KS2 only schools with a minimum participation rate of 34% for AM/SLT or TP (as applicable) among Year 6 pupils were included in the analysis. Meanwhile, for KS4 only schools with a minimum participation rate of 27% among Year 11 pupils were included for AM/SLT and 6% for TP (as very few schools offering TP to KS4 pupils had a high percentage of pupils taking part in this route).

Pupil-level analysis

Like the school-level analysis, the pupil-level analysis also assessed the impact of the NTP separately for Year 6 and Year 11 pupils for the three different pupil groups (all pupils, PP pupils, PLA pupils). This meant that we investigated the impact of the NTP (any route, any subject) on English and maths outcomes at pupil level for Year 6 and Year 11 using linear mixed effects models, with scaled score or point score as the outcome, intervention group and any other appropriate covariates as fixed effects and school as the random effect.

For Year 6 this resulted in 6 linear mixed effects models (including the weights described in 'Matching') to analyse the impact of the NTP on:

- KS2 maths outcomes for all pupils (pupil level)
- KS2 maths outcomes for PP pupils (pupil level)
- KS2 maths outcomes PLA pupils (pupil level)

- KS2 English outcomes all pupils (pupil level)
- KS2 English outcomes for PP pupils (pupil level)
- KS2 English outcomes for PLA pupils (pupil level)

Similarly for Year 11 this resulted in 6 linear mixed effects models (including the weights described in 'Matching') to analyse the impact of the NTP on:

- KS4 maths outcomes for all pupils (pupil level)
- KS4 maths outcomes for PP pupils (pupil level)
- KS4 maths outcomes PLA pupils (pupil level)
- KS4 English language outcomes all pupils (pupil level)
- KS4 English language outcomes for PP pupils (pupil level)
- KS4 English language outcomes for PLA pupils (pupil level)

Limitations

Without randomisation, it is inevitable that any evaluation of the NTP will have several limitations. We have described the key limitations below as they provide important context in which to consider the results presented later in this report, as well as the conclusions we are able to draw.

Dilution

There are two sources of dilution in the evaluation which make it likely that the effect sizes we observe will be underestimates.

Firstly, as with previous evaluations of the NTP, not all pupils included in the intervention group for the school-level analysis received tutoring. Our dilution threshold reduces the level of dilution present but does not eliminate it. This means that the intervention effects identified for all the school-level analyses (including for pupil characteristics and long-term impact of the NTP) will be underestimates.

Secondly, in the third year of NTP, there was no information available regarding the tutored subject. This means at both school and pupil level, the intervention groups could have received tutoring in a different subject to the outcome being measured. For example, pupils who only received tutoring in English would be included in the intervention group for maths outcomes even though they did not receive maths tuition. Based on the results of the second year of the NTP where subject data was available for the AM and TP routes, we think this may be a larger issue for English than for maths, as a higher percentage of pupils received maths tuition (see Table 1). It is therefore possible that any differences in impact between maths and English tuition could be accounted for by differences in dilution. However, we are aware that SLT made up the majority of the

tuition offered in year 2 of the NTP, and no subject data was available for SLT so the division of tutoring between subjects may look quite different. There is no indication, based on the NTP Year 2 IPE report (Lynch et al., 2023), that schools prioritised maths more frequently than English for SLT tutoring: 40% of surveyed schools said they prioritised maths for SLT tutoring and 43.4% prioritised English⁷.

Table 1: Percentage of pupils selected for tutoring in maths or English in Year 2 of the NTP

Among pupils selected for AM/TP tutoring	AM/TP Maths	AM/TP English
Percentage of KS2 pupils selected for each subject	68%	32%
Percentage of KS4 pupils selected for each subject	58%	42%

Source: Randstad tuition participation data.

Pupil-level selection bias

Schools could select pupils for tutoring who they felt would benefit from support. This creates the problem of pupil-level selection bias as schools are likely to choose pupils for tutoring who are struggling most and would otherwise achieve lower attainment outcomes than their peers. Although we have attempted to reduce selection bias by restricting the pupil-level analysis to pupils who were selected for tutoring in the second year of the NTP, it is likely that some negative selection bias remains. This is because pupils who were selected for both the second and third years of the NTP are likely to have more difficulties than pupils who were selected for the second year but not the third year of the NTP. The pupil-level sensitivity analysis will test if this is the case. If the effect size in the sensitivity analysis is smaller than the effect size in the main analysis, then it is likely that selection bias remains, and the effect sizes observed in the main analysis are likely to underestimate the ‘true’ effect size.

Multiple simultaneous tests

The analyses described in this report use multiple simultaneous tests on the same datasets. While it is possible to apply statistical corrections to account for multiple comparisons these corrections have not been applied here. This is because 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

⁷ Schools could say they prioritised any number of subjects; it was not a question about whether they prioritised maths or English more highly for tutoring.

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.

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 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). This means the results presented in this report need to be considered in the context of multiple testing, as it is likely that some statistically significant effects would not survive correction for multiple comparisons. This is particularly relevant for the pupil and school characteristic analyses, where a large number of tests were conducted with little prior evidence as to what effect sizes we might reasonably expect within these groups. In addition, some of these subgroup analyses only contained very small samples of participants, making it more likely spurious effects could be found.

Sample characteristics

This chapter shows the composition of pupils included in the school- and pupil-level analyses by key pupil characteristics, further broken down by whether pupils received NTP tutoring in 2022-23. Additionally, for the school-level samples the average tutoring hours amongst NTP pupils and proportion of NTP pupils at each school is shown.

It is worth noting that these are the initial samples and pupil number will differ in the final school- and pupil-level models. In the final pupil-level models the comparison group was restricted to pupils at schools not participating in the NTP (main analysis) or pupils at schools participating in the NTP (sensitivity analysis). Additionally, a small number of pupils were removed at the matching stage to enforce common support restrictions, for some school- and pupil-level models. The characteristics of the pupils in a particular model can be viewed in the matching balance tables, which are included in the technical appendix.

School-level evaluation samples

In total 603,538 Year 6 pupils were included in the KS2 school-level evaluation sample (Table 2). Amongst these pupils 175,378 (29%) were PP and 115,085 (19%) were PLA. NTP tutoring was delivered to 29% (N=175,726) of the sample, a slight increase compared to the 27% tutored in the Year 6 sample for last year's (2021-22) NTP evaluation.

The sample characteristics of intervention pupils⁸ in the main school-level analysis are shown in the third column of Table 2. A higher proportion of Year 6 intervention pupils were PP than in the whole KS2 sample (38% versus 29%) and intervention pupils more likely to be in the most deprived 40% of IDACI scores (54% versus 40% in the whole KS2 sample). As PP and IDACI quintile were included in the matching process and as regression covariates it is unlikely that these imbalances caused bias in the school-level results. Other characteristics were broadly similar in the intervention group compared to the whole KS2 sample.

Amongst Year 6 pupils 42% of pupils that received NTP tutoring were PP, compared to 29% in the whole sample. Pupils from more deprived regions were also more likely to receive tutoring: those in the most deprived 40% of IDACI scores represented 50% of tutored pupils. The evidence suggests that schools prioritised PP pupils and pupils from more deprived areas to receive tutoring. However, the proportion of PP pupils decreased slightly compared with last year from 46% to 42% indicating that disadvantaged pupils may have been a slightly lower priority for tutoring this year. PLA and SEN pupils also

⁸ The school-level intervention group was all pupils at schools where a minimum percentage of pupils received NTP tutoring. See the description of the school-level analysis in the methods section.

appear to have been prioritised to some extent, although less so than PP pupils. There was no indication that gender, EAL status or ethnicity were associated with tutoring receipt amongst the Year 6 sample.

Table 2: Number and percentage of Year 6 pupils included in the KS2 school-level evaluation sample split by pupil characteristics and NTP selection

	Total in sample	Intervention group	Selected for the NTP	Not selected for the NTP
Total number of pupils	603,538 (100%)	121,718 (100%)	175,726 (100%)	427,812 (100%)
Male	303,913 (50%)	61,049 (50%)	86,003 (49%)	217,910 (51%)
Female	299,625 (50%)	60,669 (50%)	89,723 (51%)	209,902 (49%)
PP pupils	175,378 (29%)	46,044 (38%)	74,585 (42%)	100,793 (24%)
Non-PP pupils	428,160 (71%)	75,674 (62%)	101,141 (58%)	327,019 (76%)
PLA pupils	115,085 (19%)	25,733 (21%)	44,672 (25%)	70,413 (16%)
Non-PLA pupils	488,453 (81%)	95,985 (79%)	131,054 (75%)	357,399 (84%)
PP and/or PLA	239,039 (40%)	58,255 (48%)	96,066 (55%)	142,973 (33%)
Pupils with SEN	107,154 (18%)	23,824 (20%)	38,306 (22%)	68,848 (16%)
Pupils without SEN	496,384 (82%)	97,894 (80%)	137,420 (78%)	358,964 (84%)
Pupils with EAL	120,431 (20%)	26,422 (22%)	35,871 (20%)	84,560 (20%)
Pupils without EAL	483,107 (80%)	95,296 (78%)	139,855 (80%)	343,252 (80%)
White	445,056 (74%)	88,314 (73%)	128,189 (74%)	316,867 (75%)
Black	31,871 (5%)	7,883 (7%)	11,231 (6%)	20,640 (5%)
Chinese	2,917 (<1%)	557 (<1%)	518 (<1%)	2,399 (1%)
Other Asian	67,599 (11%)	13,366 (11%)	18,588 (11%)	49,011 (12%)
Mixed ethnicity	39,707 (7%)	7,947 (7%)	11,961 (7%)	27,746 (7%)
Other ethnicity	11,206 (2%)	2,698 (2%)	3,770 (2%)	7,436 (2%)
IDACI quintile 1 (most deprived)	118,354 (20%)	36,378 (30%)	46,844 (27%)	71,510 (17%)

	Total in sample	Intervention group	Selected for the NTP	Not selected for the NTP
IDACI quintile 2	118,872 (20%)	28,954 (24%)	40,432 (23%)	78,440 (18%)
IDACI quintile 3	120,352 (20%)	22,762 (19%)	34,456 (20%)	85,896 (20%)
IDACI quintile 4	122,516 (20%)	19,403 (16%)	30,118 (17%)	92,398 (22%)
IDACI quintile 5 (least deprived)	123,444 (20%)	14,221 (12%)	23,876 (14%)	99,568 (23%)

Note: Cell entries are number (%) of pupils. Percentages may not sum to 100 due to rounding.

Source: National Pupil Database 2022/2023 (School Census, KS2).

The KS4 school-level evaluation sample included 528,250 Year 11 pupils in total, of which 125,217 (24%) were PP and 102,804 (19%) were PLA (see Table 3). NTP tutoring was delivered to 34% (N=180,054) of the sample, a slight increase on the 33% tutored in the Year 11 sample for last year's NTP evaluation.

More Year 11 intervention pupils were PP than in the whole KS4 sample (30% versus 24%) and more intervention pupils were in the most deprived 40% of IDACI scores (54% versus 39%). Other characteristics were similar in the intervention group compared to the whole KS4 sample.

31% of Year 11 tutored pupils were PP, compared to 24% in the whole sample. This is a slight decline since the previous year of NTP, where 35% of Year 11 tutored pupils were PP. Pupils from more deprived regions were again more likely to received tutoring, with those in the most deprived 40% of IDACI scores represented 48% of tutored pupils in Year 11. At KS4 white pupils and pupils with EAL were slightly underrepresented amongst NTP pupils, compared to the whole sample.

Table 3: Number and percentage of Year 11 pupils included in the KS4 school-level evaluation sample split by pupil characteristics and NTP selection

	Total in sample	Intervention group	Selected for the NTP	Not selected for the NTP
Total number of pupils	528,250 (100%)	116,070 (100%)	180,054 (100%)	348,196 (100%)
Male	266,545 (50%)	57,622 (50%)	89,343 (50%)	177,202 (51%)
Female	261,705 (50%)	58,448 (50%)	90,711 (50%)	170,994 (49%)

	Total in sample	Intervention group	Selected for the NTP	Not selected for the NTP
PP pupils	125,217 (24%)	35,169 (30%)	56,568 (31%)	68,649 (20%)
Non-PP pupils	403,033 (76%)	80,901 (70%)	123,486 (69%)	279,547 (80%)
PLA pupils	102,804 (19%)	25,144 (22%)	39,818 (22%)	62,986 (18%)
Non-PLA pupils	425,446 (81%)	90,926 (78%)	140,236 (78%)	285,210 (82%)
PP and/or PLA	191,073 (36%)	49,908 (43%)	80,135 (45%)	110,938 (32%)
Pupils with SEN	72,464 (14%)	16,816 (14%)	27,457 (15%)	45,007 (13%)
Pupils without SEN	455,786 (86%)	99,254 (86%)	152,597 (85%)	303,189 (87%)
Pupils with EAL	86,027 (16%)	23,243 (20%)	32,758 (18%)	53,269 (15%)
Pupils without EAL	442,223 (84%)	92,827 (80%)	147,296 (82%)	294,927 (85%)
White	385,897 (74%)	80,330 (70%)	126,460 (71%)	259,437 (76%)
Black	31,122 (6%)	9,165 (8%)	13,624 (8%)	17,498 (5%)
Chinese	2,058 (<1%)	397 (<1%)	554 (<1%)	1,504 (<1%)
Other Asian	59,577 (11%)	14,460 (13%)	21,429 (12%)	38,148 (11%)
Mixed ethnicity	31,586 (6%)	7,184 (6%)	11,284 (6%)	20,302 (6%)
Other ethnicity	9,332 (2%)	2,608 (2%)	3,714 (2%)	5,618 (2%)
IDACI quintile 1 (most deprived)	102,651 (19%)	33,265 (29%)	45,585 (25%)	57,066 (16%)
IDACI quintile 2	104,893 (20%)	28,449 (25%)	40,652 (23%)	64,241 (18%)
IDACI quintile 3	106,109 (20%)	21,859 (19%)	35,501 (20%)	70,608 (20%)
IDACI quintile 4	107,017 (20%)	18,219 (16%)	31,339 (17%)	75,678 (22%)
IDACI quintile 5 (least deprived)	107,580 (20%)	14,278 (12%)	26,977 (15%)	80,603 (23%)

Note: Cell entries are number (%) of pupils. Percentages may not sum to 100 due to rounding.

Source: National Pupil Database 2022/2023 (School Census, KS4).

The proportion of NTP pupils and average tutoring hours per NTP pupil (ATH-NTP)⁹ amongst Year 6 pupils at schools that participated in the NTP is shown in Figure 3 below. Schools adopted a wide range of approaches, in terms of how much tutoring was delivered and whether tutoring was focused on a small number of pupils or spread more widely throughout Year 6. However, schools commonly delivered between 5 and 20 hours on average to a minority (10%-40%) of pupils.

The distribution of tutoring hours amongst Year 11 pupils at schools that participated in the NTP is shown in Figure 4 below. Similarly to amongst Year 6 pupils, the amount and spread of tutoring varies a lot between schools. Many schools opted to deliver an average of between 5 and 20 hours to a minority (0-40%) of pupils.

To make these plots consistent with the samples used in the dosage analysis, schools with an average tutoring hours amongst all pupils (ATH-all) of over 40 hours are not included. Partly because of this, there are few schools with an ATH-NTP of over 35 hours shown, although there are at least some schools in every ATH-NTP band (rows of the heatmap). This means that the inferences reported in the dosage analysis (see the next two chapters) are less reliable for higher ranges of tutoring hours per tutored pupil.

⁹ Tutoring hours are across all subjects and NTP routes (SLT/AM/TP) that pupils received during the academic year.

Figure 3: Proportion of Year 6 pupils receiving tutoring and average tutoring hours per NTP tutored pupil (ATH-NTP) amongst Year 6 pupils at schools taking part in the NTP. Cell counts indicate number of schools.

Average tutoring hours per NTP pupil, ATH-NTP	>75	9	13	16	22	11	0	0	0	0	0
	70-75	X	X	X	X	5	X	0	0	0	0
	65-70	X	0	X	4	X	X	0	0	0	0
	60-65	X	10	11	14	10	4	X	0	0	0
	55-60	0	6	5	8	10	5	9	0	0	0
	50-55	X	7	9	7	5	5	4	X	0	0
	45-50	4	16	10	8	11	19	10	5	X	0
	40-45	7	10	18	14	25	19	13	8	9	4
	35-40	15	19	27	20	20	22	22	14	10	33
	30-35	39	69	68	51	48	59	33	24	10	56
	25-30	22	54	62	65	73	54	44	25	34	76
	20-25	47	79	81	98	110	74	86	42	37	91
	15-20	180	327	381	313	302	226	189	130	89	257
	10-15	147	300	347	305	275	207	161	102	71	190
	5-10	154	284	314	299	207	202	125	83	53	137
	0-5	122	146	186	154	110	98	58	32	32	91
		0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
		% of pupils receiving NTP									

Note: cells counts of 1,2 or 3 schools have been suppressed ('X'), in line with DfE/ONS policies on statistical disclosure control.

Source: National Pupil Database 2022/2023 (School Census)

Figure 4: Proportion of Year 11 pupils receiving tutoring and average tutoring hours per NTP tutored pupil (ATH-NTP) amongst Year 11 pupils at schools taking part in the NTP. Cell counts indicate number of schools.

