Education Endowment Foundation

Aspire to STEM

Evaluation report

June 2026

Louis Hodge, James Merewood,
Maria Jose Guevara, Helen Murphy,
Miguel Subosa, and Elena Rosa Speciani

Education Endowment Foundation

The Education Endowment Foundation (EEF) is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, colleges, and early years settings to improve teaching and learning for 2 to 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence**. Reviewing the best available evidence on teaching and learning and presenting in an accessible way.

- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.

- **Putting evidence to use**. Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by The Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were reendowed with an additional £137m from government, allowing us to continue our work until at least 2032.

For more information about the EEF or this report please contact:

Education Endowment Foundation
5th Floor, Millbank Tower,
21–24 Millbank,
London,
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Department for Education

What Works Network

Table of contents

About the evaluator

The programme was independently evaluated by RAND Europe. RAND Europe is a not-for-profit policy research organisation that helps to improve policy and decision making through research and analysis.

The evaluation team included: Louis Hodge; James Merewood; Maria Jose Guevara; Helen Murphy; and Elena Rosa Speciani (currently RAND Europe) and former RAND Europe employees including: Miguel Subosa; Finn Oades; Yulia Shenderovich; Miriam Broeks; Eleftheria Iakuvidou; Lillian Flemons; Sashka Dimova; Jack Pollard; Dr Julie Bélanger; Hannes Jarke; Andreas Culora; Lydia Lymperis; and Sonia Ilie.

The lead evaluator is Elena Rosa Speciani (September 2022 – present). Previously this was Miriam Broeks (August 2020 – August 2022), Sonia Ilie (July 2019 – July 2020), and Dr Alex Sutherland (inception – June 2019).

Contact details:

Elena Rosa Speciani
RAND Europe,
Eastbrook,
Shaftesbury Road,
Cambridge,
CB2 8BF, UK

Email: **erspeciani@randeurope.org**

Acknowledgements

We would like to thank Kata Milay (RAND) for her support during the project, and for all the support from the Education Endowment Foundation evaluation managers, including Guillermo Rodriguez Guzmán, Diotima Rapp, Faizaan Sami, Pheobe Fisher, and Vipin Yadav for their support at various stages of the project. We would also like to thank previous members of the team, including Miguel Subosa, Finn Oades, Yulia Shenderovich, Miriam Broeks, Eleftheria Iakuvidou, Lillian Flemons, Sashka Dimova, Jack Pollard, Dr Julie Bélanger, Hannes Jarke, Andreas Culora, Lydia Lymperis, Sonia Ilie, and Alex Sutherland.

Executive summary

The project

'Aspire to STEM' is a programme aimed to increase pupil attainment, and improve STEM (science, technology, engineering, and mathematics) teacher retention, in both primary and secondary schools in England. The programme was delivered through geographically based partnerships of schools intended to foster local communities of practice. Each partnership received funding for bespoke continuing professional development (CPD), careers education, STEM enrichment activities delivered by STEM Ambassadors,[1] and access to online resources. To ensure activities and resources were tailored to specific needs and goals, each partnership undertook a needs analysis and developed an action plan to inform the selection of activities. Partnerships made spending decisions regarding these activities, with final approval from STEM Learning.

STEM Learning developed and led the programme, with it being delivered through approved providers, including educational leads, subject experts, external partners, and STEM Ambassadors. The Aspire to STEM delivery was fully funded by the Department for Education (DfE) as part of their Teaching and Leadership Innovation Fund (TLIF). The Education Endowment Foundation (EEF) funded the evaluation of the programme.

This evaluation used a quasi-experimental design, employing weighted difference-in-differences with kernel-based propensity score matching. This approach allows for a comparison of how outcomes evolve in the treated group before and after the intervention, relative to a comparable control group selected based on having similar observable characteristics. A total of 207 schools (92 primary and 115 secondary) participated in the intervention. This evaluation though focused only on the effects in secondary schools. In addition to the 115 intervention schools, a further 302 schools were identified as eligible for the intervention and used to create the comparison group. As Aspire to STEM was a whole-school intervention, all pupils were included in the evaluation. Implementation and process evaluation (IPE) activities included surveys, interviews, and focus groups with teachers, headteachers, and educational leads, alongside a document review. Cohort 1 recruitment began in Spring Term 2018 and Cohort 2 in Autumn Term 2018, with delivery ending in March 2020.

Several revisions to the original study plan were required during the project. Planned analyses of pupil attainment were not possible due to the cancellation of national examinations during the COVID-19 pandemic, and the insufficient number of primary schools (and teachers there in) meant that teacher outcomes were analysed for secondary schools only. Analysis was undertaken to identify if the programme had a detectable effect on the progression to STEM A level courses and on science teacher retention in secondary schools.

Table 1: Key conclusions

Key conclusions	
1.	Due to COVID-19, the intended student attainment outcomes for this evaluation (key stage 2 and key stage 4 attainment in mathematics and STEM subjects) were not collected. Key conclusions are therefore based on progression to A level, supported by secondary outcomes and qualitative data from the IPE.
2.	The evaluation found that attending an Aspire to STEM secondary school slightly increased the likelihood of key stage 4 pupils progressing to taking a STEM subject at A level two years later, by 1.42 percentage points. This is equivalent to around an additional 1 in 70 year 11 pupils taking up a STEM A level.
3.	There was no difference in science teacher turnover between secondary schools which received Aspire to STEM and those that did not. This implies bespoke CPD provision does not influence science teachers' retention.
4.	Participating schools recognised the importance of partnership working but struggled to fulfil the programme's early-stage partnership milestones, including conducting the needs analysis and action planning.
5.	CPD was the most visible and coherent element of Aspire to STEM and was seen as aligning closely with teachers' perceived professional development needs. While engagement in CPD was comparatively high, teachers also reported substantial barriers to sustained participation, including time and workload pressures.

[1] STEM Ambassadors in this programme were external volunteers from industry or academia who support school activities in this subject.

Additional findings

This evaluation presents a mixed but informative picture of Aspire to STEM's impact and implementation. The impact evaluation found that pupils attending Aspire to STEM secondary schools had an increased likelihood of progressing to STEM A levels, but science teachers working in these schools do not experience any change in their rates of turnover[2] and wastage.[3] The IPE findings help explain these outcomes by examining how different programme components were delivered in practice.

Pupils who completed Key Stage 4 in an Aspire to STEM school were found to be 1.42 percentage points more likely to go on and study a STEM subject at A level, two years later. There is some uncertainty around this central estimate, with a confidence interval (CI) ranging from 0.68 to 2.18 percentage points. The central estimate corresponds to approximately a 15% increase on the baseline number of pupils completing at least one STEM A level. This result represents a detectable positive (non-zero) effect, and the assumptions underpinning the estimation approach have been rigorously tested. However, reliability is limited by group equivalence being necessarily based only on observable characteristics and outcome data was only available for a single post-intervention time period. By contrast, no difference in science teacher turnover or wastage was found. Overall, the programme increased pupil progression to STEM A levels but had no measurable effect on science teacher retention.

Implementation fidelity varied across components. CPD was the most consistently delivered and coherent strand of the programme and was widely perceived as relevant to teachers' professional needs. Teachers reported gains in subject knowledge, curriculum clarity, confidence in practical work, and access to high-quality resources, which many linked to increased pupil engagement and aspiration in STEM. These findings align with how the programme was intended to work, suggesting that improvements in classroom practice can support positive pupil outcomes even when other elements are delivered unevenly.

Engagement with other components, particularly STEM Ambassadors, careers education, and enrichment, was limited due to administrative barriers, staffing pressures, and delays in programme roll-out, which reduced overall dosage and constrained partnership working, especially in schools facing greater resource challenges. As a result, several intended pathways in the Theory of Change, notably those related to employer engagement and enrichment, were only partially realised. The absence of a teacher retention effect is consistent with the wider evidence base. Although CPD was valued, teachers cited workload pressures, limited access to cover, and staffing shortages were dominant influences on career decisions. These systemic factors, exacerbated by Covid-19 disruption, appear to have outweighed the potential retention benefits of the programme.