Average tutoring hours per NTP pupil, ATH-NTP	>75	11	6	X	X	0	0	0	0	0	0
	70-75	X	0	0	0	0	0	0	0	0	0
	65-70	X	X	X	X	X	X	0	0	0	0
	60-65	X	X	6	X	X	X	0	0	0	0
	55-60	X	X	0	0	X	X	0	0	0	0
	50-55	X	X	X	4	4	X	X	0	0	0
	45-50	8	X	X	X	X	0	X	X	0	0
	40-45	5	X	4	X	X	4	X	X	X	4
	35-40	8	5	X	X	4	X	X	5	4	22
	30-35	16	12	11	8	6	4	6	X	6	32
	25-30	19	12	14	7	9	12	X	9	12	36
	20-25	29	33	22	13	17	12	17	12	16	42
	15-20	72	61	52	40	38	39	31	27	30	77
	10-15	76	102	75	66	53	57	47	38	35	71
	5-10	101	85	89	78	60	56	61	41	39	34
	0-5	69	60	50	38	23	16	16	7	8	19
		0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
		% of pupils receiving NTP									

Note: cells counts of 1,2 or 3 schools have been suppressed ('X'), in line with DfE/ONS policies on statistical disclosure control.

Source: National Pupil Database 2022/2023 (School Census)

Pupil-level evaluation samples

The KS2 pupil-level sample consisted of 115,301 Year 6 pupils (Table 3) who received NTP tutoring in the previous year (2021-22). More than half of these pupils (53%) went on to receive further tutoring during Year 6 (2022-23). The sample had a high proportion of PP pupils (51%) and pupils from more deprived areas: 51% were from the most deprived 40% of areas. This might be expected, given that both characteristics are associated with tutoring receipt (see Table 2) and all pupils in this sample were previously selected for NTP tutoring.

The KS4 pupil-level sample consisted of 62,171 Year 11 pupils who received NTP tutoring in the previous year. Similarly to the Year 6 sample, 55% of pupils received further tutoring in Year 11 (2022-23). Again, there was a high proportion of PP pupils (42%) and pupils from more deprived areas (50% were from the most deprived 40% of areas) in the sample.

The proportion of PP pupils is slightly higher amongst Year 6 pupils that were tutored in 2022-23 (56%) compared to the whole Year 6 sample (51%). Similarly, 45% of Year 11 pupils tutored in 2022-23 were PP, compared to 42% in the whole sample. Pupils tutored in 2022-23 were also from slightly more deprived areas than the whole sample, for both the Year 6 and Year 11 samples. This means that PP pupils and pupils from more deprived areas were not only prioritised for tutoring in 2021-22, but also for further tutoring in 2022-23.

The remaining characteristics displayed in Table 3 (gender, PLA, SEN, EAL, and ethnicity) show no evidence of being associated with tutoring receipt in 2022-23, in either of the samples.

Table 4: Number and percentage of pupils in the pupil-level evaluation sample that received NTP tutoring in 2021-22 split by pupil characteristics and 2022-23 NTP selection.

	KS2 total in sample	KS2 selected for the NTP (%)	KS2 not selected for the NTP (%)	KS4 total in sample	KS4 selected for the NTP (%)	KS4 not selected for the NTP (%)
Total number of pupils	115,301 (100%)	60,660 (100%)	54,641 (100%)	62,171 (100%)	34,487 (100%)	27,684 (100%)
Male	57,054 (49%)	29,497 (49%)	27,557 (50%)	31,732 (51%)	17,612 (51%)	14,120 (51%)
Female	58,247 (51%)	31,163 (51%)	27,084 (50%)	30,439 (49%)	16,875 (49%)	13,564 (49%)
PP pupils	58,587 (51%)	33,779 (56%)	24,808 (45%)	26,342 (42%)	15,548 (45%)	10,794 (39%)

	KS2 total in sample	KS2 selected for the NTP (%)	KS2 not selected for the NTP (%)	KS4 total in sample	KS4 selected for the NTP (%)	KS4 not selected for the NTP (%)
Non-PP pupils	56,714 (49%)	26,881 (44%)	29,833 (55%)	35,829 (58%)	18,939 (55%)	16,890 (61%)
PLA pupils	35,300 (31%)	19,176 (32%)	16,124 (30%)	16,144 (26%)	9,058 (26%)	7,086 (26%)
Non-PLA pupils	80,001 (69%)	41,484 (68%)	38,517 (70%)	46,027 (74%)	25,429 (74%)	20,598 (74%)
PP and/or PLA	74,049 (64%)	41,463 (68%)	32,586 (60%)	34,584 (56%)	20,003 (58%)	14,581 (53%)
Pupils with SEN	29,179 (25%)	15,935 (26%)	13,244 (24%)	12,324 (20%)	6,874 (20%)	5,450 (20%)
Pupils without SEN	86,122 (75%)	44,725 (74%)	41,397 (76%)	49,847 (80%)	27,613 (80%)	22,234 (80%)
Pupils with EAL	23,229 (20%)	12,266 (20%)	10,963 (20%)	11,263 (18%)	6,532 (19%)	4,731 (17%)
Pupils without EAL	92,072 (80%)	48,394 (80%)	43,678 (80%)	50,908 (82%)	27,955 (81%)	22,953 (83%)
White	83,505 (73%)	43,619 (72%)	39,886 (74%)	43,261 (71%)	23,520 (69%)	19,741 (72%)
Black	7,866 (7%)	4,528 (8%)	3,338 (6%)	5,020 (8%)	3,132 (9%)	1,888 (7%)
Chinese	282 (<1%)	156 (<1%)	126 (<1%)	161 (<1%)	87 (<1%)	74 (<1%)
Other Asian	11,883 (10%)	6,035 (10%)	5,848 (11%)	7,186 (12%)	3,960 (12%)	3,226 (12%)
Mixed ethnicity	8,129 (7%)	4,373 (7%)	3,756 (7%)	4,107 (7%)	2,373 (7%)	1,734 (6%)
Other ethnicity	2,649 (2%)	1,465 (2%)	1,184 (2%)	1,396 (2%)	819 (2%)	577 (2%)
IDACI quintile 1 (most deprived)	32,735 (28%)	18,654 (31%)	14,081 (26%)	17,080 (27%)	10,506 (30%)	6,574 (24%)
IDACI quintile 2	27,018 (23%)	14,660 (24%)	12,358 (23%)	14,606 (23%)	8,386 (24%)	6,220 (22%)
IDACI quintile 3	22,263 (19%)	11,386 (19%)	10,877 (20%)	12,095 (19%)	6,495 (19%)	5,600 (20%)

	KS2 total in sample	KS2 selected for the NTP (%)	KS2 not selected for the NTP (%)	KS4 total in sample	KS4 selected for the NTP (%)	KS4 not selected for the NTP (%)
IDACI quintile 4	18,767 (16%)	9,334 (15%)	9,433 (17%)	10,170 (16%)	5,186 (15%)	4,984 (18%)
IDACI quintile 5 (least deprived)	14,518 (13%)	6,626 (11%)	7,892 (14%)	8,220 (13%)	3,914 (11%)	4,306 (16%)

Note: Percentages may not sum to 100 due to rounding.

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4), Randstad tuition participation data.

Main results maths: What was the impact of the NTP on maths outcomes in 2022-23?

Key Findings

- We found evidence that participation in the NTP was associated with improved maths outcomes at KS2 for all pupils, PP pupils and PLA pupils, which were equivalent to around 1 month's additional progress.
- The effect sizes seen for maths outcomes at KS4 were smaller and in most cases were not statistically significant.
- There is some indication at both KS2 and KS4 that PLA pupils may have benefited slightly more from tutoring than other pupils.
- The optimum average maths tutoring per NTP pupil was estimated at 31 hours, although there is a lot of uncertainty around this estimate. For KS2 maths, the optimum average tutoring hours per NTP pupil was initially estimated to be 46 hours. Due to lack of subject information, we do not know exactly how many of these tutoring hours were in maths: we assumed 68% based on the split of tutoring provided via the AM/TP routes last year, to reach the estimate of 31 hours.
- At KS4, we did not find any relationship between average tutoring hours per NTP pupil and maths attainment. This could be indicative of a different dosage dependence at KS4, but it could also be explained by the fact there are fewer schools in KS4 so the analysis may not have enough data to detect an effect.
- We did not detect any differences in impact between different geographic regions.

This chapter describes the results of the analysis exploring the school and pupil-level impacts of the NTP on KS2 and KS4 maths outcomes and includes the pupil-level sensitivity analysis. It also investigates how the impact of the NTP on maths outcomes varies according to tutoring dosage (specifically how best to distribute tutoring hours at a school level). The final section explores variation by geographic region. We anticipate that the school and pupil-level analyses should complement one another, with the school-level analysis giving the most reliable indication of the direction of the effect and, assuming the sensitivity analysis does not indicate lingering selection bias, the pupil-level analysis the most reliable indication of the effect size at pupil level.

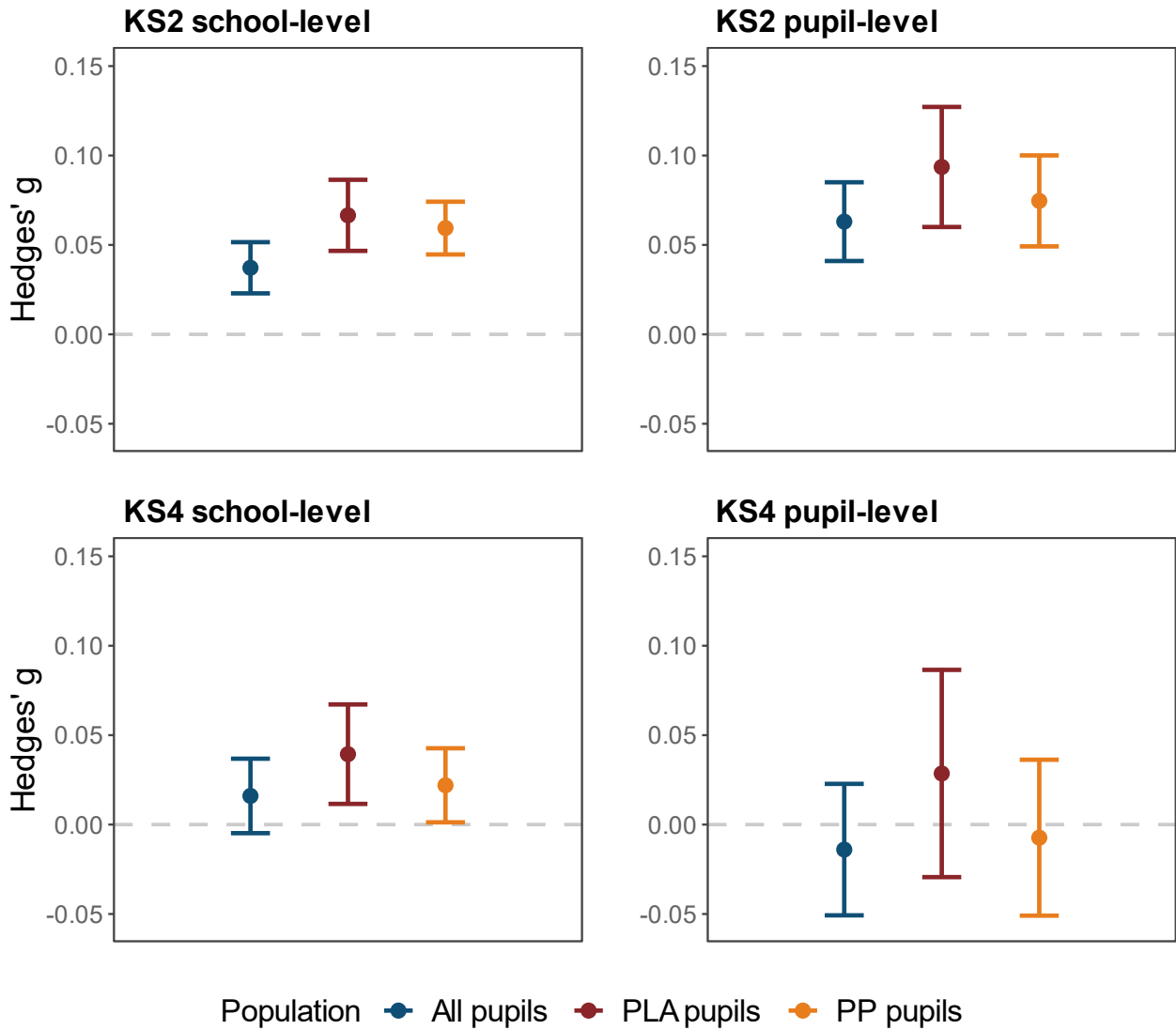
What was the impact of the NTP on maths outcomes for all pupils, PP pupils and PLA pupils?

Our results indicate that participation in the NTP was associated with small improvements in maths outcomes at KS2: see Figure 5 and Table 4 for full statistical details, including effect sizes and confidence intervals. Results were similar between the school-level and pupil-level approaches, although the pupil-level effect sizes were slightly larger. All KS2 results were highly statistically significant, meaning that tutoring is likely to have benefitted KS2 maths attainment. There is also some indication that PLA pupils may have benefitted more from tutoring than Year 6 pupils in general.

The impact of NTP tutoring seen at KS4 was lower than that at KS2; effect sizes were consistent with a very small positive effect or no effect. The KS4 school-level analysis indicated a small positive impact, although amongst all pupils the result was not statistically significant. Effect sizes for the KS4 pupil-level analysis were close to zero and not statistically significant. Year 11 PLA pupils appear to have benefitted more from tutoring slightly more than their peers, similarly to the results at KS2.

Overall, the effect sizes seen in these analyses were small or very small, ranging from -0.014 in the KS4 pupil-level analysis (all pupils) to 0.094 in the KS2 pupil-level analysis (PLA pupils). Using the EEF scale of effect sizes these effects would be equivalent to 1 months' additional progress or less (EEF, 2023).

Figure 5: The impact of the NTP on KS2 and KS4 maths outcomes.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4), Randstad tuition participation data.

Table 5: The impact of the NTP on KS2 and KS4 maths outcomes; complete results from the linear mixed effects models

Outcome	Sample	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	All pupils	0.28 (0.17, 0.39)	0.037 (0.023, 0.052)	0	0.000
School-level KS2	PLA pupils	0.46 (0.32, 0.6)	0.067 (0.047, 0.086)	1	0.000
School-level KS2	PP pupils	0.46 (0.35, 0.58)	0.059 (0.045, 0.074)	1	0.000
Pupil-level KS2	All pupils	0.43 (0.28, 0.58)	0.063 (0.041, 0.085)	1	0.000
Pupil-level KS2	PLA pupils	0.62 (0.4, 0.84)	0.094 (0.06, 0.127)	1	0.000
Pupil-level KS2	PP pupils	0.53 (0.35, 0.71)	0.075 (0.049, 0.1)	1	0.000
School-level KS4	All pupils	0.03 (-0.01, 0.07)	0.016 (-0.005, 0.037)	0	0.132
School-level KS4	PLA pupils	0.05 (0.02, 0.09)	0.039 (0.012, 0.067)	0	0.006
School-level KS4	PP pupils	0.04 (0, 0.08)	0.022 (0.001, 0.043)	0	0.038
Pupil-level KS4	All pupils	-0.03 (-0.09, 0.04)	-0.014 (-0.051, 0.023)	0	0.456
Pupil-level KS4	PLA pupils	0.04 (-0.04, 0.11)	0.029 (-0.029, 0.087)	0	0.334
Pupil-level KS4	PP pupils	-0.01 (-0.09, 0.06)	-0.007 (-0.051, 0.036)	0	0.742

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4), Randstad tuition participation data.

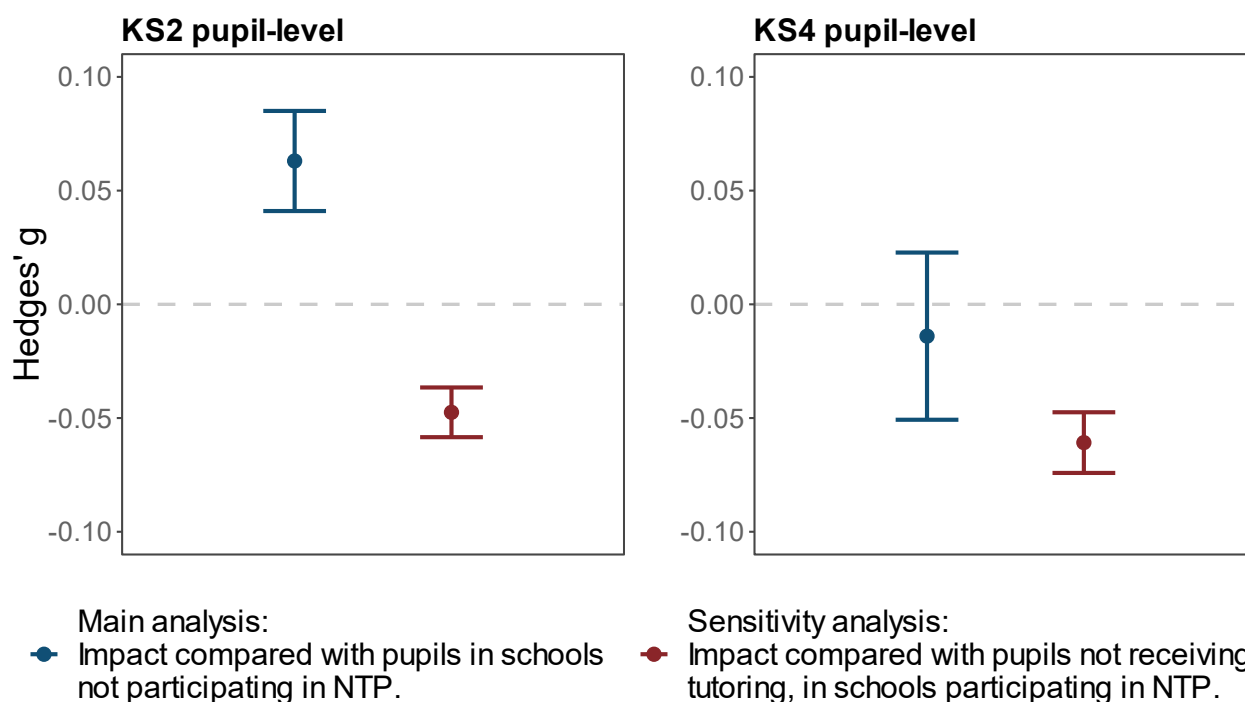
Before sharing the results from the pupil-level sensitivity analysis, we present a recap of who is included in the pupil-level main and sensitivity analysis in Table 5 below. The difference between the two analyses is whether comparison pupils attended a school that participated in NTP Year 3 (see the methods section for details).

Table 6: Summary of which pupils are in the intervention and comparison groups for the pupil-level main and sensitivity analysis

	Pupil-level analysis	Were pupils tutored in NTP Year 2?	Were pupils tutored in NTP Year 3?	Did pupil's school participate in NTP Year 3?
Intervention group	Main	Yes	Yes	Yes
Intervention group	Sensitivity	Yes	Yes	Yes
Comparison group	Main	Yes	No	No
Comparison group	Sensitivity	Yes	No	Yes

Results from the pupil-level sensitivity analysis are shown in Figure 6 below. On these plots the blue bars show the impact of the NTP when the comparison group is pupils attending schools that did not take part in the NTP (i.e. simply the main pupil-level effect sizes from Table 4). The red bars show the impact when pupils in the comparison group were not themselves selected for NTP in 2022-23, but their school did take part (the sensitivity analysis). For the sensitivity analyses the impacts are substantially more negative than for the main pupil-level analyses, especially at KS2. This result is consistent with schools selecting pupils for tutoring who are most in need of support. It means that although restricting the sample of the pupil-level analysis to pupils who participated in the NTP in 2021-22 has helped to reduce selection bias, some lingering selection bias is still present within this analysis. All effect sizes from the main pupil-level analyses are therefore likely to be negatively biased, meaning the true positive impact is higher than that estimated in this report. This may explain why the effect sizes in the pupil level analysis are not much larger than those seen in the school-level analysis despite the school-level analysis being subject to significant dilution.

Figure 6: Pupil level sensitivity analysis for KS2 and KS4 maths outcomes.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4), Randstad tuition participation data.

The difference between the main and sensitivity analysis estimates provides some insight into the degree of bias in the main analysis: a greater difference is likely to indicate more bias. However, it is not possible to quantify the exact degree of negative bias in the main analysis from Figure 6 (e.g. it is not simply equal to the sensitivity analysis estimate) as the degree of bias in the main analysis depends on how many comparison pupils would have been selected for tutoring had it been available at their school, which is an unknown.

How does the impact of the NTP on maths outcomes vary according to tutoring dosage?

The dosage analysis investigated the impact of how a school spreads their fixed budget of tutoring hours: smaller amounts of tutoring for many pupils or more tutoring for fewer pupils. Unlike the previous section, this analysis included all schools that took part in the NTP in 2022/23 (regardless of how many pupils were tutored) and excluded schools that did not take part.

We now recap two important quantities in the dosage analysis, which were introduced in the methods section:

ATH-NTP, the average tutoring hours amongst pupils that received NTP tutoring at a school.

ATH-all, the average tutoring hours amongst all pupils at a school.

In this analysis higher ATH-NTP means focusing more tutoring hours on a smaller number of pupils. This is because we have 'fixed' (conditioned on) a school's tutoring budget (given by ATH-all), so ATH-NTP can only be increased by tutoring fewer pupils. The analysis excluded schools with a very high tutoring budget (ATH-all > 40).

At KS2 it was decided via statistical testing that the relationship between ATH-NTP and KS2 maths attainment was quadratic (see technical appendix). This means focusing a fixed tutoring budget on fewer pupils improved maths attainment up to a point of maximum benefit and after that point attainment declined (relative to the maximum) as illustrated in Figure 7. Note this does not mean more tutoring made attainment worse after the maximum: it means further focusing a school's fixed tutoring budget on fewer pupils made attainment worse (on average, across all pupils) after this point. The 'optimum' value of ATH-NTP (producing the maximum of the maths attainment curve) occurred at 46 hours. 0.67 of a standard score point was gained on average amongst pupils¹⁰ at a school moving from 1 to 46 hours¹¹.

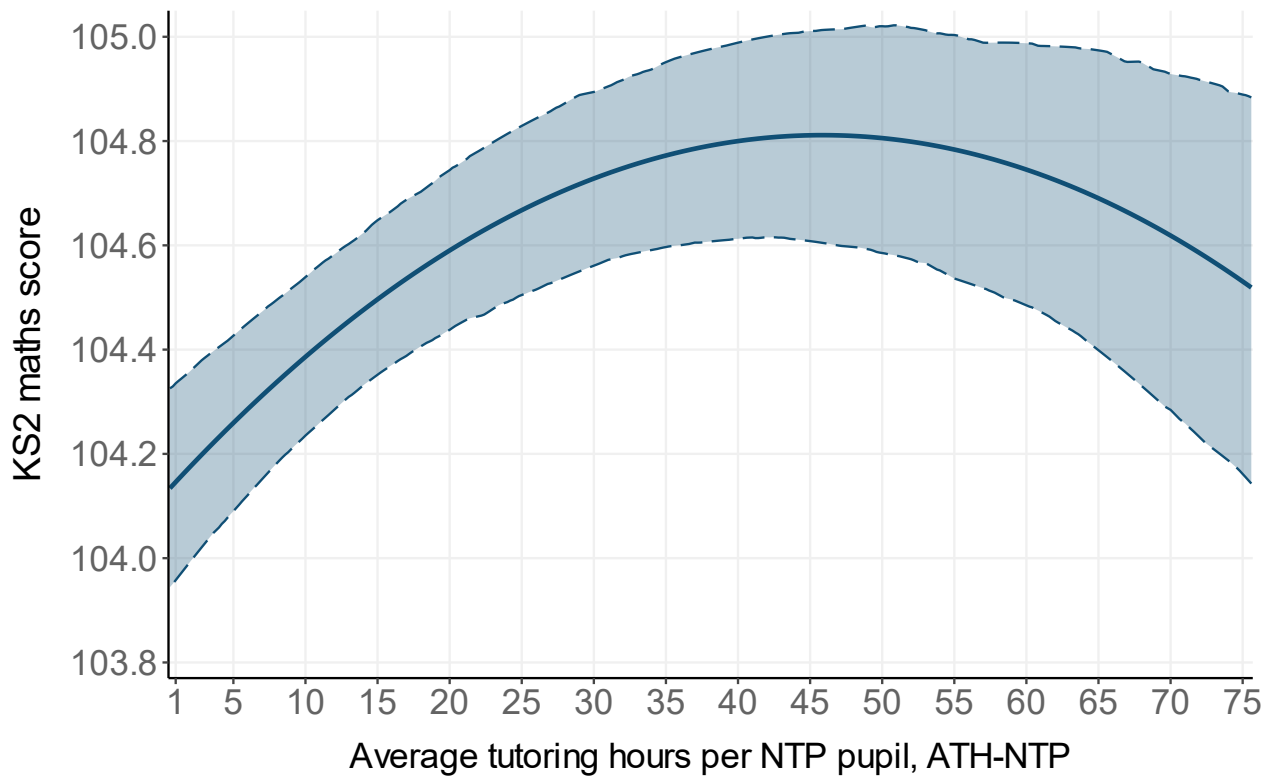
As no subject information was available for a pupil's tutoring hours, this maximum estimated above is likely to be an overestimate for maths tutoring specifically. Based on the split of tutoring provided via the AM/TP routes in the second year of NTP, we assume that 68% of the tuition provided is in maths, this would suggest 31 hours of tutoring are needed for optimum impact. There is uncertainty around this estimate arising from multiple sources, including sampling uncertainty and lack of tutoring subject information. We suggest the actual optimum is likely to lie above 20 hours, so schools could consider focusing their maths tutoring budget on fewer pupils if they are beneath this number. This is proposed as a 'rule of thumb', as it is not possible to statistically derive an interval (e.g. confidence interval) for the optimum ATH-NTP.

At KS4 no relationship (linear or quadratic) was found between ATH-NTP and maths attainment. This study therefore finds no evidence that focusing tutoring on fewer pupils has any impact (positive or negative) on KS4 maths attainment. This could be indicative of a different dosage dependence at KS4, but it could also be explained by the fact there are fewer schools in KS4 so the analysis may not have enough data to detect a statistically significant linear or quadratic effect – in other words, an effect may exist but this analysis was unable to detect it.

¹⁰ Not all individual pupils would benefit: Year 6 pupils that do not receive tutoring, so that other pupils can receive more hours, may have lowered attainment. It is a 'cost-benefit' trade-off, with overall pupil attainment increasing.

¹¹ This is the maximum benefit that might be expected; schools would generally see less benefit than this if they focused tuition on fewer pupils, because their ATH-NTP is already part way between 1 and 46 hours.

Figure 7: Pupil KS2 maths scores determined by average tutoring hours delivered at their school amongst NTP pupils (ATH-NTP).



Note 1: predicted scores are obtained from the fitted model by fixing continuous covariates at their mean and categorical variables at their mode, then creating predictions for values of ATH-NTP between 0 and 75.

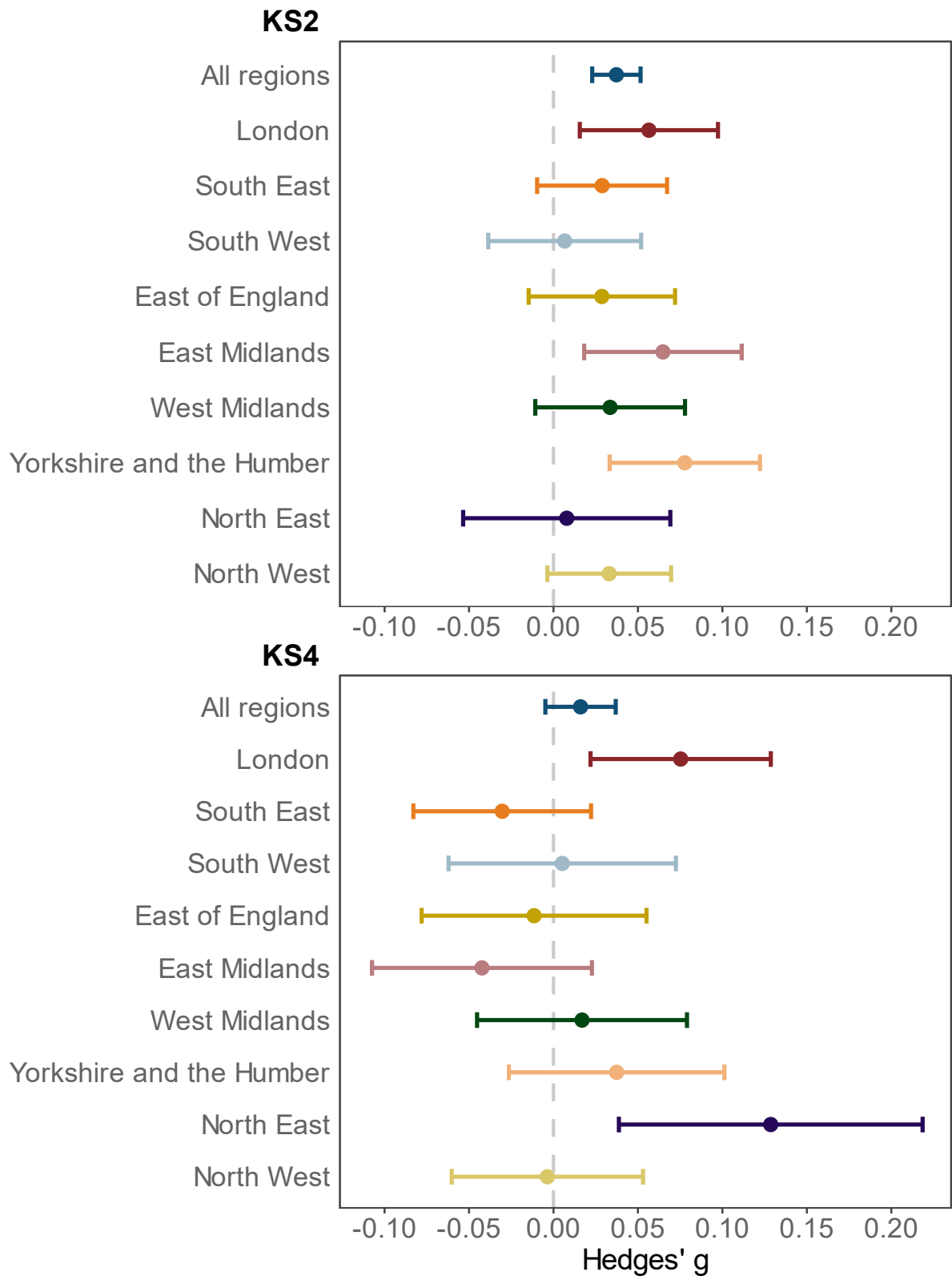
Note 2: the shaded region around the line represents a 95% confidence interval for the KS2 maths score of a pupil at each point along the line.

Source: National Pupil Database 2022/2023 (School Census, KS2)

How does the impact of the NTP on maths outcomes vary according to geographic region?

The impact of the NTP on maths attainment within each of the nine geographic regions in England is illustrated in Figure 8 below, with all statistical results reported in Table 6. Although effect sizes appear to vary considerably from region to region, confidence intervals around these effect sizes are wide (especially at KS4) and generally overlap with the confidence interval for all regions in England (i.e. the main school-level result). This means that observed differences in the effect of the NTP between regions are likely to be due to chance.

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



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 7: The impact of the NTP on maths outcomes at school level by region; complete results from the linear mixed effects models

Outcome	Region	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	London	0.43 (0.12, 0.74)	0.056 (0.016, 0.097)	1	0.007
School-level KS2	South East	0.22 (-0.07, 0.51)	0.029 (-0.01, 0.067)	0	0.143
School-level KS2	South West	0.05 (-0.29, 0.39)	0.007 (-0.039, 0.052)	0	0.774
School-level KS2	East of England	0.22 (-0.11, 0.55)	0.029 (-0.015, 0.072)	0	0.196
School-level KS2	East Midlands	0.48 (0.14, 0.83)	0.065 (0.018, 0.111)	1	0.007
School-level KS2	West Midlands	0.26 (-0.08, 0.59)	0.034 (-0.011, 0.078)	0	0.139
School-level KS2	Yorkshire and the Humber	0.6 (0.25, 0.94)	0.078 (0.033, 0.122)	1	0.001
School-level KS2	North East	0.06 (-0.39, 0.5)	0.008 (-0.053, 0.069)	0	0.802
School-level KS2	North West	0.25 (-0.03, 0.52)	0.033 (-0.004, 0.07)	0	0.078
School-level KS4	London	0.16 (0.05, 0.27)	0.075 (0.022, 0.129)	1	0.006
School-level KS4	South East	-0.06 (-0.17, 0.05)	-0.03 (-0.083, 0.022)	0	0.259
School-level KS4	South West	0.01 (-0.12, 0.14)	0.005 (-0.062, 0.072)	0	0.880
School-level KS4	East of England	-0.02 (-0.15, 0.11)	-0.011 (-0.078, 0.055)	0	0.736
School-level KS4	East Midlands	-0.08 (-0.2, 0.04)	-0.042 (-0.107, 0.023)	0	0.206

Outcome	Region	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS4	West Midlands	0.03 (-0.09, 0.16)	0.017 (-0.045, 0.079)	0	0.595
School-level KS4	Yorkshire and the Humber	0.07 (-0.05, 0.2)	0.037 (-0.026, 0.101)	0	0.253
School-level KS4	North East	0.24 (0.07, 0.41)	0.129 (0.039, 0.219)	2	0.007
School-level KS4	North West	-0.01 (-0.12, 0.1)	-0.004 (-0.06, 0.053)	0	0.901

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Main results English: What was the impact of the NTP on English outcomes in 2022-23

Key Findings

- We found evidence that participation in the NTP was associated with small statistically significant improvements in English outcomes at KS2 for PP pupils and PLA pupils, but not all pupils. As with maths, smaller effects were seen for English outcomes at KS4 and only the school-level results reached statistical significance. In all cases the effect sizes equated to less than one month's additional progress.
- As with maths outcomes, there is some indication at both KS2 and KS4 that PLA pupils may have benefited slightly more from tutoring than other pupils in terms of English outcomes.
- The optimum average English tutoring per NTP pupil was estimated at 20 hours, although there is a lot of uncertainty around this estimate. For KS2 reading, the optimum average tutoring hours per NTP pupil was initially estimated to be 63 hours. Due to lack of subject information, we do not know exactly how many of these tutoring hours were in English: we assumed 32% based on the split of tutoring provided via the AM/TP routes last year, to reach the estimate of 20 hours.
- At KS4, we did not find any relationship (linear or quadratic) between average tutoring hours per NTP pupil and English attainment. This could be indicative of a different dosage dependence at KS4, but it could also be explained by the fact there are fewer schools in KS4 so the analysis may not have enough data to detect a statistically significant quadratic effect.
- We did not detect any differences in impact between different geographic regions.