Cost

The indicative average cost of Aspire to STEM was around £4,805 per school, or £7.29 per pupil. This figure is only indicative, as it is based on limited information and makes several assumptions. Further details can be found in the 'Costs' section.

Impact

Table 2: Summary of impact on outcome(s)

Outcome	Effect size (95% CI)	Estimated months' progress	The EEF security rating	No. of schools	P-value
Progression to STEM A level	0.23 (0.11 – 0.35)	N/A	N/A	280	0.007
STEM teacher turnover	-0.00 (-0.35 – 0.34)	N/A	N/A	281	0.974
STEM teacher wastage	0.02 (-0.18 – 0.22)	N/A	N/A	281	0.859

[2] Turnover refers to teachers leaving their current school.
[3] Wastage refers to teachers leaving the profession entirely.

Introduction

Background

'Aspire to STEM' was a complex whole-school intervention designed to improve pupil attainment and teacher retention in science, technology, engineering, and mathematics (STEM) subjects across both primary and secondary schools. Aspire to STEM operated through clusters of schools in the same geographical area forming partnerships and sharing resources to build a sustainable 'community of practice' in STEM, adapted to their local needs and context. Aspire to STEM was delivered by STEM Learning, a well-established organisation for the provision of school-led STEM continuing professional development (CPD).

The UK has experienced a persistent shortage in STEM skills over the past decade. In 2014, it was reported that employers were unable to find enough future employees suitably qualified in STEM to meet demand (Straw and Macleod, 2013), with the issue still evident in more recent analyses (Greaves and Brawley, 2025). In 2025, 49% of engineering and technology firms reported difficulties with recruitment because of skill shortages, with these shortages estimated to cost the UK economy £1.5 billion a year (Greaves and Brawley, 2025). These challenges extend into the teaching profession, where recruitment shortfalls had been the most pronounced in STEM subjects (Snell *et al.,* 2024). At the time that the Aspire to STEM programme was introduced, only 47% of the annual recruitment target for physics teachers was met. Similarly, the recruitment of teachers in mathematics (71%), computing (73%), and chemistry (79%) were all below their respective targets (DfE, 2018). The latest Initial Teacher Training census for the 2025/2026 academic year shows that this is no longer the case, with recruitment in STEM subjects exceeding the target by 8 percentage points (DfE, 2025a). Previous shortfalls had resulted in a high number of non-specialist teachers teaching STEM subjects who may struggle to deliver high-quality instruction and inspire the next generation of STEM graduates, exacerbating STEM skills shortages in the future (Thompson-Lee *et al.,* 2025). Aspire to STEM aims to address these challenges by both further improving teacher retention and improving student outcomes in STEM subjects.

Evidence suggests that the types of approaches that Aspire to STEM utilised, including improving leadership and engaging schools in sustained, active, and collaborative CPD, are likely to lead to impact on teachers, with less evidence for impact on pupil outcomes. CPD is regarded as a 'key lever for improving teaching' (Opfer, 2016: p. 3) with sustained, active, and collaborative CPD is likely to lead to changes in teacher practice and pupil outcomes (Opfer, 2016).

While the evidence is limited, it indicates that CPD-based programmes may have the potential to improve student outcomes, particularly in the take-up of STEM subjects in key stage 4. Evaluation of the Institute of Physics' Stimulating Physics Network, which employed a similar multi-year collaborative CPD approach for secondary school physics teachers (particularly non-specialists), found notable increases in pupil uptake of STEM A levels. Specifically, participating schools saw a 16.6% increase in A level physics entries compared to a 1.3% increase in matched schools, and a 29.2% rise in female entries compared with a 13.0% increase nationally (Thomson and Plaister, 2022). Outcomes related directly to student learning have been estimated to increase with CPD by an overall effect size of 0.09, indicating a small but educationally meaningful improvement in attainment that is similar in magnitude to other widely used school-level interventions (Fletcher-Wood and Zuccollo, 2020). Similarly, when teachers in primary schools were given the ability to design, lead, and evaluate a mathematics CPD programme in their schools, staff reported visible changes to the students' understanding and engagement with the subject, although this evidence is only anecdotal (McNeill *et al.,* 2014).

Further evidence relating to non-academic outcomes remains inconclusive, with some CPD interventions reporting no changes in pupil behaviour or concentration, while others identify improvements in confidence and self-efficacy. Although the reported impacts appear more consistently positive in STEM subjects (Fletcher-Wood and Zuccollo, 2020). The evidence on student outcomes does have important methodological limitations and should therefore be interpreted with appropriate caution (Sims and Fletcher-Wood, 2018). Several CPD programmes have not demonstrated improvements in pupil outcomes (Garet *et al.,* 2011; Garet *et al.,* 2016).

Research in the United States suggests that high teacher turnover negatively affects student attainment (Atteberry *et al.,* 2017). Just under a third of early career teachers have left teaching by their fifth year in the profession, with only 52% still teaching after 14 years (DfE, 2025b). At the time this programme was commissioned, the teacher retention rate was at a similar level of just under a third of teachers still in service five years after qualification, but was trending downwards (DfE, 2019). There is further evidence that indicated that retention rates were lower for science teachers when this study was commissioned, with their odds of leaving being 26% higher than non-science teachers (Allen and Sims, 2017).

In England, this trend has resulted in incentive schemes, in the form of bursaries and scholarships for Initial Teacher Training. These are larger for trainees in chemistry, computing, mathematics, and physics.[4] While it has been found in other countries, financial incentives, such as loan forgiveness and cash awards, have been less effective in the recruitment of STEM teachers due to a range of higher paid jobs available to them (Thompson-Lee *et al.,* 2025), evidence suggests that DfE funding incentives have produced different results. It was found that a £10,000 increase in a subjects' bursary resulted in an overall increase in trainees of 17.6%, including across STEM subjects (Dawson *et al.*, 2023).

CPD may be an effective way of retaining existing teachers and so reducing the need for recruitment and associated costs. A 2017 study found that participation in STEM Learning CPD, where teachers were able to attend professional development courses and form connections with others in their profession, was associated with an increase in the odds of staying in their profession by 160% (Allen and Sims, 2017). Although not focused exclusively on STEM, reports from previous Teaching and Leadership Innovation Fund (TLIF) projects provide further evidence of this positive effect. Eight TLIF projects delivered high-quality CPD to teachers and school leaders in priority areas and found a positive impact on teacher retention. Participants were significantly more likely than other teachers and leaders to: remain in the teaching profession; continue working in challenging schools; and stay at the same school (Straw *et al.,* 2022).

Support and investment in high-quality staff through continued CPD is a key factor in the second part of the DfE's 'every child achieving and thriving' mission (DfE, 2026). Embedding continued CPD, initiated by school leaders, aims to create a school-level culture that is rewarding for teachers, thereby encouraging them to stay within the profession. Aspire to STEM has the potential to contribute towards this goal as well as the wider challenges of training and retaining STEM teachers (Bélanger and Broeks, 2016). While some of STEM Learning's previous CPD initiatives have been evaluated, they have not been evaluated with robust counterfactual designs that focus on pupil outcomes. This research therefore fills an important evidence gap on the effectiveness of CPD and enrichment programmes in improving pupil engagement in STEM subjects.

Beyond teacher recruitment and retention, Aspire to STEM has the potential to contribute to a substantial number of policy developments announced in 2024 through improvements in students' STEM outcomes. Both the Curriculum and Assessment Review[5] and Establishment of Skills England[6] developments aim to determine and address skills shortages in maths and technical education, for which Aspire to STEM may be able to contribute to addressing these shortages. Additionally, the 'Plan for Change'[7] and Modern Industrial Strategy[8] both aim to increase Britain's global standing in clean energy and kick-start economic growth through sectors including advanced manufacturing, defence, digital technologies, and life sciences. Both policies include aims to increase enrolment in apprenticeships, technical education, and higher education STEM degrees, for which STEM-based CPD has been proven to have an impact on increasing STEM enrolment beyond key stage 4 (Thomson and Plaister, 2022).