This chapter describes the results of the analysis exploring the school and pupil-level impacts of the NTP on KS2 and KS4 English outcomes and includes the pupil-level sensitivity analysis. As with Chapter 4, it also investigates how the impact of the NTP on English outcomes varies according to tutoring dosage and geographic region.

What was the impact of the NTP on English outcomes for all pupils, PP pupils and PLA pupils?

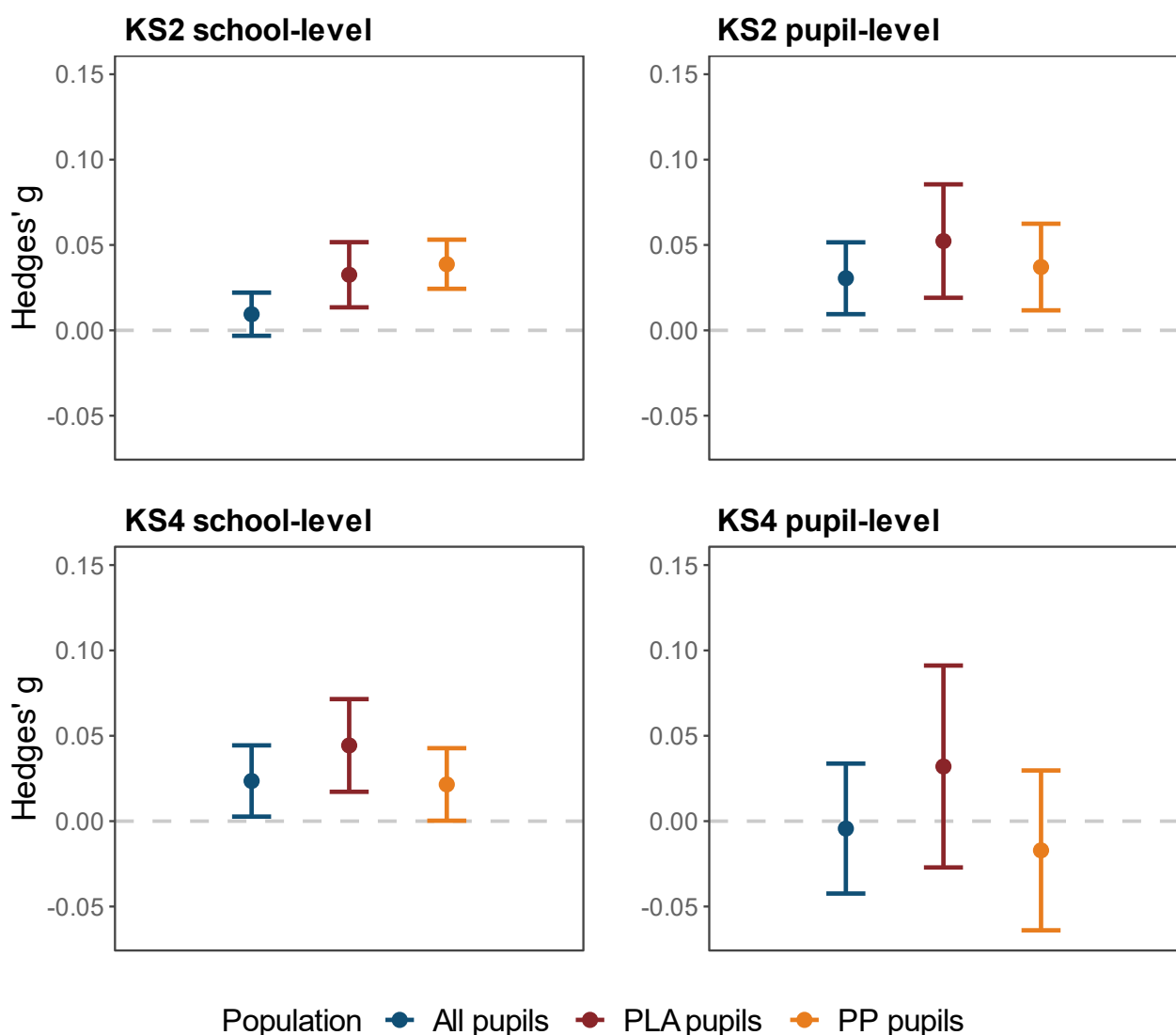
Our results indicate that participation in the NTP was associated with small improvements in English outcomes (see Figure 9 and Table 6 for full statistical details including effect sizes and confidence intervals) at both KS2 and KS4. The improvements seen in English outcomes were similar in KS2 and KS4. As with maths (see Chapter 4),

there is some indication that PLA pupils may benefit more from tutoring in terms of English outcomes compared with all pupils and PP pupils.

Overall, the effect sizes were very small, ranging from 0.004 for KS4 pupil-level All pupils to 0.052 for KS2 pupil-level PLA pupils. Using the EEF scale of effect sizes these effects would be equivalent to less than one months' additional progress (EEF, 2023).

The effect sizes for the KS2 school-level analysis were statistically significant for PP and PLA pupils, and for all three pupil groups in the KS2 pupil-level analysis. At KS4 the school-level effects were statistically significant. By contrast, the effect sizes for the KS4 pupil-level analysis were close to zero and not statistically significant.

Figure 9: The impact of the NTP on English outcomes at KS2 and KS4.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 8: The impact of the NTP on English outcomes at KS2 and KS4; complete results from the linear mixed effects models

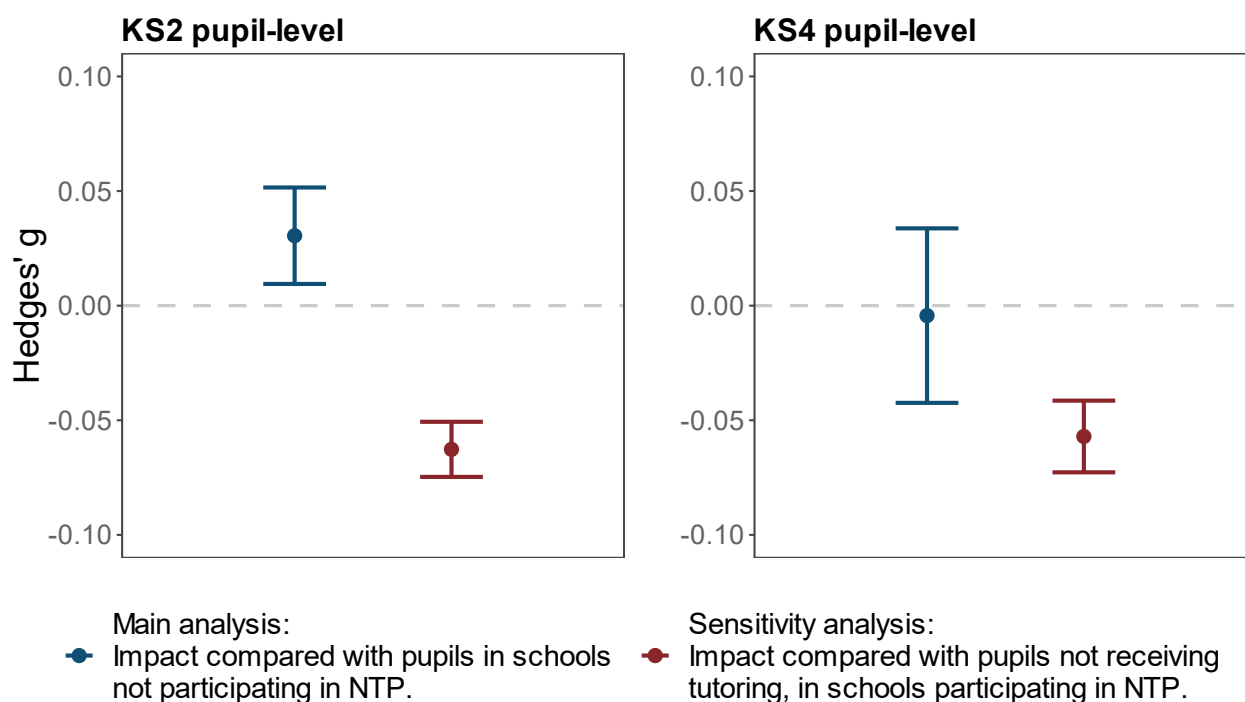
Outcome	Sample	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	All pupils	0.07 (-0.03, 0.17)	0.009 (-0.003, 0.022)	0	0.144
School-level KS2	PLA pupils	0.23 (0.1, 0.37)	0.033 (0.013, 0.052)	0	0.001
School-level KS2	PP pupils	0.30 (0.19, 0.42)	0.039 (0.024, 0.053)	0	0.000
Pupil-level KS2	All pupils	0.22 (0.07, 0.37)	0.031 (0.009, 0.052)	0	0.004
Pupil-level KS2	PLA pupils	0.36 (0.13, 0.58)	0.052 (0.019, 0.085)	0	0.002
Pupil-level KS2	PP pupils	0.27 (0.09, 0.46)	0.037 (0.012, 0.062)	0	0.004
School-level KS4	All pupils	0.04 (0, 0.08)	0.024 (0.003, 0.044)	0	0.027
School-level KS4	PLA pupils	0.06 (0.02, 0.09)	0.044 (0.017, 0.071)	0	0.001
School-level KS4	PP pupils	0.04 (0, 0.08)	0.022 (0, 0.043)	0	0.047
Pupil-level KS4	All pupils	-0.01 (-0.07, 0.06)	-0.004 (-0.042, 0.034)	0	0.823
Pupil-level KS4	PLA pupils	0.04 (-0.03, 0.11)	0.032 (-0.027, 0.091)	0	0.289
Pupil-level KS4	PP pupils	-0.03 (-0.11, 0.05)	-0.017 (-0.064, 0.03)	0	0.474

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Results of the pupil-level sensitivity analysis for English are like those seen for maths. They show small negative effects sizes for the impact of tutoring on English outcomes (Figure 10), as opposed to the small positive effects seen in the main pupil-level analysis (Figure 9). Again, these results are consistent with lingering selection bias being present in the analysis despite us restricting the pupil-level analysis to pupils who had previously been selected for tutoring in the second year of the NTP. However, it is not possible to

quantify the exact degree of bias in the main pupil-level analysis from Figure 10, as explained in the maths sensitivity analysis.

Figure 10: Pupil-level sensitivity analysis for KS2 and KS4 English outcomes.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4), Randstad tuition participation data.

How does the impact of the NTP on English outcomes vary according to tutoring dosage?

At KS2 it was observed through statistical testing that the relationship between ‘average tutoring hours per NTP pupil’ (ATH-NTP) and KS2 reading attainment could be approximated by a quadratic function (see technical appendix), similar to the relationship for KS2 maths attainment. Again, this means focusing a fixed tutoring budget on fewer pupils improves reading attainment up to a point of maximum benefit and after that point attainment declines, as illustrated in Figure 11. The maximum of the reading attainment curve occurred at 63 hours, with 0.62 of a standard score point gained on average amongst pupils¹² at a school moving from 1 to 63 hours¹³. Again, as no subject information was available for a pupil’s tutoring hours, this maximum is likely to be an

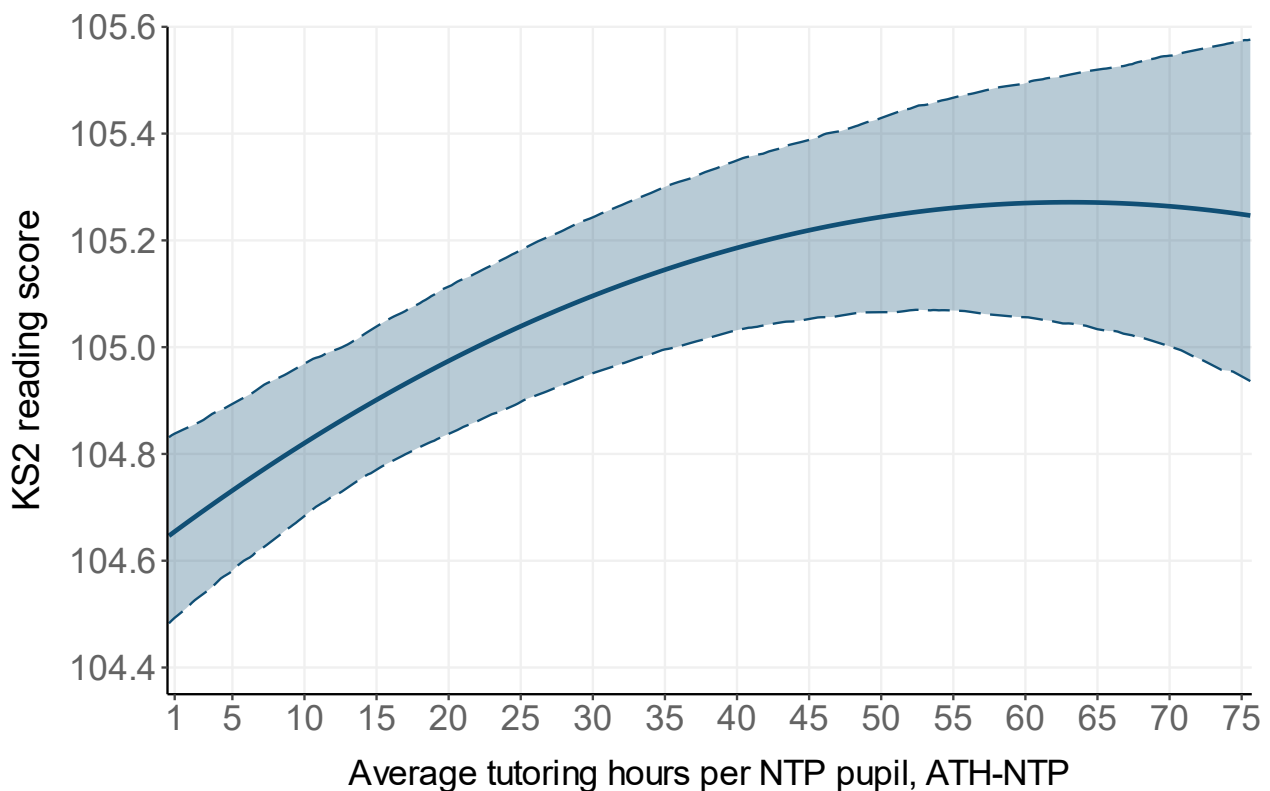
¹² Not all individual pupils would benefit: Year 6 pupils that do not receive tutoring, so that other pupils can receive more hours, may have lowered attainment. It is a ‘cost-benefit’ trade-off, with overall pupil attainment increasing.

¹³ This is the maximum benefit that might be expected; schools would generally see less benefit than this if they focused tuition on fewer pupils, because their ‘average tutoring hours per NTP pupil’ is already part way between 1 and 63 hours.

overestimate for reading tutoring specifically. If based on the split of tutoring provided via the AM/TP routes last year and we assume that 32% of the tuition provided is in English, then this would suggest 20 hours of tuition are needed for optimum impact. There is clearly uncertainty around this estimate. We suggest the actual optimum ATH-NTP is likely to lie above 15 hours (as a 'rule of thumb'), so schools could consider focusing their English tutoring budget on fewer pupils if they are beneath this number.

Reflecting the dosage analysis for maths, no relationship (linear or quadratic) was found between ATH-NTP and English Language attainment for KS4. This study therefore finds no evidence that focusing tutoring on fewer pupils has any impact (positive or negative) on KS4 English Language attainment. Again, this could be indicative of a different dosage dependence at KS4, but it could also be explained by the fact there are fewer schools in KS4.

Figure 11: Pupil KS2 reading scores determined by average tutoring hours delivered at their school amongst NTP pupils (ATH-NTP).



Note 1: predicted scores are obtained from the fitted model by fixing continuous covariates at their mean and categorical variables at their mode, then creating predictions for values of ATH-NTP between 0 and 75.

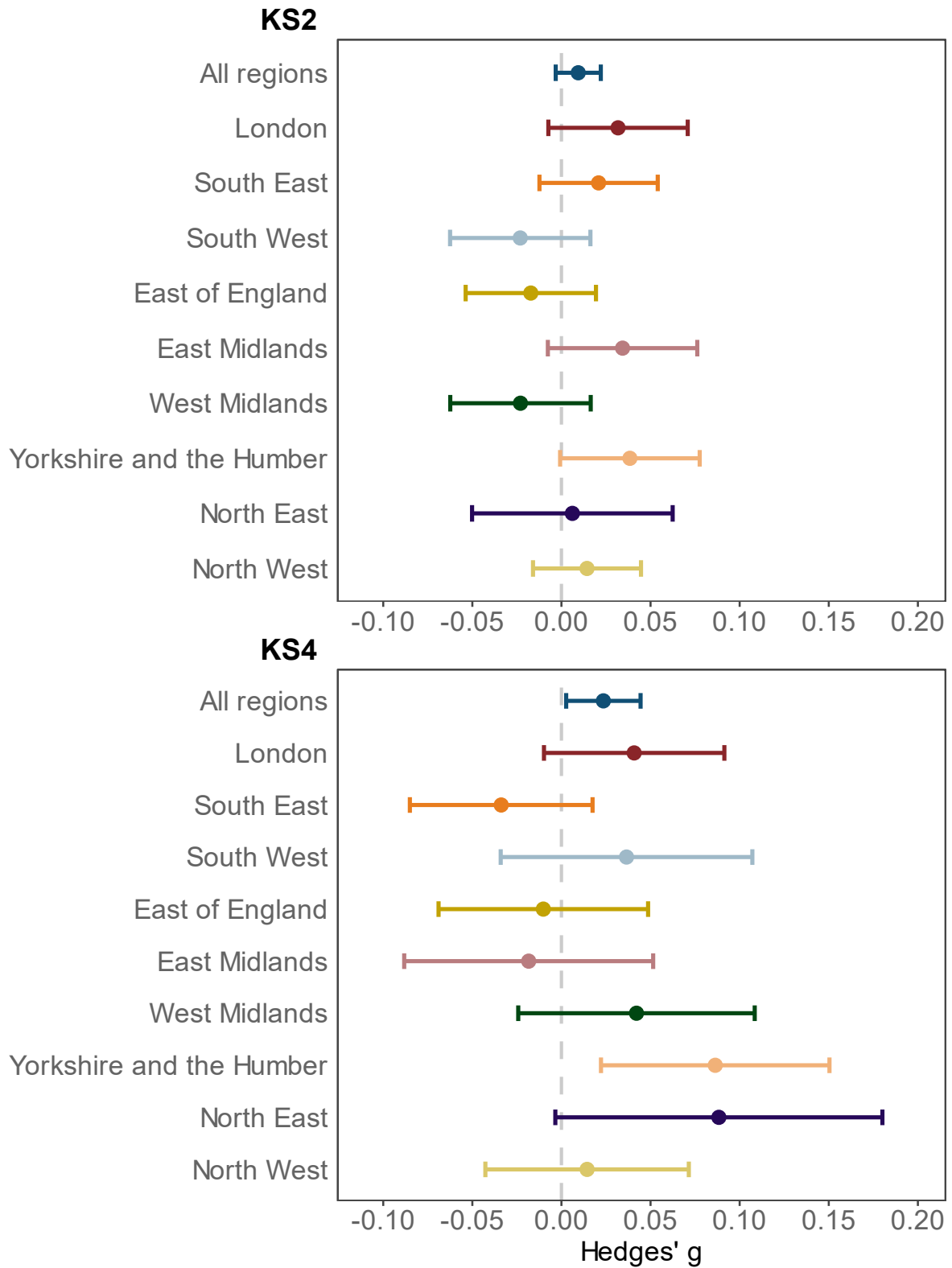
Note 2: the shaded region around the line represents a 95% confidence interval for the KS2 reading score of a pupil at each point along the line.

Source: National Pupil Database 2022/2023 (School Census, KS2)

How does the impact of the NTP on English outcomes vary according to geographic region?

The impact of the NTP on reading attainment within each of the nine geographic regions in England is illustrated in Figure 12 below, with all statistical results reported in Table 8. Although effects sizes appear to vary considerably from region to region, confidence intervals around these effect sizes are wide (especially at KS4) and generally overlap with the confidence interval for all regions in England (i.e. the main school-level result). This means that the observed differences in the effect of the NTP on English attainment between regions are likely to be due to chance.

Figure 12: The impact of the NTP on English outcomes at school level by region.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 9: The impact of the NTP on English outcomes at school level by region; complete results from the linear mixed effects models.

Outcome	Region	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	London	0.24 (-0.06, 0.54)	0.032 (-0.007, 0.071)	0	0.112
School-level KS2	South East	0.16 (-0.1, 0.42)	0.021 (-0.012, 0.054)	0	0.219
School-level KS2	South West	-0.18 (-0.5, 0.13)	-0.023 (-0.062, 0.016)	0	0.250
School-level KS2	East of England	-0.13 (-0.42, 0.15)	-0.017 (-0.054, 0.019)	0	0.357
School-level KS2	East Midlands	0.27 (-0.06, 0.59)	0.034 (-0.008, 0.076)	0	0.109
School-level KS2	West Midlands	-0.18 (-0.48, 0.13)	-0.023 (-0.062, 0.016)	0	0.253
School-level KS2	Yorkshire and the Humber	0.3 (-0.01, 0.61)	0.038 (-0.001, 0.078)	0	0.055
School-level KS2	North East	0.05 (-0.38, 0.48)	0.006 (-0.05, 0.062)	0	0.831
School-level KS2	North West	0.11 (-0.12, 0.34)	0.014 (-0.016, 0.045)	0	0.354
School-level KS4	London	0.08 (-0.02, 0.17)	0.041 (-0.01, 0.091)	0	0.116
School-level KS4	South East	-0.06 (-0.16, 0.03)	-0.034 (-0.085, 0.017)	0	0.198
School-level KS4	South West	0.07 (-0.06, 0.2)	0.037 (-0.034, 0.107)	0	0.313
School-level KS4	East of England	-0.02 (-0.12, 0.09)	-0.01 (-0.069, 0.049)	0	0.735
School-level KS4	East Midlands	-0.03 (-0.16, 0.09)	-0.018 (-0.088, 0.051)	0	0.607
School-level KS4	West Midlands	0.08 (-0.04, 0.2)	0.042 (-0.024, 0.108)	0	0.216

Outcome	Region	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS4	Yorkshire and the Humber	0.16 (0.04, 0.28)	0.086 (0.022, 0.15)	1	0.009
School-level KS4	North East	0.16 (-0.01, 0.33)	0.088 (-0.003, 0.18)	1	0.064
School-level KS4	North West	0.03 (-0.08, 0.13)	0.014 (-0.043, 0.071)	0	0.622

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Results by pupil characteristics: How does the school-level impact of the NTP on maths and English outcomes vary according to pupil characteristics?

Key Findings

- We did not detect any differences in impact for the NTP according to pupil characteristics for either maths or English at KS2 or KS4. This suggests that different pupil groups benefitted similarly from participating in the NTP.

This chapter 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 each model the comparison is between a group of pupils with a certain characteristic (e.g. males) at intervention schools and pupils from that same group at comparison schools.

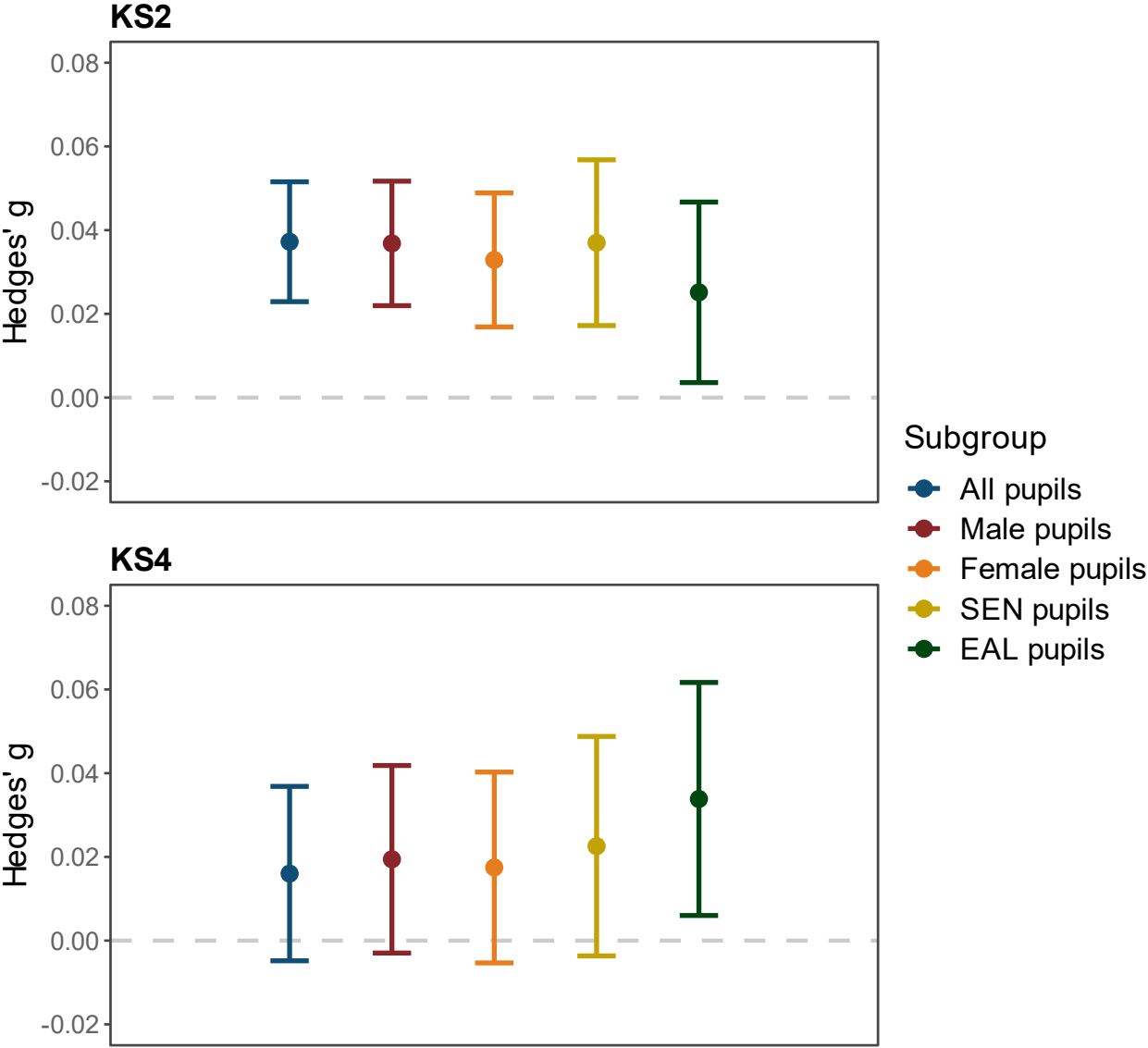
How does the school-level impact of the NTP on maths outcomes vary according to pupil characteristics?

Figure 13 below shows the school-level impact of the NTP on maths attainment for several subgroups of pupils: males, females, SEN pupils and EAL pupils (all statistical results are shown in Table 9). The main school-level impact estimates for all pupils are included as a point of reference. At KS2 effect sizes were small and positive for all subgroups of Year 6 pupils. All results were statistically significant at KS2, meaning there was evidence of a positive impact of tutoring in each subgroup. There is no evidence that this impact was different between subgroups: effect sizes were similar and confidence intervals largely overlapped.

At KS4 very small but positive impacts on maths attainment were seen in all subgroups of Year 11 pupils. Results were not statistically significant in most subgroups¹⁴. Similarly to KS2, the impact of tutoring was similar in each subgroup and confidence intervals largely overlapped with the confidence interval for all pupils.

¹⁴ The result for EAL pupils was statistically significant, but this does not in itself indicate a greater benefit of tutoring for EAL pupils, as the confidence interval for EAL pupils still largely overlaps with those of other subgroups.

Figure 13: The impact of the NTP on maths outcomes at school level by pupil characteristics.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 10: The impact of the NTP on maths outcomes at school level by pupil characteristics; complete results from the linear mixed effects models

Outcome	Subgroup	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	Male Pupils	0.283 (0.17, 0.4)	0.037 (0.022, 0.052)	0	0.000
School-level KS2	Female Pupils	0.246 (0.13, 0.37)	0.033 (0.017, 0.049)	0	0.000
School-level KS2	SEN Pupils	0.307 (0.14, 0.47)	0.037 (0.017, 0.057)	0	0.000
School-level KS2	EAL Pupils	0.19 (0.03, 0.35)	0.025 (0.004, 0.047)	0	0.022
School-level KS4	Male Pupils	0.039 (-0.01, 0.08)	0.019 (-0.003, 0.042)	0	0.089
School-level KS4	Female Pupils	0.034 (-0.01, 0.08)	0.017 (-0.005, 0.04)	0	0.133
School-level KS4	SEN Pupils	0.043 (-0.01, 0.09)	0.023 (-0.004, 0.049)	0	0.092
School-level KS4	EAL Pupils	0.07 (0.01, 0.13)	0.034 (0.006, 0.062)	0	0.017

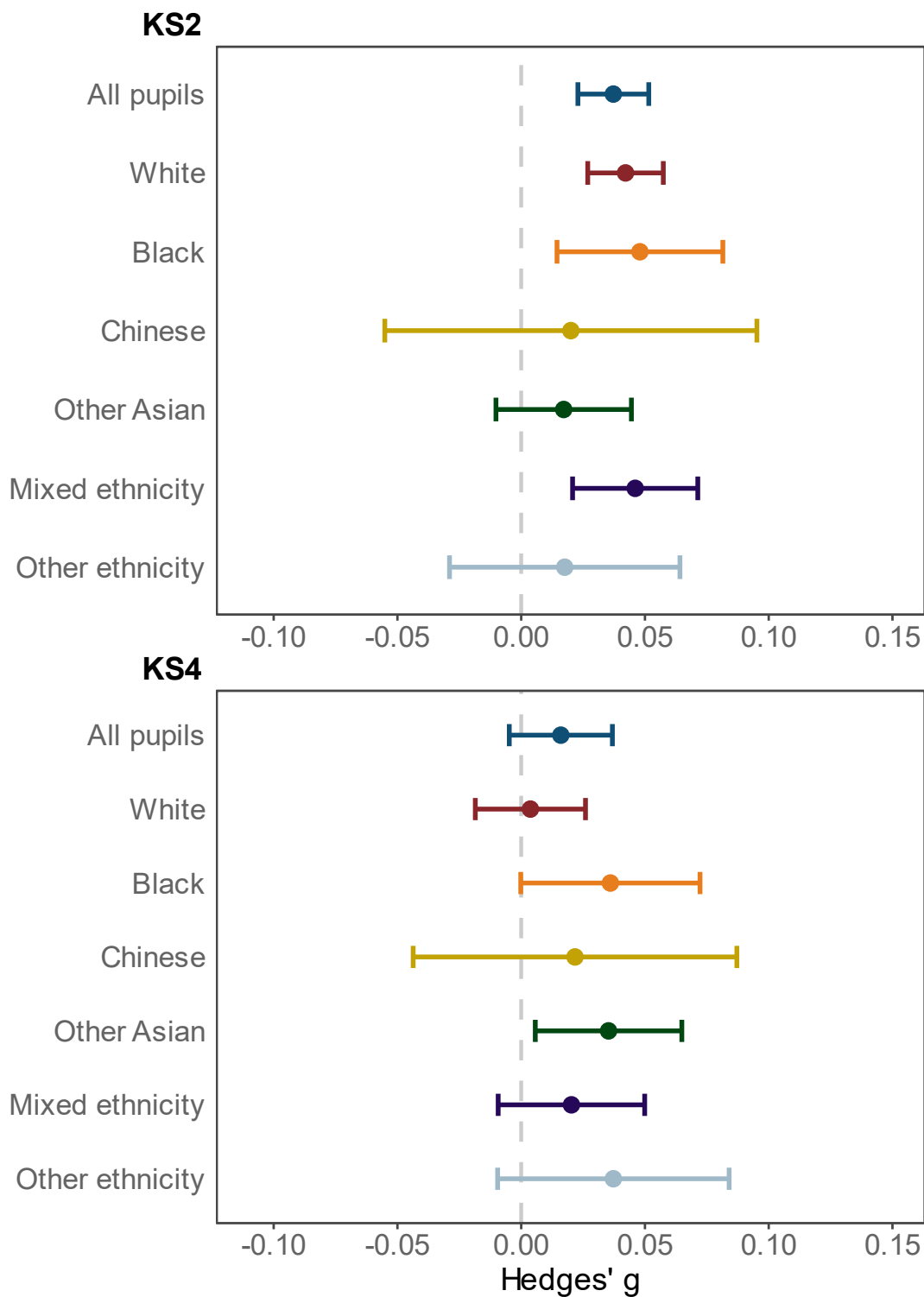
Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

The school-level impact of NTP tutoring on maths attainment amongst pupils of different ethnicities is shown in Figure 14 (all statistical results are shown in Table 10). Small positive effect sizes were observed at KS2 for Year 6 pupils of each ethnicity, which were statistically significant for most ethnicities. Where they were not statistically significant (Chinese, other Asian, other ethnicity) this is likely to be due to the small number of pupils with that ethnic background in the data, rather than a substantive difference in impact. The impact of NTP tutoring at KS2 was similar for pupils of each ethnic background and confidence intervals overlapped with the confidence interval for all pupils.

Amongst KS4 Year 11 pupils the impact on maths attainment was positive and very small for each ethnicity and results were not statistically significant in most cases. There was

no evidence that the impact of tutoring on KS4 maths attainment was different for pupils of a particular ethnicity: confidence intervals for each ethnicity largely overlapped with the confidence interval for all pupils.