Aspire to STEM was implemented as a part of the TLIF, a three-year initiative in England that aimed to support high-quality professional development for teachers and school leaders. The programme focused on primary and secondary schools with the greatest need (schools rated 'Requires improvement' or 'Inadequate' by the Office for Standards in Education,

---

[4] See: **www.gov.uk/government/publications/funding-initial-teacher-training-itt/funding-initial-teacher-training-itt-academic-year-2026-to-2027#school-direct-salaried-route**
[5] The Curriculum and Assessment Review was commissioned in 2024, and the final report was published in November 2025. Available at: **www.gov.uk/government/publications/curriculum-and-assessment-review-final-report**
[6] Skills England was established in July 2024 in shadow form within the Department for Education (DfE). Available at: https://www.gov.uk/government/organisations/skills-england
[7] The Plan for Change 2024 outlines the UK Government's ambitious vision for national renewal, focusing on five key missions aimed at addressing economic and social challenges. Available at: https://www.gov.uk/government/publications/plan-for-change
[8] The Modern Industrial Strategy is the UK's ten-year plan launched in 2025, aimed at boosting economic growth by focusing on eight key sectors. Available at: https://www.gov.uk/government/collections/the-uks-modern-industrial-strategy-2025

Children's Services and Skills [Ofsted] and also located in an Opportunity Area,[9] although 'Good' schools could join with DfE approval and schools without a current Ofsted rating, such as newly converted academies, were also considered eligible).

School recruitment was undertaken through local networks and driven by geographic considerations and so the programme was not suitable to be evaluated through a randomised controlled trial (RCT). Instead, a quasi-experimental design (QED) was used to assess the impact of Aspire to STEM on key outcomes of interest. Specifically, a difference-in-differences approach was combined with inverse probability weighting (IPW). This design compares changes in school-level outcomes over time between the schools that enrolled in Aspire to STEM and eligible schools that did not enrol. IPW is used to improve the compatibility of the control group. The evaluation also included an implementation and process evaluation (IPE), which explored how Aspire to STEM was implemented across the different school clusters, and the barriers and facilitators to successful implementation.

Intervention

Aspire to STEM consists of bespoke CPD supplemented by STEM enrichment activities, focusing on four areas: i) improved leadership to support STEM teaching; ii) great teaching of STEM subjects; iii) increased science capital within disadvantaged communities; and iv) careers and information guidance, more details can be observed in Figure 1. However, the programme is flexible and was adapted to the needs of each school following an initial 'needs analysis'.

As part of this programme, each cluster of schools received: i) funding (£25,000) for bespoke CPD for teachers and leaders; ii) career guidance for students; iii) access to STEM inspiration and enrichment activities delivered by a network of STEM Ambassador volunteers; and iv) access to online resources via the STEM Learning 'STEM Club infrastructure and eLibrary'.

The following updated Template for Intervention Description and Replication (TIDieR) framework for the programme has been agreed on as part of the initial Intervention Delivery and Evaluation Analysis (IDEA) workshop in April 2018 and revised further in ongoing communication with the Aspire to STEM team.

TIDieR framework

1. Brief name

Aspire to STEM.

2. Why: Rationale, theory, and/or goal of essential elements of the intervention

Aspire to STEM aims to raise aspirations and improve student STEM academic outcomes through improved STEM teaching. The programme team believed that improved STEM teaching more than anything else can improve pupil opportunity and drive social mobility. As can be seen in the Theory of Change (Figure 1), Aspire to STEM was designed to lead to improvements across three levels:

1. **Leaders.** Improved leadership, new relationships with STEM employers, better engagement with parents, families, and the community;
2. **Teachers.** Increased confidence, motivation and competence in teaching STEM, better knowledge and ability to use real-life and industry contexts, improved teacher retention;
3. **Students.** Raised aspirations and ultimately, improved STEM outcomes.

[9] An Opportunity Area is a geographically defined place identified by the government as having persistently low levels of social mobility, where additional, targeted support is prioritised to improve the life chances of children and young people through education. Introduced in 2017 as part of the government's social mobility strategy, Opportunity Areas focused on a set of local authority districts (LADs) with weak educational outcomes and high levels of disadvantage, bringing together schools, early years providers, employers, local authorities, and community organisations to address barriers from early years through to employment and skills, with targeted investment and place-based partnership working at their core. See: https://www.gov.uk/government/publications/social-mobility-and-opportunity-areas

The programme aims to increase uptake of STEM among students and ultimately improve student STEM academic outcomes through improved STEM teaching, leadership, and community engagement.

3. Who: Recipients of the intervention

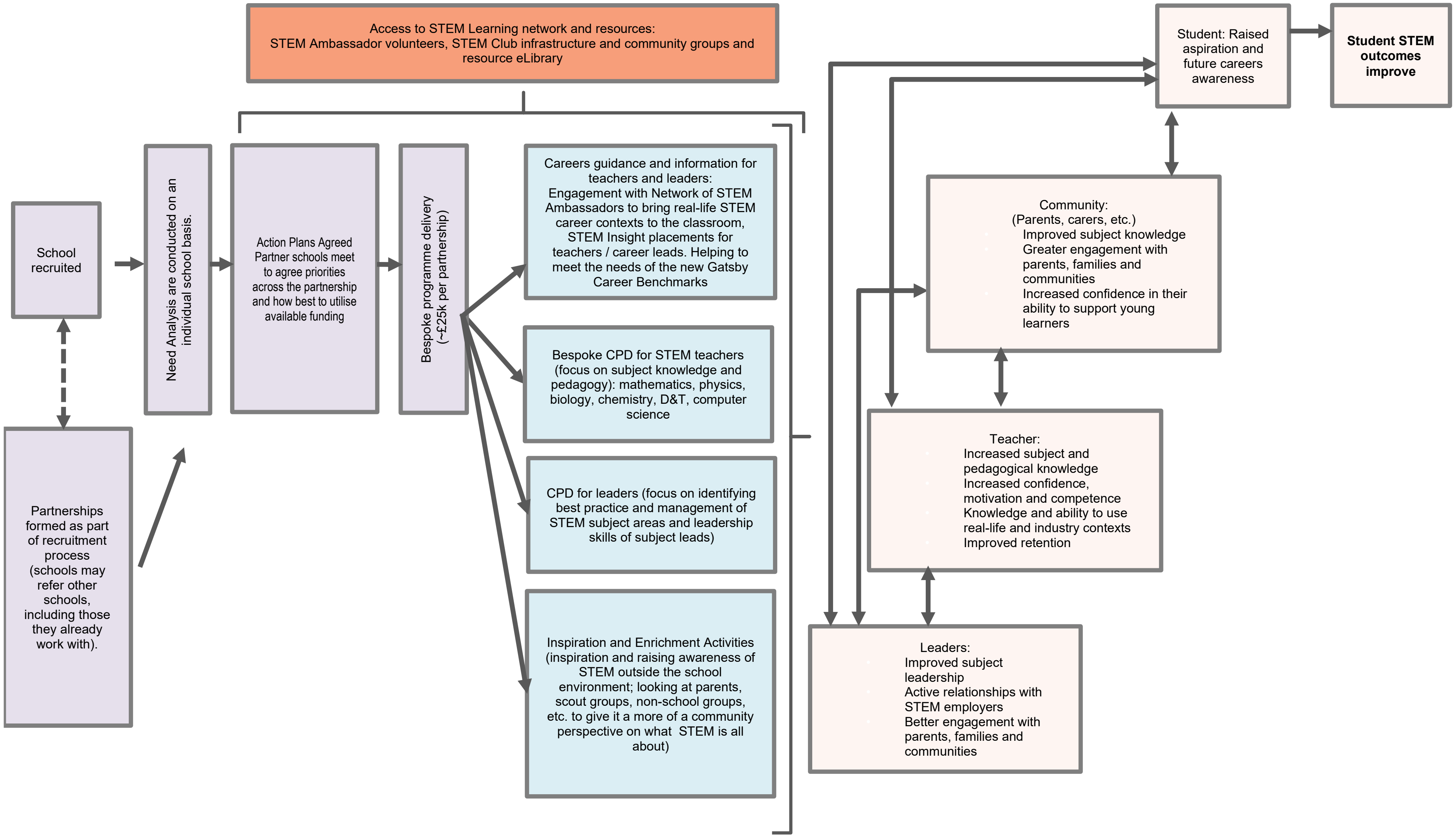
The programme targeted teachers and leaders of STEM subjects and school senior leaders in both primary and secondary schools. Aspire to STEM was implemented as a part of the Teaching and Leadership Innovation Fund (TLIF), a three-year initiative in England to support high-quality professional development for teachers and school leaders in England that need it most.[10] TLIF was offered in Opportunity Areas (OAs) and in Local Authority Districts (LADs), which were rated lowest in England (identified as LADs 5 or 6). Opportunity Areas were announced in 2016/2017 as areas 'identified as the most challenged when it comes to social mobility' by the DfE.[11] These areas were identified drawing on the Social Mobility Index and the Achieving Excellence Areas Index (DfE, 2017: p. 1). LADs rated 5 and 6 by DfE are LADs with the lowest scores on composite indicators that include measures of pupil attainment, such as Progress 8 scores, and 'capacity to improve indicators' identified by DfE, such as the number of teacher trainees per 10,000 pupils (DfE, 2016: pp. 10–11).