Figure 14: The school-level impact of the NTP on maths outcomes for pupils from different ethnic backgrounds.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 11: The school-level impact of the NTP on maths outcomes for pupils from different ethnic backgrounds; complete results from the linear mixed effects models

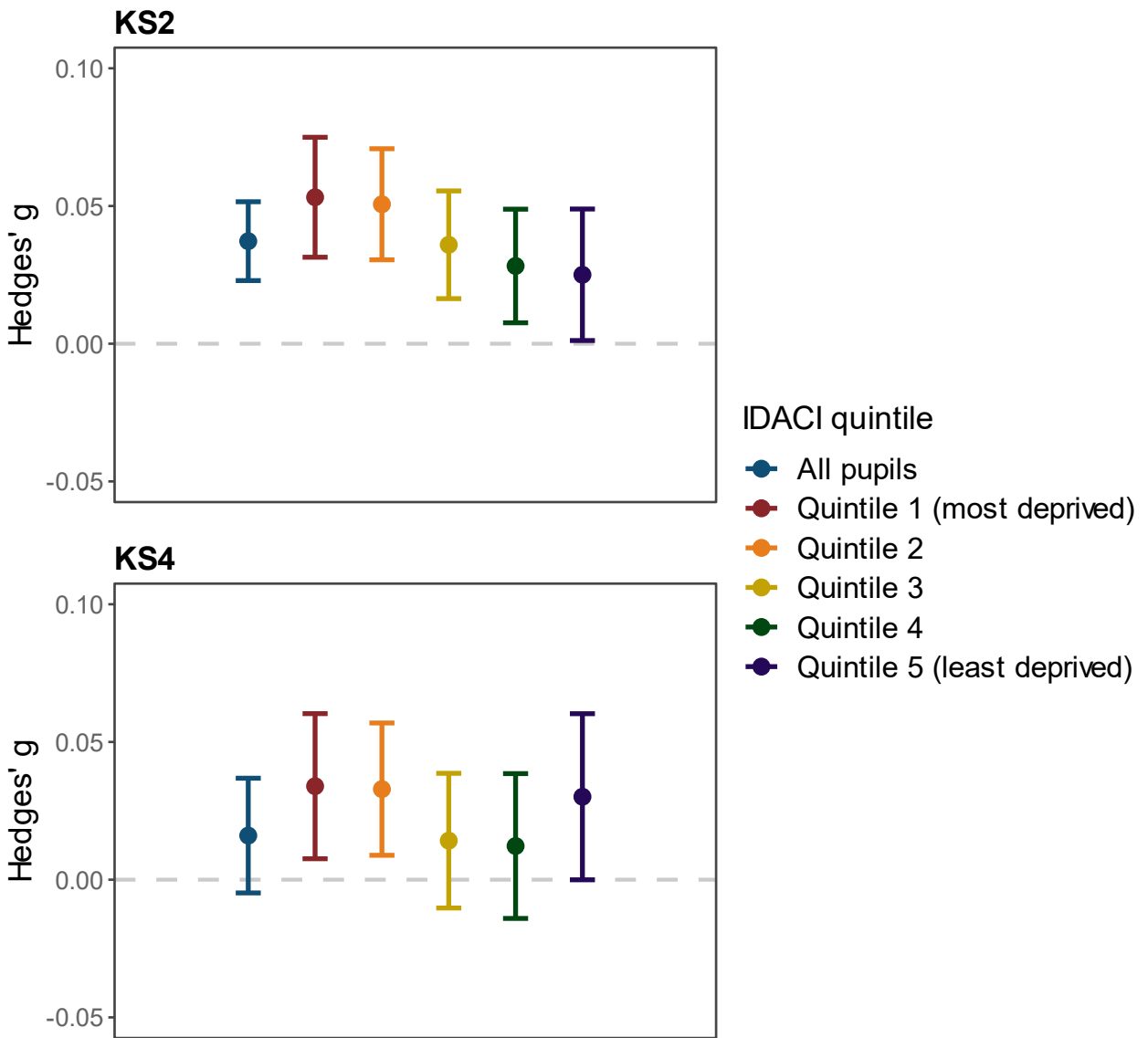
Outcome	Ethnicity	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	White	0.317 (0.2, 0.43)	0.042 (0.027, 0.057)	0	0.000
School-level KS2	Black	0.354 (0.11, 0.6)	0.048 (0.014, 0.081)	0	0.005
School-level KS2	Chinese	0.15 (-0.41, 0.71)	0.02 (-0.055, 0.095)	0	0.601
School-level KS2	Other Asian	0.13 (-0.08, 0.34)	0.017 (-0.01, 0.045)	0	0.219
School-level KS2	Mixed ethnicity	0.362 (0.16, 0.56)	0.046 (0.021, 0.071)	0	0.000
School-level KS2	Other ethnicity	0.132 (-0.22, 0.48)	0.018 (-0.029, 0.064)	0	0.459
School-level KS4	White	0.007 (-0.04, 0.05)	0.004 (-0.019, 0.026)	0	0.743
School-level KS4	Black	0.072 (0, 0.14)	0.036 (0, 0.072)	0	0.051
School-level KS4	Chinese	0.047 (-0.09, 0.19)	0.022 (-0.044, 0.087)	0	0.515
School-level KS4	Other Asian	0.074 (0.01, 0.14)	0.035 (0.006, 0.065)	0	0.020
School-level KS4	Mixed ethnicity	0.041 (-0.02, 0.1)	0.02 (-0.009, 0.05)	0	0.179
School-level KS4	Other ethnicity	0.077 (-0.02, 0.17)	0.037 (-0.01, 0.084)	0	0.119

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4), Randstad tuition participation data.

The school-level impact of NTP tutoring on KS2 and KS4 maths attainment for pupils living in areas with different levels of deprivation (IDACI quintiles) is shown in Figure 15 (all statistical results are shown in Table 11). At KS2, the impact of tutoring on maths attainment was small and positive for Year 6 pupils in every IDACI quintile. Results were statistically significant in every IDACI quintile, providing evidence that tutoring benefitted KS2 maths attainment. There may be slight evidence of a monotonic trend at KS2: effect sizes decrease as the level of deprivation in the pupil's area decreases (higher IDACI quintiles). This potential trend should be interpreted with caution, however, as confidence intervals still largely overlap between IDACI quintiles.

For the KS4 models, small positive effect sizes were observed for Year 11 pupils in all IDACI quintiles. There was a mixture of statistically significant and non-statistically significant results. Statistical significance (or lack thereof) is unlikely to reflect a substantive difference in impact for any one IDACI quintile, as once again effect sizes were similar and confidence intervals largely overlapped with the interval for all pupils.

Figure 15: The school-level impact of the NTP on maths outcomes for pupils from areas with different levels of deprivation.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 12: The school-level impact of the NTP on maths outcomes for pupils from areas with different levels of deprivation; complete results from the linear mixed effects models

Outcome	IDACI	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	IDACI quintile 1 (most deprived)	0.42 (0.25, 0.59)	0.053 (0.031, 0.075)	0	0.000
School-level KS2	IDACI quintile 2	0.39 (0.24, 0.55)	0.051 (0.03, 0.071)	0	0.000
School-level KS2	IDACI quintile 3	0.27 (0.12, 0.42)	0.036 (0.016, 0.055)	0	0.000
School-level KS2	IDACI quintile 4	0.21 (0.06, 0.36)	0.028 (0.008, 0.049)	0	0.007
School-level KS2	IDACI quintile 5 (least deprived)	0.18 (0.01, 0.35)	0.025 (0.001, 0.049)	0	0.040
School-level KS4	IDACI quintile 1 (most deprived)	0.07 (0.01, 0.12)	0.034 (0.008, 0.06)	0	0.012
School-level KS4	IDACI quintile 2	0.07 (0.02, 0.11)	0.033 (0.009, 0.057)	0	0.007
School-level KS4	IDACI quintile 3	0.03 (-0.02, 0.08)	0.014 (-0.01, 0.039)	0	0.256
School-level KS4	IDACI quintile 4	0.02 (-0.03, 0.08)	0.012 (-0.014, 0.039)	0	0.363
School-level KS4	IDACI quintile 5 (least deprived)	0.06 (0, 0.12)	0.03 (0, 0.06)	0	0.051

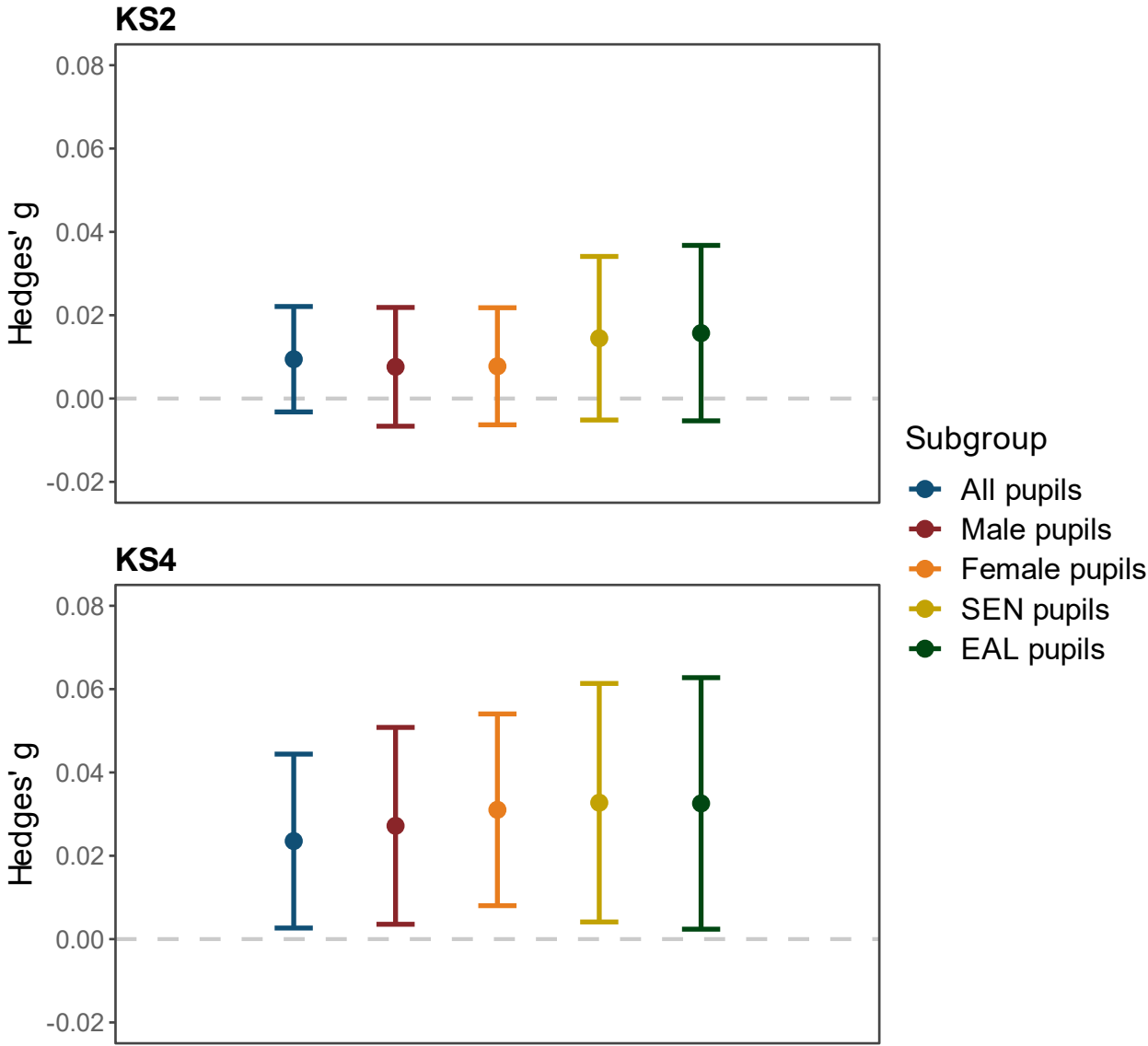
Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

How does the school-level impact of the NTP on English outcomes vary according to pupil characteristics?

Figure 16 below shows the impact of the NTP on English attainment for males, females, SEN pupils and EAL pupils (all statistical results are given in Table 12. The main school-level impact estimates for all pupils are included as a point of reference. At KS2 the impact on reading was small and positive for all subgroups of Year 6 pupils. However, the results at KS2 were not statistically significant for any subgroup. Effect sizes were similar in each subgroup and confidence intervals largely overlapped, so the impact of tutoring was similar in each subgroup.

At KS4 very small positive impacts were seen in all the subgroups of Year 11 pupils plotted in Figure 16. Results were statistically significant in all cases, indicating that tutoring is likely to have benefitted pupils in each subgroup. The impact of tutoring was similar in between subgroups and confidence intervals largely overlapped with the confidence interval for all pupils.

Figure 16: The impact of the NTP on English outcomes at school level by pupil characteristics.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 13: The impact of the NTP on English outcomes at school level by pupil characteristics; complete results from the linear mixed effects models

Outcome	Subgroup	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	Male Pupils	0.06 (-0.05, 0.17)	0.008 (-0.007, 0.022)	0	0.296
School-level KS2	Female Pupils	0.059 (-0.05, 0.17)	0.008 (-0.006, 0.022)	0	0.280
School-level KS2	SEN Pupils	0.123 (-0.04, 0.29)	0.014 (-0.005, 0.034)	0	0.148
School-level KS2	EAL Pupils	0.121 (-0.04, 0.28)	0.016 (-0.005, 0.037)	0	0.144
School-level KS4	Male Pupils	0.049 (0.01, 0.09)	0.027 (0.004, 0.051)	0	0.024
School-level KS4	Female Pupils	0.056 (0.01, 0.1)	0.031 (0.008, 0.054)	0	0.008
School-level KS4	SEN Pupils	0.057 (0.01, 0.11)	0.033 (0.004, 0.061)	0	0.025
School-level KS4	EAL Pupils	0.059 (0, 0.11)	0.033 (0.002, 0.063)	0	0.035

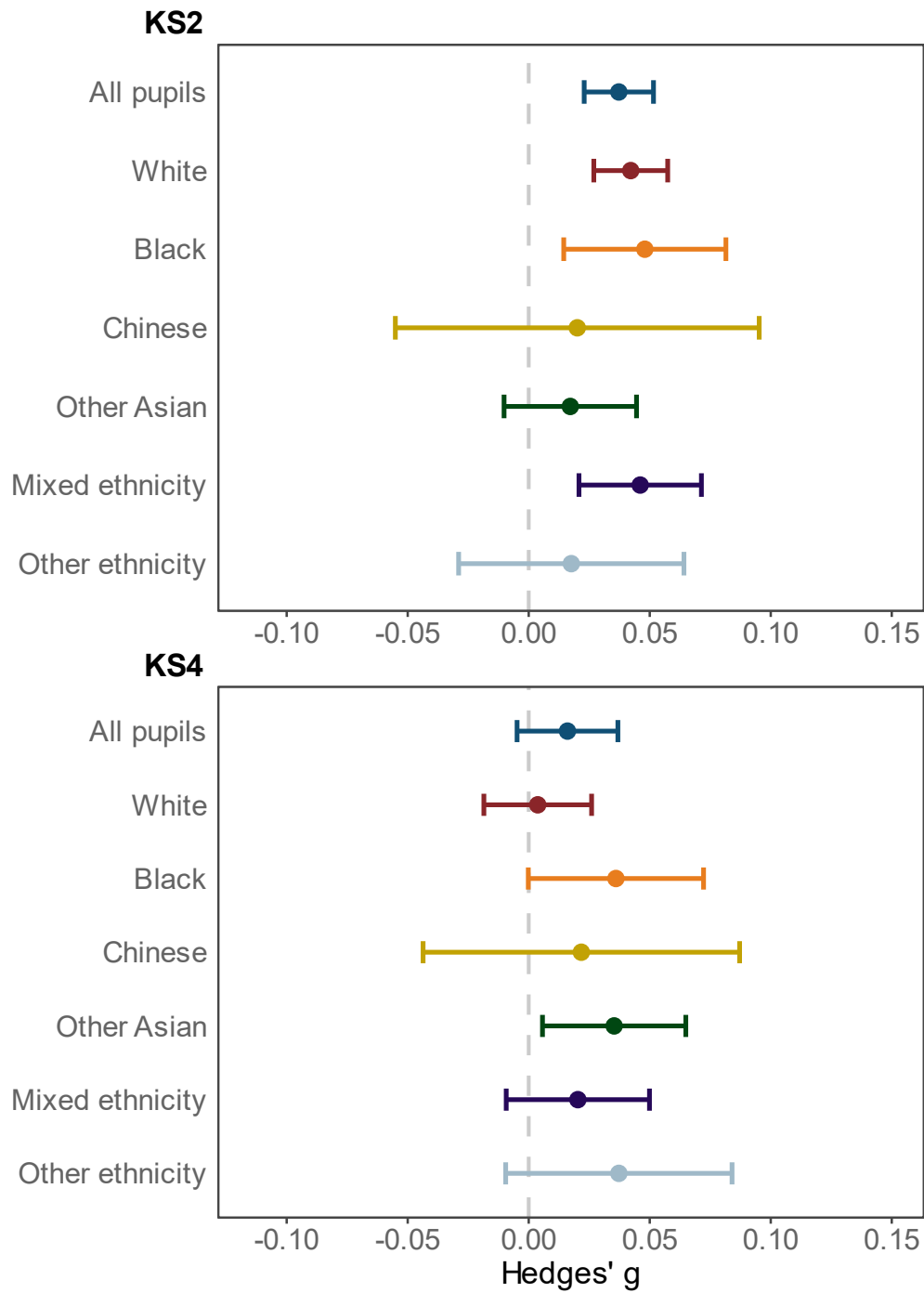
Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

The school-level impact of NTP tutoring on English attainment amongst Year 6 pupils of different ethnicities is shown in Figure 17 (all statistical results are shown in Table 13). Small positive or negative effect sizes were observed at KS2 for Year 6 pupils across the different ethnic groups, none of which were statistically significant. Although effect sizes for Chinese pupils and pupils from 'other' ethnicities were slightly negative, confidence intervals were wide due to the small number of pupils in these groups and overlapped

with the interval for all pupils. There is therefore no evidence of a different impact of the NTP on English attainment in these (or any other) ethnic groups at KS2.

Amongst KS4 Year 11 pupils the school-level impact on English attainment was positive and small for each ethnicity. Not all results were statistically significant, although it is more important to note that confidence intervals for each ethnic group largely overlapped with the interval for all pupils. There was therefore no evidence that the impact of tutoring on KS4 English attainment was different for pupils of a particular ethnicity.

Figure 17: The school-level impact of the NTP on English outcomes for pupils from different ethnic backgrounds.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 14: The school-level impact of the NTP on English outcomes for pupils from different ethnic backgrounds; complete results from the linear mixed effects models

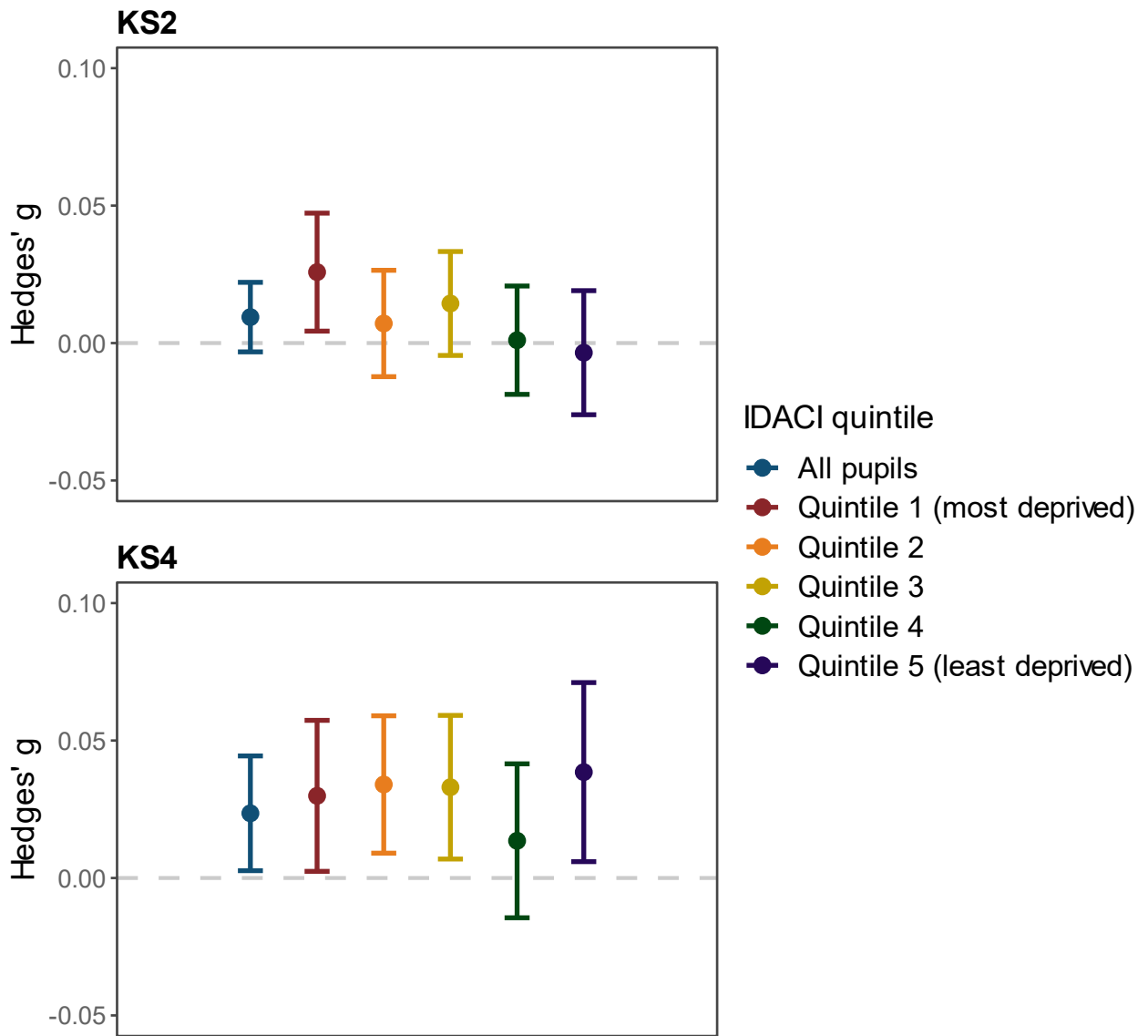
Outcome	Ethnicity	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	White	0.063 (-0.04, 0.17)	0.008 (-0.005, 0.022)	0	0.239
School-level KS2	Black	0.214 (-0.03, 0.46)	0.029 (-0.004, 0.062)	0	0.083
School-level KS2	Chinese	-0.284 (-0.86, 0.3)	-0.037 (-0.114, 0.039)	0	0.338
School-level KS2	Other Asian	0.195 (-0.01, 0.4)	0.026 (-0.001, 0.053)	0	0.060
School-level KS2	Mixed ethnicity	0.07 (-0.14, 0.28)	0.009 (-0.017, 0.035)	0	0.507
School-level KS2	Other ethnicity	-0.059 (-0.43, 0.31)	-0.008 (-0.055, 0.04)	0	0.754
School-level KS4	White	0.039 (0, 0.08)	0.021 (-0.001, 0.044)	0	0.060
School-level KS4	Black	0.02 (-0.05, 0.09)	0.011 (-0.027, 0.05)	0	0.564
School-level KS4	Chinese	0.105 (-0.05, 0.26)	0.054 (-0.023, 0.132)	0	0.171
School-level KS4	Other Asian	0.027 (-0.03, 0.09)	0.015 (-0.018, 0.048)	0	0.375
School-level KS4	Mixed ethnicity	0.08 (0.02, 0.14)	0.043 (0.01, 0.076)	0	0.010
School-level KS4	Other ethnicity	0.037 (-0.05, 0.13)	0.02 (-0.029, 0.069)	0	0.420

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

The school-level impact of NTP tutoring on KS2 and KS4 English attainment for pupils living in areas with different levels of deprivation is shown in Figure 18 below (all statistical results are shown in Table 14). At KS2, the impact of tutoring on maths attainment was small and positive for Year 6 pupils in most IDACI quintiles; it was small and negative for pupils in the least deprived quintile. Results were not statistically significant in most IDACI quintiles. As for the KS2 maths models, there was a slight suggestion of a trend: effect sizes decrease as the level of deprivation in the pupil's area decreased (higher IDACI quintiles). This should again be interpreted with caution, as there is a high degree of overlap between the confidence intervals in different IDACI quintiles.

For the KS4 models, there were small positive effect sizes for NTP tutoring amongst Year 11 pupils in all IDACI quintiles. Results were statistically significant for most IDACI quintiles: the impact in quintile 4 was not, but this is unlikely to indicate a differential impact. Effect sizes were similar for pupils in different IDACI quintiles and confidence intervals largely overlapped with the interval for all pupils.

Figure 18: The school-level impact of the NTP on English outcomes for pupils from areas with different levels of deprivation.



Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Table 15: The school-level impact of the NTP on English outcomes for pupils from areas with different levels of deprivation; complete results from the linear mixed effects models

Outcome	IDACI	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
School-level KS2	IDACI quintile 1 (least deprived)	-0.026 (-0.19, 0.14)	-0.004 (-0.026, 0.019)	0	0.761
School-level KS2	IDACI quintile 2	0.008 (-0.14, 0.16)	0.001 (-0.019, 0.021)	0	0.918
School-level KS2	IDACI quintile 3	0.112 (-0.04, 0.26)	0.014 (-0.005, 0.033)	0	0.136
School-level KS2	IDACI quintile 4	0.056 (-0.1, 0.21)	0.007 (-0.012, 0.026)	0	0.471
School-level KS2	IDACI quintile 5 (most deprived)	0.202 (0.03, 0.37)	0.026 (0.004, 0.047)	0	0.019
School-level KS4	IDACI quintile 1 (least deprived)	0.068 (0.01, 0.13)	0.039 (0.006, 0.071)	0	0.021
School-level KS4	IDACI quintile 2	0.024 (-0.03, 0.07)	0.014 (-0.014, 0.042)	0	0.344
School-level KS4	IDACI quintile 3	0.06 (0.01, 0.11)	0.033 (0.007, 0.059)	0	0.013
School-level KS4	IDACI quintile 4	0.062 (0.02, 0.11)	0.034 (0.009, 0.059)	0	0.008
School-level KS4	IDACI quintile 5 (most deprived)	0.054 (0, 0.1)	0.03 (0.002, 0.057)	0	0.033

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)

Long term impact: What is the longer-term impact of the NTP delivered in 2021-22 on maths and English attainment outcomes for pupils in Year 6 and Year 11 in 2022-23?

Key Findings

- We found some evidence that participation in SLT/AM in 2021-22 had a sustained positive impact on maths outcomes for KS2 pupils in 2022-23 but not at KS4.
- We were not able to detect any evidence of a positive impact of participating in the NTP in 2021-22 (any route) on KS2 or KS4 English outcomes.

This section presents the results from a series of models investigating the school-level impact of NTP tutoring in 2021-22 on maths and English attainment in 2022-23. Analysed pupils were in Year 6 and Year 11 in 2022-23, so were in Year 5 and Year 10 respectively when they received tutoring. Results are reported for the SLT/AM routes (combined) and then separately for the TP route. Estimates are provided for the whole sample and additionally the subgroups of PLA and PP pupils.

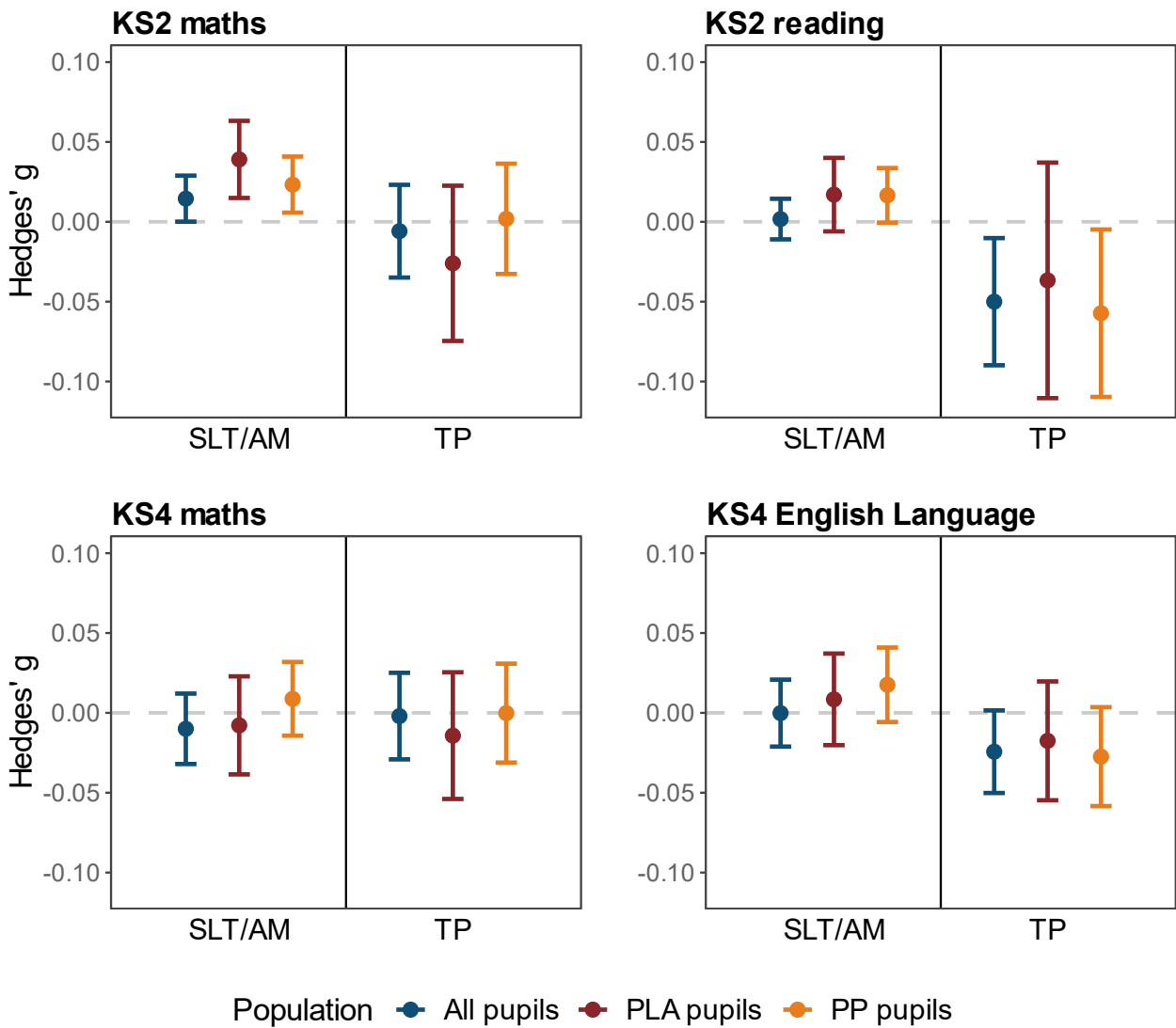
The school-level impact of 2021-22 NTP tutoring on 2022-23 KS2 and KS4 English and maths attainment is shown in Figure 19 (all statistical results are shown in Table 15). Impacts on 2022-23 maths attainment were generally close to zero at both KS2 and KS4, for all tutoring routes and pupil subgroups. However, there were small, positive and statistically significant impacts on KS2 maths attainment seen for all groups of pupils taking the SLT/AM route. This is consistent with the observation that impacts for year 3 of the programme were greatest for KS2 maths.

There was no evidence of a positive impact of 2021-2022 NTP tutoring on KS2 and KS4 English attainment, across all tutoring routes and pupil subgroups. There is a statistically significant negative effect size for the impact of TP on 2022-23 KS2 reading amongst all pupils. However, this is more likely to represent no impact than an actively detrimental impact: a small amount of school-level bias could account for a negative effect size on this scale.

The general impression across all results shown in Table 15 is of very small impacts that are not distinguishable (or barely distinguishable) from zero except for KS2 maths delivered via the SLT route. There is some indication that the SLT/AM route generally performs better than TP at KS2, but it is not possible to be sure of this trend due to the large confidence intervals around the TP effect sizes.

Complete results for the long-term impact analysis can be found in Appendix A.

Figure 19: The impact of participation in the NTP in 2021-22 on English and maths outcomes at KS2 and KS4 in 2022-23



Source: National Pupil Database 2021/2022 (School Census) and 2022/2023 (School Census, KS2, KS4)

Discussion

Overall, we found consistent evidence across all three subgroups of pupils that the NTP was effective for improving maths and English attainment at KS2 and some more limited evidence at KS4. All improvements seen were small. This is consistent with findings from previous research indicating that tutoring tends to be more effective for younger pupils; especially for English (Nickow, Oreopoulos and Quan, 2020; EEF, 2021a, 2021b). This may be because tutoring is more able to help fill learning gaps for primary school pupils, whereas for older pupils, there may be more significant gaps in their learning because they have accumulated over a longer period, especially for disadvantaged pupils, making it harder for them to be addressed through catch-up interventions such as tutoring sessions (E-ACT, No date). In addition, it has also previously been hypothesised that this is due to critical learning periods for language in a child's development (Nickow, Oreopoulos and Quan, 2020), where the brain is most plastic and able to learn up to around the age of 12 (i.e., the end of primary school/early years of secondary school) (see Gariépy, Bailey and Holochwost, 2020; Hartshorne, Tenenbaum and Pinker, 2018).

Another potential explanation for differences in tutoring effectiveness between primary and secondary school pupils is differences in tutoring delivery. For example, the Year 3 implementation and process evaluation indicated that primary schools were likely to deliver tutoring during afternoons once core subject teaching had finished (Lynch et al., 2023). Primary schools were also more likely than secondary schools to have rotated the times tutoring took place to ensure pupils did not always miss the same lesson, which may have supported pupil engagement with the sessions. It is also possible that tutoring may have been delivered online more frequently in secondary schools, was offered in more different subjects (potentially increasing the dilution effects), or that there were differences in the alignment with the school curriculum which could also have affected outcomes. However, unlike previous research we found larger effect sizes for maths than reading for KS2 pupils (see Nickow, Oreopoulos and Quan, 2020).