To be eligible for TLIF, schools in these areas had to be rated 'Requires improvement' or 'Inadequate' (3 or 4) in their most recent Ofsted inspection.[12] Schools rated as 'Good' were able to join the programme but only with DfE approval.[13] A number of schools in the eligible areas did not have an Ofsted rating at the time of recruitment, for instance because the school became an academy; these schools were considered as eligible to take part in the programme.

To facilitate creation of partnerships, STEM Learning prioritised recruitment by examining Google Maps to focus on areas where many eligible schools were located in geographical proximity of each other, meaning that isolated schools otherwise eligible would not be approached because of location. The recruitment process also aimed to keep all schools within an education authority together as these schools were likely to have existing collaborations.

[10] See: **www.gov.uk/guidance/teaching-and-leadership-innovation-fund**
[11] See: www.gov.uk/government/news/social-mobility-package-unveiled-by-education-secretary
[12] See: **www.gov.uk/guidance/teaching-and-leadership-innovation-fund-programmes-for-teachers-and-school-leaders#eligibility**
[13] The evaluation recorded the Ofsted ratings for schools included in the study.

Figure 1: Theory of Change



4. What: Physical or informational materials used in the intervention

Once the schools joined the programme, they had access to STEM Learning's network and resources: STEM Ambassador volunteers, STEM Club infrastructure and community groups and resource eLibrary. Participating schools were offered a variety of face-to-face and online CPD activities, complemented by teaching resources identified by the schools with the support of educational leads. Additionally, the programme targeted pupil aspirations through specific inspiration and enrichment activities. All materials were specifically tailored to the needs of each partnership.

5. What: Procedures, activities, and/or processes used in the intervention

Aspire to STEM created clusters of schools in the same geographical area to share resources and support, forming a partnership and building a sustainable 'community of practice' in STEM that was adapted to their local needs and context. Each partnership was supported by an educational lead and an education expert who worked with schools on their needs analysis and action planning and helped the partnership plan and implement the required activities.

All schools were offered a two-year support package, ending March 2020, with each partnership having access to £25,000. The funds were earmarked to cover the costs of CPD, with choices on spending being made by partnerships, with support of their educational leads and experts, and final sign-off being approved by STEM Learning.

As a key step of their participation, schools completed a needs analysis and action plan and agreed on the activities for the partnership of between three and ten schools. Once the action plan was agreed with the educational lead for each school in the partnership, schools could access several types of activities and resources in line with their action plan goals, including:

- bespoke CPD for STEM teachers (both on subject knowledge and subject pedagogy);
- bespoke CPD for STEM leaders;
- guidance on STEM careers for teacher and leaders; and
- inspiration and enrichment activities.

6. Who: Intervention providers/implementers

Aspire to STEM was delivered by providers approved by STEM Learning, including the National STEM Learning Centre subject expert team, STEM Learning's Network of Science Learning Partnerships, external educational experts, and trusted partner organisations (e.g. Maths Hubs, Institute of Physics), as well as 'STEM Ambassadors' who are STEM professionals that are registered and trained by STEM Learning.

7. How: Mode of delivery

Online and face-to-face.

8. Where: Location of the intervention

Aspire to STEM was delivered in a variety of places: on-site at school or a partnering school; Science Learning Partnership or regional centre; and at the National STEM Learning Centre.

9. When and how much: Duration and dosage of the intervention

Schools were recruited in two cohorts with schools being placed in partnerships with other schools in the same cohort. To ensure the feasibility of intervention roll-out, recruitment was staggered over two time points: recruitment for Cohort 1 started in Spring Term 2018; and recruitment for Cohort 2 started in Autumn Term 2018, partnered school belonged to the same Cohort. Implementation ended for both cohorts in March 2020.

Despite staggered starts, in practice, some Cohort 2 schools moved through the implementation process faster and started CPD earlier than some Cohort 1 schools. For example, by November 2018 only three out of the 18 partnerships involved in

Cohort 1 had fully completed their needs analysis and action plans in all partnership schools or agreed initial partnership activity, and four out of 18 had formally agreed partnership-level priorities. Therefore, schools from both cohorts are analysed jointly.

10. Tailoring: Adaptation of the intervention

By its nature, Aspire to STEM is a tailored programme. Each school completed a needs analysis with the help of the STEM Learning educational lead to identify their bespoke support needs. Following this, partnership leads agreed on an action plan for their partnership and drew upon several types of activities and resources (see 'What: Procedures, activities, and/or processes used in the intervention' above). This resulted in a combination of individual school and wider partnership support packages, adapted to the needs of the schools and partnerships.

Aspire to STEM was offered to both primary and secondary schools, but there are critical differences between the CPD needs of these different settings. In primary schools, it is common for teachers to be generalist and degrees in STEM subjects are not expected. However, in secondary schools, teachers are expected to have relevant education and/or experience in STEM subjects. Given the differences between STEM experiences and CPD needs, the support offered by Aspire to STEM differed between primary and secondary schools.

11. How well (planned): Strategies to maximise effective implementation

Each partnership was allocated a dedicated educational lead. Both the schools and educational leads had access to a number of support and assistance services (see 'What: Procedures, activities, and/or processes used in the intervention' above).

Evaluation objectives

Protocols for quasi-experimental studies are uncommon as quasi-experimental studies are conducted retrospectively and generally require more complex analyses with multiple decision points (Hedges, 2017). However, we created a study plan with pre-specified model sensitivity analyses as per Anders *et al.* (2017).

Research questions

This study seeks to compare school-level outcomes between schools that implement Aspire to STEM and schools running business as usual STEM CPD and enrichment activities. We had originally proposed to also explore key stage 2 and key stage 4 outcomes but due to data unavailability, owing to the cancellation of 2019/2020 and 2020/2021 key stage 2 and key stage 4 examinations in England as a result of COVID-19, these were not included in the final study. The final study outcomes are therefore, the original secondary outcomes—enrolment in STEM A levels, and teacher turnover and wastage.

Below we outline the research questions according to the latest study plan (previous research questions can be found in previously published study plans).[14]

Impact evaluation questions:[15]

1. Does Aspire to STEM lead to a higher proportion of students opting to take STEM A levels, compared to 'business as usual'?
2. Does Aspire to STEM lead to fewer (STEM) teachers leaving their school (teacher turnover), compared to 'business as usual' in secondary schools?
3. Does Aspire to STEM lead to fewer (STEM) teachers leaving the profession (teacher wastage), compared to 'business as usual' in secondary schools?

---

[14] See: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/aspire-to-stem
[15] The cancelation of exams as a result of the COVID-19 pandemic resulted in the primary impact question (Does Aspire to STEM lead to improvements in pupil attainment in key stage 2 maths, key stage 4 maths, and other Key Stage 4 STEM subjects compared to 'business as usual'?) not being explored as part of the evaluation as key stage 2 or key stage 4 outcome data was not available.

IPE questions:

4. How well was Aspire to STEM implemented in schools?
5. What are the typical components of Aspire to STEM?
6. What are the barriers and enablers for implementation of Aspire to STEM?

These questions were further divided into five more specific inquiries:

1. What was the level of fidelity in the intervention schools?
2. What was the level of dosage in the intervention schools?
3. What factors and initial conditions appear to explain variation in the dosage and fidelity of implementation?
4. What appear to be the necessary conditions for success of the intervention?
5. What were the barriers to delivery?

Ethics and evaluation registration

The evaluation was reviewed by RAND U.S. Human Subjects Protection Committee (HSPC) and approved on 12 June 2018.

The study was not registered as this is not an RCT. There were no Memorandums of Understanding for the evaluation given data for this evaluation was at school level and therefore, did not require collection of personal, identifiable information. Teacher retention data, which was received from DfE, was at the teacher level but was de-identified before being shared with RAND.

None of the evaluation team had any conflicts of interest and all members of the study team approved this protocol prior to publication.

Data protection

School staff invited to take part in interviews or surveys were provided with a project information sheet and Privacy Notice to ensure that participants make an informed choice to participate in the research. Oral consent was obtained from school staff participating in the interviews. For the online surveys, we regarded the completion of the survey as adequate evidence of consent (implicit consent).

As part of the interviews in case study schools, RAND obtained personal data from school staff (i.e. email and name) in order to contact participants. We collected this data on the basis of legitimate interest. RAND acted as data controllers in this capacity. The monitoring data shared by STEM Learning was provided at school level and anonymised, therefore not personally identifiable. The data was shared via Syncplicity, a secure file sharing service, compliant with the General Data Protection Regulation.

This publication includes analysis of the National Pupil Database (NPD) and the School Workforce Census (SWC). https://www.gov.uk/guidance/data-protection-how-we-collect-and-share-researchdata.

The DfE is responsible for the collation and management of the NPD and SWC, and is the data controller of NPD and SWC data. Any inferences or conclusions derived from the NPD and/or SWC in this publication are the responsibility of RAND Europe and not DfE.

This work contains statistical data from the Office for National Statistics (ONS) which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets, which may not exactly reproduce National Statistics aggregates.

RAND Europe adopts good industry practices regarding the protection of personal data as part of its obligations as a data controller under the Data Protection Act 1998 and takes appropriate technical and organisational measures conformant with ISO 27001 to protect personal data. Individuals targeted by the study have the right to oppose, have access to, rectify, or remove personal or sensitive personal data held by RAND Europe.

Project team

Recruitment and delivery team: STEM Learning

- Project officer: Erin Grey (STEM Learning) (previously Andy Thirwell).
- Aspire to STEM lead: Wayne Jarvis (STEM Learning).
- Research and evaluation lead: Ben Dunn (previously Dr. Irina Kudenko).
- Research assistant: Ben Dunn, Erin Grey.

Evaluation team: RAND Europe

- Project leader: Elena Rosa Speciani (RAND Europe) (previously Miriam Broeks, Dr Alex Sutherland, and Sonia Ilie all previously RAND Europe).
- Project manager: James Merewood (RAND Europe) (previously Miguel Subosa, Andreas Culora, Miriam Broeks, Hannes Jarke, and Yulia Shenderovich, all previously RAND Europe).
- Reporting team: Elena Rosa Speciani, James Merewood, Louis Hodge, Maria Jose Guevara, Helen Murphy (RAND Europe), and Miguel Subosa (previously RAND Europe).
- Core fieldwork and analysis team: Miriam Broeks, Eleftheria Iakuvidou, Sashka Dimova, Jack Pollard, Judith Ajebon, Giovanni Amodeo, Sonia Ilie, Lillian Flemons, Miriam Broeks, and Andreas Culora (all previously RAND Europe).

Advisors

- Dr Kata Mihaly (RAND).

Methods

Evaluation design

Table 3: Evaluation design

Evaluation design		Weighted difference-in-differences using kernel-based propensity scores to generate a matched sample
Unit of randomisation		School
No. of units included in analysis (intervention, comparison)		115 secondary schools; 302 comparison secondary schools
Outcomes	Variables	1. Pupil progression 2. Science teacher attrition
	Measures (instrument, scale, source)	1. Number of pupils entered STEM A levels two years after leaving key stage 4, aggregated to the school level, NPD 2. a) Science teachers leaving the school (turnover), aggregated to school level, SWC b) Science teachers leaving the profession (wastage), aggregated to school level, SWC

The impact of Aspire to STEM on key outcomes was evaluated using a quasi-experimental design with analysis at the school level, using administrative data. This is the first evaluation of its kind for this intervention and compares changes in outcomes over time between schools that enrolled in Aspire to STEM and eligible schools that did not enrol.

All schools that had joined the programme by 16 October 2018 were considered enrolled in the programme ('quasi-randomisation' date; Anders, 2017). All schools enrolled on the quasi-randomisation date were included in the analyses as 'treatment' schools, even if they dropped out from delivering the programme. Although the propensity score matching (PSM) process reduced the analysis sample to 73 treated and 208 control secondary schools. Schools were eligible to join the programme after the quasi-randomisation date but were not included in the analysis. Counterfactual schools were schools that did not participate in Aspire to STEM. They did not receive any incentives as there were no evaluation requirements of them due to the use of administrative data for the evaluation.

To ensure the feasibility of intervention roll-out, recruitment was staggered over two time points: recruitment for Cohort 1 started in Spring Term 2018; and recruitment for Cohort 2 started in Autumn Term 2018. Schools were placed in partnerships with other schools in the same cohort, and all schools that had enrolled by Autumn Term 2018 were considered enrolled on Aspire to STEM for the purposes of the evaluation. Both interventions ended on 31 March 2020. We analyse both cohorts together, without differential treatment timing, in order to remain closer to the original planned analysis.

The intended outcomes for the study reflected both the aims of the intervention (see Theory of Change in Figure 1 above) and the anticipated long-term benefits for pupils arising from having more skilled/better resourced STEM teachers and leaders working in schools. The timelines for anticipated outcomes were selected in consultation with STEM Learning based on their experience. However, while the primary outcome was designed to be school-level pupil attainment in STEM subjects, due to disruptions arising from the COVID-19 pandemic, Key Stage 2 and Key Stage 4 outcome data was not available for analysis. **The final study outcomes are therefore the original secondary outcomes—enrolment in STEM A levels,[16] and teacher turnover and wastage** (see 'Outcomes' section for further discussion). To capture behaviour

[16] STEM AS levels was also considered but the relatively small number of pupils taking AS levels compared to A levels would have introduced bias.

changes as a result of the pandemic, additional IPE data collection was conducted in 2024. Further discussions of the impact of Covid-19 on the evaluation can be found in the amended study plan.[17]

The original study plan outlined the use of the Toolkit for Weighting and Analysis of Non-equivalent Groups (TWANG) algorithm to conduct the matching (Griffin *et al.,* 2014). However, in practice, the use of the TWANG algorithm presented two issues:

i) it could not be performed with the required variables as the RAND-propriety R package could not easily be downloaded and run within the ONS Secure Research Service (SRS) environment, and

ii) the algorithmic approach meant researchers had less flexibility to prioritise the closeness of a match on key variables of interest to the Education Endowment Foundation (EEF), such as proportion of students eligible for free school meals (FSM).

The practical issues with TWANG became insurmountable, so RAND Europe and the EEF updated the matching approach. Following recommendations made in Anders *et al.* (2017), different propensity score matching strategies were tested and kernel matching was selected as it resulted in the smallest mean absolute standardised percent of bias and largest sample size, compared to the alternatives considered.[18] Matching algorithms were implemented using the *psmatch2* command in Stata. Further details of the new approach can be found in Appendix C.