Our results are also broadly consistent with the results we saw for SLT in the impact evaluation for the second year of the NTP. This previous evaluation found evidence that participation in SLT led to small improvements in maths outcomes at both KS2 and KS4 (Lucas et al., 2023). There was also some more limited evidence at a school level that SLT led to improvements in English outcomes at both KS2 and KS4. The consistency of this evidence between these two impact evaluations is somewhat promising, especially for KS2 pupils, and builds on previous research indicating that small group tutoring is effective (e.g. EEF, 2021a, 2021b). At KS4 and for KS2 English the small effect sizes and lack of consistent statistically significant results means that we cannot completely exclude the possibility that these effects may be artifacts of selection bias despite our considerable attempts to remove it, rather than genuine effects of the NTP on pupil outcomes. For example, Weidmann and Miratrix (2020) showed 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 which

is larger than any of the effect sizes detected for the NTP to date. However, for KS2 maths the fact that we find statistically significant effects for nearly all subgroups, each of which has a separate match, coupled with the statistically significant longer-term effects of KS2 maths tutoring mean and the consistency with the year 2 evaluation results mean that we can be more confident that this is indeed a real effect and not an artifact of school-level selection bias.

This year we were not able to separate our evaluation by tutoring route. That said, the results of the Year 3 implementation and process evaluation (Lynch et al., 2023) which explored experiences of the NTP with a sample of participating schools indicated that among survey respondents, the vast majority of schools (85%) were participating in SLT. If this survey is representative of the proportions of schools which used each route across England, then this information combined with the positive results for SLT from last year's impact evaluation may indicate that the positive effects seen here could be primarily attributable to SLT.

Overall, the effect sizes observed here are of a broadly similar magnitude to those detected in the Year 2 impact evaluation (Lucas et al., 2023) i.e., they equated to one months' additional progress or less. This means that, as with the results we saw in the previous impact evaluation, the effect sizes are smaller than we might expect based on evidence about the effectiveness of small group tuition (Ritter et al., 2009; Dietrichson *et al.*, 2017; Nickow, Oreopoulos and Quan, 2020; EEF, 2021a). However, as we have noted previously (see Lucas et al., 2023), there are factors related to the evaluation design (e.g., specific target population, ability to conduct pre- and post-tests) and tutoring implementation which make it more likely that these previous studies would detect a larger effect than reported here. In addition, although we have attempted to mitigate dilution effects by imposing a minimum threshold for the percentage of tutored pupils in intervention schools, dilution remains an issue within the present analysis (see limitations). This means that as with the year 2 impact evaluation, the effect sizes reported here are likely to underestimate the true effect. Furthermore, the lack of subject information adds to the dilution at both school and pupil level as it means that, for example, the intervention group for evaluating maths outcomes is likely to include some pupils that only received tutoring in English. Once these factors are taken into account, the effect sizes detected are likely consistent with previous research.

A key challenge for this and previous evaluations of the NTP has been developing an appropriate comparison group for the pupil-level analysis as schools choose which pupils receive tutoring. This means that they are likely to have selected pupils most in need of support, resulting in negative selection bias (i.e., without tutoring these pupils would be expected to have lower attainment scores than their peers). To reduce selection bias within the pupil-level analysis, we restricted the analysis to pupils who were selected for tutoring in the second year of the NTP. This meant our main pupil-level analysis compared pupils who also received tutoring in the third year of the NTP with pupils who did not because their school did not take part in the NTP in the third year. We found small

positive effects of also receiving tutoring in Year 3 of the NTP on pupil outcomes at KS2. In contrast the pupil-level sensitivity analysis, which was essentially a within-school comparison of pupils who were and were not selected for tutoring in the third year of the NTP, found small negative effects. A similar pattern of results was seen at KS4 although the effect sizes were smaller in all cases than at KS2. These results provide a strong indication that selection bias is still present in the pupil-level analysis and therefore more strongly indicate that our estimates are likely to underestimate the true effect sizes.

As noted above and in earlier reports (Lucas et al., 2023), it is also likely that tutoring implementation factors such as differences in tutoring dosage could account for differences in outcomes between the NTP impact evaluations and previous studies of tutoring. This year the dosage analysis explored how best to divide a fixed number of tutoring hours between pupils to achieve the optimum impact at school level. Our analysis suggests that concentrating a higher number of tutoring hours on a smaller number of pupils is likely to be most beneficial overall, even though this means that some pupils will not receive tutoring. This suggests that headteachers should focus on identifying pupils that would benefit from longer-term support when considering how best to spend tutoring funds. However, while we estimated an optimum number of impact hours of tutoring overall, we are unable to observe how many of these hours were in the relevant subject due to the lack of subject information. Instead, we estimated the number of maths or English hours based on NTP year 2 AM/TP data, which introduced additional uncertainty into the final estimates. More research would be needed to optimise tutoring dosage and maximise the benefits for pupils and schools.

In addition to tutoring dosage, we still do not yet have a full understanding of how to optimise other aspects of tutoring such as session duration, mode of delivery (online vs in person), aligning sessions with the school curriculum, and time of delivery, to get the best results for schools as pupils. For example, research conducted by Ofsted (GOV.UK, 2023a) found that tutoring was most effective when it was embedded in schools as part of their day-to-day processes and schools had a clear strategy for implementing tutoring and engaging parents/carers. Unsurprisingly, the quality of the tutor and tutoring sessions was also found to be key to the success of the sessions. Ofsted's work also found that some of the weakest tutoring was provided online, potentially providing insights into the optimum mode of delivery. However, the work undertaken by Ofsted was not an impact evaluation so could not measure the effectiveness of tutoring. Given that schools had a lot of flexibility over tutoring delivery, we would expect that there will be variation in how schools choose to implement tutoring which could in turn impact outcomes.

The shift in focus of the NTP to reducing the persistent attainment gap for socioeconomically disadvantaged pupils this year means that understanding any longer-term benefit of tutoring is key. To close the attainment gap, progress made via tutoring would need to be sustained over time. Here we were only able to detect a longer-term benefit of having received SLT/AM in the previous academic year on maths outcomes for KS2 pupils (though we note no subject information was available for SLT). No other

longer-term positive effects were found despite the positive effects seen for SLT in the year 2 impact evaluation. In addition, the proportion of socioeconomically disadvantaged pupils (i.e., PP pupils) selected to receive tutoring this year was smaller than last year (see Sample Characteristics and Lucas *et al.*, 2023). This makes it more challenging to close the attainment gap for these pupils as they are not necessarily receiving the support. If the goal is to close the attainment gap for socioeconomically disadvantaged pupils, then support needs to be targeted at them. However, we know that schools will not have the funds to support all disadvantaged pupils (see Lynch *et al.*, forthcoming) and that there may be other pupils including PLA pupils who may stand to benefit as much from tutoring.

Recommendations

An objective of the NTP is to move towards a model of embedding tutoring within schools. With this in mind, we make the following recommendations:

- Senior leaders should consider which pupils are likely to benefit from regular tutoring and concentrate the tutoring hours they are able to offer on a smaller number of pupils. An optimum number of hours for greatest impact within a limited budget is likely to lie above 20 hours per pupil, closer to the EEF recommendation of 30 hours of tuition.
- DfE should seek to further build the evidence base around how to optimise the delivery of tutoring to maximise impact. This includes understanding the 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).
- If there is going to be continued funding for tutoring, consideration should be given to focussing it on maths and English in primary school. Evidence for the impact of primary maths tuition was strong in this evaluation (including a sustained impact), while previous evidence suggests strong benefits of English tuition in this age group.
- DfE should consider alternative approaches to closing the socioeconomic attainment gap. There is evidence that tutoring has had immediate benefits for pupils that participated, and that this may be slightly higher for PLA and PP pupils. However, it hasn't been effective as a mechanism for closing the socioeconomic performance gap because disadvantaged pupils have not been selectively targeted, and there is only limited evidence of sustained benefit (in KS2 maths).

References

Andrews, J. (2023) Recovering from the Covid-19 pandemic: analysis of Star Assessments. Available at: <https://renaissance.widen.net/view/pdf/sz0pq7dkzf/Recovering-FromTheCovid19PandemicMay2023.pdf> (Accessed: 22 July 2024).

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: 22 July 2024).

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#data-Block-f072b9bb-f8a9-4119-bc67-08da8695004d-tables> (Accessed: 22 July 2024).

Dietrichson, J., Bøg, M., Filges, T. and Jørgensen, A.-M. (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>.

E-ACT (No Date) Keep up not catch up: building a strategy for inclusion and achievement. Available at: <https://www.e-act.org.uk/wp-content/uploads/2022/11/Keep-up-not-catch-up-FINAL.pdf> (Accessed: 24 May 2024).

EEF (2021a) One to one tuition. Available at: <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/one-to-one-tuition>. (Accessed: 22 May 2024).

EEF (2021b) Small group tuition. Available at: <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/small-group-tuition> (Accessed: 21 March 2024).

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 (Accessed: 22 July 2024).

EEF (2023) Teaching and learning: Early Years Toolkit guide. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/toolkit/Toolkit_guide_v1.2_-_2023.pdf (Accessed: 24 May 2024).

Gariépy, J.-L., Bailey, D.B. and Holochwost, S.J. (2020) 'Critical periods', *Encyclopedia of Infant and Early Childhood Development* (Second Edition). Edited by J.B. Benson, pp. 347–357. Available at: <https://doi.org/10.1016/B978-0-12-809324-5.23626-2>.

GOV.UK (2022) Key stage 2 attainment: academic year 2021/22, GOV.UK. Available at: <https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment/2021-22> (Accessed: 26 June 2024).

GOV.UK (2023a) Independent review of tutoring in schools: phase 2 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-2-findings> (Accessed: 22 May 2024).

GOV.UK (2023b) National Tutoring Programme: guidance for schools, 2022 to 2023. Available at: <https://www.gov.uk/government/publications/national-tutoring-programme-guidance-for-schools-2022-to-2023/national-tutoring-programme-guidance-for-schools-2022-to-2023> (Accessed: 22 May 2024).

GOV.UK (2024) National Tutoring Programme: guidance for schools – academic year 2023/24, GOV.UK. Available at: <https://www.gov.uk/government/publications/national-tutoring-programme-guidance-for-schools-academic-year-202324> (Accessed: 25 June 2024).

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>.

Hartshorne, J.K., Tenenbaum, J.B. and Pinker, S. (2018) 'A critical period for second language acquisition: evidence from 2/3 million English speakers', *Cognition*, 177, pp. 263–277. Available at: <https://doi.org/10.1016/j.cognition.2018.04.007>.

Lucas, M., Moore, E., Morton, C., Staunton, R. and Welbourne, S. (2023) Independent evaluation of the National Tutoring Programme year 2: impact evaluation. Available at: https://assets.publishing.service.gov.uk/media/6530d24692895c0010dcba04/Independent_Evaluation_of_the_National_Tutoring_Programme_Year_2_Impact_Evaluation.pdf (Accessed: 25 June 2024).

Lynch, S., Aston, K., Bradley, E., Morton, C., Del Pozo Segura, J.M. and Lord, P. (2023) Evaluation of the National Tutoring Programme year 3: implementation and process evaluation. Available at: https://assets.publishing.service.gov.uk/media/653fd1926de3b9000da7a64b/National_Tutoring_Programme_year_3_-_implementation_and_process_evaluation_-_research_report.pdf (Accessed: 21 March 2024).

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: 22 July 2024).

Nelson, J. and Sharp, C. (2020) Schools' responses to Covid-19: key findings from the wave 1 survey. Available at: https://www.nfer.ac.uk/media/5n5knw2l/schools_responses_to_covid_19_key_findings_from_the_wave_1_survey.pdf (Accessed: 22 July 2024).

Nickow, A., Oreopoulos, P. and Quan, V. (2020) The impressive effects of tutoring on PreK-12 learning: a systematic review and meta-analysis of the experimental evidence. Available at: <https://doi.org/10.3386/w27476>.

Poet, H., Lord, P., Styles, B., Oppedisano, V., Zhang, M. and Dorsett, R. (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> (Accessed: 22 July 2024).

Poet, H., Lord, P., Styles, B., Oppedisano, V., Zhang, M. and Dorsett, R. (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> (Accessed: 22 July 2024).

Ritter, G.W., Barnett, J.H., Denny, G.S. and Albin, G.R. (2009) 'The effectiveness of volunteer tutoring programs for elementary and middle school students: a meta-analysis', *Review of Educational Research*, 79(1), pp. 3–38. Available at: <https://doi.org/10.3102/0034654308325690>.

Twist, L., Jones, E. and Treleaven, O. (2022) The impact of Covid-19 on pupil attainment. Available at: https://www.nfer.ac.uk/media/w2sllon1/the_impact_of_covid_19_on_pupil_attainment.pdf (Accessed: 22 July 2024).

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>.

Welbourne, S., Smith, A., Staunton, R., Lord, P. and Aston, K. (2023) Evaluation of the National Tutoring Programme Year 3 study plan. Available at: https://assets.publishing.service.gov.uk/media/6542365c9e05fd000dbe7b29/National_Tutoring_Programme_year_3_-_implementation_and_process_evaluation_-_study_plan.pdf (Accessed: 26 June 2024).

Appendix A: complete results from the long-term impact analysis

Table 16: The impact of participation in the NTP in 2021-22 on English and Maths outcomes at KS2 and KS4 in 2022-23; complete results from the linear mixed effects models

Outcome	NTP route	Subgroup	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
KS2 maths score	SLT/AM	All pupils	0.11 (0, 0.22)	0.014 (0, 0.029)	0	0.049
KS2 maths score	SLT/AM	PLA pupils	0.27 (0.1, 0.44)	0.039 (0.015, 0.063)	0	0.002
KS2 maths score	SLT/AM	PP pupils	0.18 (0.04, 0.32)	0.023 (0.006, 0.041)	0	0.009
KS2 maths score	TP	All pupils	-0.04 (-0.26, 0.17)	-0.006 (-0.035, 0.023)	0	0.691
KS2 maths score	TP	PLA pupils	-0.18 (-0.52, 0.16)	-0.026 (-0.075, 0.023)	0	0.294
KS2 maths score	TP	PP pupils	0.01 (-0.25, 0.28)	0.002 (-0.033, 0.036)	0	0.916
KS2 reading score	SLT/AM	All pupils	0.01 (-0.09, 0.11)	0.002 (-0.011, 0.014)	0	0.794
KS2 reading score	SLT/AM	PLA pupils	0.12 (-0.04, 0.29)	0.017 (-0.006, 0.04)	0	0.148
KS2 reading score	SLT/AM	PP pupils	0.13 (0, 0.27)	0.017 (0, 0.034)	0	0.057

Outcome	NTP route	Subgroup	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
KS2 reading score	TP	All pupils	-0.39 (-0.7, -0.08)	-0.05 (-0.09, -0.01)	0	0.014
KS2 reading score	TP	PLA pupils	-0.26 (-0.79, 0.26)	-0.037 (-0.11, 0.037)	0	0.330
KS2 reading score	TP	PP pupils	-0.45 (-0.86, -0.04)	-0.057 (-0.11, -0.005)	-	0.032
GCSE maths grade	SLT/AM	All pupils	-0.02 (-0.06, 0.02)	-0.01 (-0.032, 0.012)	0	0.375
GCSE maths grade	SLT/AM	PLA pupils	-0.01 (-0.05, 0.03)	-0.008 (-0.039, 0.023)	0	0.617
GCSE maths grade	SLT/AM	PP pupils	0.02 (-0.03, 0.06)	0.009 (-0.014, 0.032)	0	0.456
GCSE maths grade	TP	All pupils	0 (-0.06, 0.05)	-0.002 (-0.029, 0.025)	0	0.884
GCSE maths grade	TP	PLA pupils	-0.02 (-0.07, 0.03)	-0.014 (-0.054, 0.025)	0	0.481
GCSE maths grade	TP	PP pupils	0 (-0.06, 0.06)	0 (-0.031, 0.031)	0	0.992
GCSE English language grade	SLT/AM	All pupils	0 (-0.04, 0.04)	0 (-0.021, 0.021)	0	0.988

Outcome	NTP route	Subgroup	KS2 scaled score points or KS4 grade point scores (95% CI)	Hedges' g (95% CI)	Additional months progress	P value
GCSE English language grade	SLT/AM	PLA pupils	0.01 (-0.03, 0.05)	0.008 (-0.02, 0.037)	0	0.564
GCSE English language grade	SLT/AM	PP pupils	0.03 (-0.01, 0.07)	0.018 (-0.006, 0.041)	0	0.140
GCSE English language grade	TP	All pupils	-0.04 (-0.09, 0)	-0.024 (-0.05, 0.002)	0	0.065
GCSE English language grade	TP	PLA pupils	-0.02 (-0.07, 0.03)	-0.018 (-0.055, 0.02)	0	0.356
GCSE English language grade	TP	PP pupils	-0.05 (-0.1, 0.01)	-0.027 (-0.058, 0.004)	0	0.083

Source: National Pupil Database 2022/2023 (School Census, KS2, KS4)



Department
for Education

© Department for Education copyright 2024

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

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

Reference: RR1445

ISBN: 978-1-83870-568-8

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

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