After undertaking the propensity score matching process, the remaining sample of primary schools was too small and the analysis of teacher outcomes in primary schools would have been substantially underpowered. We therefore, made the decision, in consultation with the EEF, to only analyse teacher outcomes for secondary school STEM teachers. Further details on these changes can be found in the second amended study plan.[19]

Finally, during analysis it became clear that data was unavailable to complete the compliance analysis that had been intended (see 'Analysis in the presence of non-compliance' section) and some small changes needed to be made to the robustness tests due to lower than expected sample sizes (see 'Robustness test' section).

Participant selection

Aspire to STEM was implemented as part of TLIF, targeting schools located either in Opportunity Areas or in other LADs ranked among the lowest-performing in England, subject to additional school-level eligibility criteria based on Ofsted ratings. To be eligible for Aspire to STEM, schools in these areas had to be rated 'Requires improvement' or 'Inadequate' (3 or 4) in their most recent Ofsted inspection (see 'Intervention' section for further details).

During recruitment 1,031 eligible primary schools and 417 potential secondary schools (including middle schools deemed secondary) were identified. In total, 92 primary schools and 115 secondary schools were enrolled in Aspire to STEM, leaving up to 939 primary schools and 302 secondary schools for comparison. All schools that were eligible for participation were included in the sample available for matching, meaning that all schools in the final analysis were taken from the same areas.

Aspire to STEM is a whole-school intervention, meaning that all pupils and all teachers participated where Aspire to STEM was active. To define a STEM teacher we followed the definition in the original study plan, used by Allen and Sims (2017), where in secondary school, a STEM teacher is someone who in the focal period (in this case, during Aspire to STEM recruitment period, academic year 2017/2018) has taught science for greater than or equal to half of their timetabled teaching hours, and spent at least one hour a week teaching science (identified in the SWC). In the context of this evaluation, 'science' refers to, biology, chemistry, and physics, consistent with how science teachers are defined in the retained

[17] https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/Aspire_to_STEM_Study_Plan_June2020_postcovid_7May21_v2_final.pdf?v=1763977836

[18] Methods under consideration were: i) 1 nearest neighbour with no replacement; ii) 2 nearest neighbours; iii) 1 nearest neighbour with no replacement and a calliper equal to 1/4 of; iv) kernel matching; and v) coarsened exact matching.

[19] https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836

literature. This definition does *not* include other STEM-related subjects such as mathematics, computer science, engineering, or broader technology specialisms, which are classified separately in workforce data. This risks underestimating the impact of Aspire to STEM on the wider STEM teaching population (e.g. mathematics, engineering, and technology), however, this was how the original study team defined STEM teachers and is what was maintained for this evaluation. For the purposes of transparency, we refer to these teachers as science teachers throughout the report, as opposed to STEM teachers as was used in the study plans.

Outcome measures

Pupil outcomes

Aspire to STEM's Theory of Change (see Figure 1) hypothesises that students will have raised aspirations, which will improve pupil outcomes. During the set-up phase the EEF, the evaluation team, and STEM Learning agreed that aspiration would be measured by progression to STEM A level courses.

While drawing on administrative data offers a practical approach to including many schools in the evaluation, one limitation of the approach is that, focusing on the school-level results, we are not able to account for student turnover and examine to what extent the results are influenced by composition effects. Ideally, the analyses would control for prior attainment of the same cohort of students rather than controlling for earlier attainment at that age in the same school. However, this is not possible with the data available for the evaluation.

Originally, progression to STEM A level[20] was intended to be captured for pupils who undertook their A levels at a Key Stage 5 setting, which was associated with a key stage 4 setting in our sample (e.g. a sixth form college). However, constructing progression to STEM A level in this way means that the portion of the sample which attended a key stage 5 setting which was not associated with a key stage 4 setting in our sample would be lost. Therefore, a joint decision was made between RAND Europe and the EEF to instead capture the A level results of all pupils who attended a key stage 4 setting in our sample. Therefore, the progression to STEM A level represents the proportion of pupils who completed a STEM A level regardless of the Key Stage 5 setting they attended.

At the time of inception, reforms were being made to AS and A levels.[21] These reforms decoupled AS and A levels, meaning that AS grades no longer contributed to A level grades. The latest updated study plan[22] outlined that we would 'aim to incorporate AS levels' alongside A level progression for the same cohort of pupils, however, at analysis stage enrolment of STEM AS levels was reviewed and it was deemed to be significantly lower than those enrolled in STEM A levels. Specifically, between 2019 and 2021 (the two academic years of interest for pupil outcomes) 451,248 pupils enrolled in STEM A levels compared to 47,032 pupils in STEM AS levels.[23] The unequal allocation creates challenges for interpretation as the small number of pupils enrolled in AS levels renders this analysis underpowered. Therefore, to minimise bias, aid interpretation, and maintain an adequate sample size, the analysis for pupil-level outcomes was conducted using STEM A levels only.

Teacher outcomes

In line with Aspire to STEM's Theory of Change (see Figure 1) we explored the extent to which the programme affects the percent of secondary school science teachers staying in the same school. Two types of non-retention were defined:

- *turnover* (teachers leaving their current school); and
- *wastage* (teachers leaving the profession entirely).

[20] Biology, chemistry, physics, psychology, environmental science, geology, marine science, mathematics, further mathematics, computer science, applied information and communication technology, digital media and design, and software systems development.
[21] See: **www.gov.uk/government/publications/get-the-facts-gcse-and-A level-reform/get-the-facts-as-and-A level-reform**
[22] See: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/aspire-to-stem
[23] See: **https://explore-education-statistics.service.gov.uk/data-catalogue/data-set/10aeadde-eece-4e45-a9b8-8b6c2fbd5124**

The original study plan intended to include primary school teachers as well, but the low sample size resulting after matching made this analysis underpowered and so primary school teachers were excluded from the analysis. Further details regarding these changes can be found in the second amended study plan.[24]

Our analysis examined whether a science teacher employed at a school during the 2018/2019 academic year (as recorded in the November 2018 SWC) had left that school by the same point in the following academic year (2019/2020, November 2019 SWC). This definition allowed for consistency with baseline weighting variables (wastage/turnover in the previous year) with teacher-level information on wastage and turnover aggregated to the school level. Specifically, teacher *turnover* for school *m,* in period *t,* was defined as the proportion of science teachers who are no longer teaching in school *m* in period *t.* Teacher *wastage* for school *m,* in period *t,* was defined as the proportion of science teachers employed in school *m* in period *t-1* who were no longer teaching in any state-maintained school in period *t*. As the SWC does not include independent schools, any teacher moving to the independent sector is considered equivalent to leaving the teaching profession (at least in the post-period of our analysis).

Sample size

There are few clear guidelines on power calculations for quasi-experimental designs (Hedges, 2017). The study plan based the minimum detectable effect size (MDES) calculation for schools on two scenarios:

1. A 1:1 allocated individually RCT, 224 schools in total (112 schools in the treatment and control groups, respectively).
2. A 1:2 allocated individually RCT, 336 schools in total (112 schools in the treatment group, and 224 in the control group of 224).

In both scenarios, an alpha of 5% and a desired power of 80% were used. It was also assumed that 40% of variance in the outcome was explained by covariates. Power calculations were completed using PowerUp! (Maynard and Dong, 2013). Table 4 outlines the results. For a cluster randomised trial with a 1:1 allocation ratio (scenario 1), the MDES would be 0.291. In the second scenario (1:2 allocation), the MDES would be 0.253.

In the end, 115 secondary schools were enrolled in Aspire to STEM, leaving 302 schools for comparison. This is equivalent to a 1:2.6 allocation, which using the same set of assumptions as the study plan results in an MDES at analysis of 0.219.

Table 4: Minimum detectable effect sizes based on study design and analysis for secondary schools

		Study plan		Analysis
Allocation ratio		1:1	1:2	1:2.6
MDES		**0.291**	**0.253**	**0.219**
Percent of variance in outcome explained by covariates	School level	0.40	0.40	0.40
Alpha		0.05	0.05	0.05
Power		0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two
Number of schools	Intervention	112	112	115
	Control	112	224	302
	Total	224	336	417

Notes: MDES is based on hypothetical RCTs. School-level variance is assumed in all cases.

Propensity score matching

Defining predictor variables

[24] https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836

The propensity score is defined as the 'conditional probability of being exposed [to an intervention] given the observed covariates, $e(X) = P(Z = 1 \mid X)$' (Williamson *et al.,* 2011: p. 278). Propensity scores can be calculated in different ways, but a common approach is to use a logistic regression model:

$$Logit(\pi) = \alpha + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_k X_k + \varepsilon \qquad \text{Equation 1}$$

Where $Logit(\pi)$ is the log-odds of the probability that the outcome (treatment Yes/No) is equal to 1, expressed as a linear function of the predictors $X_1, X_2 \ldots X_k$ and ε (prediction error). The resulting prediction from this model, ranging between 0 and 1, constitutes the propensity score.

The primary goal in selecting predictor variables is identifying variables that are thought to be predictive of either treatment or outcome, or both. For this evaluation we included: i) school-level prior attainment; ii) being in an Opportunity Area or LADs 5 or 6; and iii) variables associated with STEM attainment (see Table 5).

Table 5: Variables associated with STEM attainment included in the propensity score matching

Variable(s)	Notes
School-level key stage 4 maths and science pupil academic attainment before the intervention	Due to the changes in the GCSE scoring from A*–U to 9–U, scores are standardised before averaging.
% FSM pupils (2017/2018)	Existing reviews of literature suggest that disadvantage and literacy are important predictors of science attainment (Nunes *et al.,* 2017)
% of English as an Additional Language (EAL) pupils (2017/2018)	
School type (2017/2018)	E.g., maintained, academy, etc.
Same-sex or mixed school (2017/2018)	There is some indication that single-sex and co-educational schooling could affect pupil attainment (Pahlke *et al.*, 2014)
School size (2017/2018)	Number of pupils attending
Other interventions implemented by STEM Learning in the past two years (2015/2016 and 2016/2017)	STEM Learning to provide a list of schools taking part in interventions of sufficient intensity to be comparable to Aspire to STEM—given the wide reach of this organisation, it is important to account for their ongoing interventions in potential comparison schools

We had initially intended to also include teacher attrition from the previous academic year as a predictor variable. but it was not possible to match data from the SWC (teacher outcomes) to data in the NPD (pupil data). For the teacher *wastage* and teacher *turnover* outcomes we conducted the primary analysis with the same counterfactual group used for pupil-level outcomes, to aid interpretation allowing for us to draw conclusions across the same schools for both the two outcomes. However, we have added an extra robustness check by re-running analysis of the teacher outcomes with an alternative counterfactual group based on propensity scores that are generated with the additional inclusion of teacher-level covariates.

Matching using the propensity score

Initially, it was proposed that matching would be conducted using TWANG (see 'Evaluation design' section above and previous study plans[25] for more details). However, this was found to not be usable inside the ONS SRS and so following discussions with the EEF, it was decided that a pragmatic approach would be to follow the guidance in Anders *et al.* (2017) and the approach taken by other quasi-experimental design evaluations for the EEF (Anders and Jerrim, 2021; Hodgen *et al.,* 2019). This involved testing the balance between treatment and counterfactual groups using the same set of variables outlined in Table 5 above within the matched sample using:

- 1:1 nearest neighbour matching without replacement;

[25] See: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836

- 1:2 nearest neighbour matching;
- 1:1 nearest neighbour matching with multiple calliper widths;
- kernel matching; and
- coarsened exact matching.

Based on these analyses we decided to use kernel matching in the final analysis as this produced the smallest mean absolute standardised percentage of bias and generated the largest sample size. More information on the matching approach is provided in Appendix C, and full details of the approach and an assessment of the standardised percentage of bias reduction under each method can be found in the updated study plan.[26]

Statistical analysis

Difference-in-differences

Once the weights for schools were identified, we then undertook a weighted difference-in-differences analysis (see Woolridge, 2012). The difference-in-differences estimator compares the difference in outcomes before and after the intervention for intervention schools to the same difference for comparison schools, with each observation weighted according to the propensity score. This reduces bias in the post-intervention differences between the treatment and comparison schools that could be the result from time-invariant differences (both observed and unobserved) between these two groups. Time-varying factors, such as nationwide changes in grading approaches, will affect both comparison and intervention schools similarly, so any difference in trends should be attributable to the intervention (i.e. the only time-varying variable that differentially affects one of the groups should be the intervention). Furthermore, multi-year data likely offers increases in statistical power by providing multiple observations of the school performance over time (McKenzie, 2012).

As any design, difference-in-differences relies on a number of assumptions, such as Stable Unit Treatment Value Assumption, meaning that the composition of intervention and comparison schools is stable over time and there is no spill-over (Stuart *et al.,* 2014; Wing *et al.*, 2018). A key assumption of difference-in-differences, which we examine visually and test statistically in Appendix E is that the difference between treatment and comparison schools is constant over time (i.e. parallel trends). For difference-in-differences it is not essential that pre-intervention trends are the same, merely that the trend lines are parallel. The initial study plan proposed a two-period analysis.[27] However, given data was available for more time periods, we chose to draw on the full longitudinal panel data available. A difference-in-differences can be formulated as a Two-Way Fixed Effects (TWFE) regression model of the following form:

$$y_{st} = \alpha + \eta_s + \theta_t + \gamma D_{st} + \boldsymbol{X}'_{st}\beta + \varepsilon_{st} \qquad \text{Equation 2}$$

where y_{st} is the outcome for school s at year t, η_s are school fixed effects, θ_t are time fixed effects, D_{st} is a treatment indicator (1 if treated, 0 otherwise), $\boldsymbol{X}'_{st}$ is a vector of school-level characteristics, and ε_{st} is an idiosyncratic error term. The coefficient of interest is γ and provides a causal estimate of the average treatment effect. School-level characteristics included as covariates were based on the list of weighting variables to provide a 'double-robust' estimation.

Robustness checks

As per Anders *et al.* (2017) we had planned to assess sensitivity to model specification by re-running analyses using different specifications. We outline each of the robustness tests undertaken in this evaluation below.

---

[26] https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836
[27] See: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/aspire-to-stem

First, we tested generating a matched comparison group using a modified set of variables when weighting schools; we tested whether there were differences in the quality of the matched sample when using pre-intervention school-level average attainment outcomes or percentage attainment outcomes. The updated study plan and Appendix E demonstrate that a lower mean standardised percentage of bias is achieved when using average attainment outcomes, and so this matching strategy was taken forward for analysis.[28]

Second, we tested an alternate specification of the analysis model for our secondary outcomes (teacher turnover and teacher wastage) which includes pre-intervention level means of both teacher outcomes in the propensity score matching strategy. The results of this robustness test aimed to demonstrate that the original matched sample, which was generated excluding teacher-level outcomes, was of a high-match quality. It was not possible to recreate this robustness test on analysis which uses pupil outcomes (i.e. the primary analysis), as teacher-level outcomes were not available in the SRS space which was used for pupil-level analysis.

Third, we ran placebo tests of the difference-in-differences specification. In these models we artificially implement the intervention one year early. This allowed us to validate the parallel trends assumption, ensuring the estimated treatment effect is not driven by pre-existing differences, random chance, or confounders.

Finally, we ran a single-level outcome regression with pre-intervention values as covariates, i.e. we conducted analysis using the outcome measure in 2018/2019, while controlling for a pre-intervention measure of the same outcome (in 2016/2017) within the regression.

Unfortunately, during the analysis phase several of the options outlined in the study plan became untenable.[29] We had planned to compare the intervention group to alternative control groups to control for differences in school motivation to engage in a demanding CPD programme by looking at two groups:

1. eligible schools not approached by STEM Learning, by the quasi-randomisation date; and
2. eligible schools approached by STEM Learning and refused/unable to participate, by the quasi-randomisation date.

However, there are a small number of schools that were not approached for recruitment (but met the criteria to participate) (1), and these schools are much more likely to be more remote and geographically isolated than the intervention schools. We also cannot rule out that they would have been interested in taking part had they been invited to do so. Additionally, there were a relatively small number of the schools approached by STEM Learning (2) that did not participate (n=41), had some had instead taken part in other TLIF interventions, including the intervention offered by Institute of Physics.

Analysis in the presence of non-compliance

Our suggested strategy to determine compliance, developed in discussions with STEM Learning, was to create a continuous indicator primarily focusing on dosage, namely: i) the amount of spending of the budget (which can be accessed at school level)[30]; and ii) the amount of teacher and leadership CPD, as these are the key activities of the programme. As outlined in the study plan,[31] we planned to combine these measures at the *school* rather than *partnership* level to ensure sufficient power.

However, it was not possible to access information on school-level budget spending at a detailed enough level to construct a compliance metric. Two forms of data were intended to inform compliance analysis on costs: i) information from STEM Learning on the breakdown of costs; and ii) data on additional spending taken from the headteacher survey. Unfortunately, the most granular level that a breakdown of costs was available for was at the cohort level. This data can be split by local or

---

[28] See: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836
[29] See: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836
[30] A budget indicator at school level would be weighted by the number of pupils, i.e. school size, such as X1= N1* Y/(N1+N2+N3), where X1 is the weighted budget, N1 is the school size, Y is the spending, and N1–N3 are the sizes of other schools in the partnership.
[31] See: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836

regional location, but it cannot be broken down to the school level, which would be required for compliance analysis to take place (since the primary analysis is conducted at the school level). Furthermore, while the data collected as part of the headteacher survey on additional spending is available at the school level, the majority of responses were qualitative in nature (e.g. ‘don’t know’ or ‘more than allocated’).

As a result, it was not possible to construct a cost-related compliance metric from the available data. A decision was made that CPD alone does not constitute compliance, and therefore, compliance analysis was not completed as part of this evaluation.

Missing data analysis

The approach we proposed relied on complete data on the outcomes. The use of school-level outcomes was planned to minimise missingness, which can be greater at the individual level. However, matching schools using longitudinal data brings further challenges in terms of inconsistencies, for example, schools may close and/or reopen over the evaluation period which can make it difficult to track outcomes over time. As outlined in the study plan,[32] if schools changed Unique Reference Number (URN) during the study period, we linked the school over time using data on URN changes from Get Information About Schools (GIAS). As this allows us to generate a dataset which track school URN changes over time, we are able to utilise outcomes from these schools rather than treating them as missing. Despite this, a small number of schools missing some pupil-level data in some years were included in the analysis (see Appendix D for more detail).

Subgroup analyses

This study does not focus on individual pupils, and the Theory of Change does not indicate the percentage of FSM pupils at a school as a key influence, so we did not propose any subgroup analyses.

Estimation of effect sizes

We drew on the Hedges’ g adapted from Hedges (1981) as given below:

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\sigma_S^2 + \sigma_{error}^2}}$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean post-intervention difference between intervention groups adjusted for baseline characteristics (estimated in our case through the difference-in-differences coefficient) and $\sqrt{\sigma_S^2 + \sigma_{error}^2}$ is an estimate of the population standard deviation (SD) at baseline. The effect size therefore, represents the proportion of the population SD attributable to the intervention (Hutchison and Styles, 2010).

In the study plan,[33] we stated that we would draw on the population variance if we were able to use data from all TLIF-eligible schools, and if not, we would draw on the sample-adjusted variance. While an alternative is offered by the What Works Clearinghouse Procedures and Standards Handbook, we were able to calculate the pooled SD, which accounts for the sample used within the evaluation and so the alternate was not needed.

IPE

Research methods

The implementation and process evaluation (IPE) sought to answer the following questions:

1. What was the level of fidelity in the intervention schools?

---

[32] See: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836
[33] See: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763977836

2. What was the level of dosage in the intervention schools?
3. What factors and initial conditions appear to explain variation in the dosage and fidelity of implementation?
4. What appear to be the necessary conditions for success of the intervention?
5. What were the barriers to delivery?

These were in alignment with the EEF guidance at the time (Humphrey *et al.,* 2015) though we note that the EEF guidance was updated before reporting, but after data collection, which means that the IPE is less aligned to the EEF current guidance than would have been the case had the evaluation been commissioned today. As such, this IPE understands fidelity as the extent to which implementers adhered to the intended treatment model, and dosage as how much of the intended intervention was delivered or received by the target participants. However, we note that typical sections now included as standard in the EEF IPE sections (e.g. business as usual, and usual practice) were not collected as part of the IPE for this evaluation.

Methods

The IPE design was informed by the EEF guidelines at the time (Humphrey *et al.,* 2015) and conversations with STEM Learning. An IDEA workshop was conducted before commencing the project to identify the mechanisms by which the intervention was expected to influence pupil and teacher outcomes. These mechanisms were laid out in a Theory of Change (see Figure 1).

To explore the said mechanisms, the IPE drew on the following data collection methods (which are summarised in Table 7 below):

- surveys with teachers, headteachers, and educational leads, conducted at the beginning and near the end of the evaluation;
- interviews and focus group discussions with staff in a sub-sample of eight case study schools;
- focus group discussions with educational leads; and
- review monitoring information collected by STEM Learning (e.g. enrolment in CPD, engagement in activities).

The IPE needed to accommodate the bespoke nature of Aspire to STEM, given that CPD events and other activities implemented as part of the programme varied depending on the needs and capacity of each partnership.

Given the quasi-experimental nature of the study, the IPE only included schools that participated in Aspire to STEM. Four schools were selected from each of the two cohorts as case study schools (hence, eight schools in total). Case study schools were selected based on the criteria indicated in Table 6 below.

Table 6: Criteria used to select case study schools

		Similarity of priorities across the partnership	
		Similar priorities	Different priorities
Duration of partnership	Previously existing partnership	Two case study schools	Two case study schools
	New partnership	Two case study schools	Two case study schools

These criteria were chosen to surface differences—or lack thereof—in barriers and facilitators encountered by schools based on the duration of their partnership with other schools and the extent to which partnership schools shared common priorities.
However, schools that dropped out from the programme after the 'quasi-randomisation' date were not included in the IPE. This limits the findings of the IPE significantly in that it could not harvest information from schools who might have found it difficult to implement Aspire to STEM and hence dropped out of the programme. Nevertheless, the IPE provides useful insights for the future implementation of similar programmes.

Table 7: IPE methods overview

Research methods	Data collection methods	Participants / data sources	Data analysis methods
Survey	Online questionnaire	Headteachers in Aspire to STEM schools at the time the survey was deployed / Autumn Term 2018 and Spring Term 2020[a]	Descriptive statistics
		Educational leads / Autumn Term 2018	Descriptive statistics
		Primary teachers and secondary school STEM teachers / Winter Term 2020	Descriptive statistics
		STEM subject leaders (faculty/department heads and senior leadership team [SLT]) / Autumn Term 2019	Descriptive statistics
Interviews	Individual interviews	STEM teachers, headteachers, STEM subject leaders, and members of the SLT—in case study schools / January 2020	Framework analysis
Focus groups	Focus group discussions	Educational leads / Autumn Term 2019	Framework analysis
Document review	Monitoring data from STEM Learning	Educational lead monthly updates and needs assessment / action plan (in case study schools) Action plans from STEM Learning[b]	Framework analysis

[a] *Two headteacher surveys were administered: one in Autumn Term 2018 (October and November) and one in Spring Term 2020 (February and March).*
[b] *Monthly updates were expected, but the evaluation team did not receive those updates.*

The evaluation team administered surveys with teachers, headteachers, and STEM subject leaders at the beginning and near the end of the intervention. These surveys collected information regarding usual practice before Aspire to STEM, as well as difficulties and enabling factors encountered by participating schools during programme implementation.

Administering surveys at two time points allowed for pre- and post-intervention comparison, and doing so with teachers, headteachers, educational leads, and STEM subject leaders allowed the evaluation team to explore how the programme affected stakeholders at different levels of the school structure. This is important since the said stakeholders have different responsibilities in educational provision and will therefore also have had different experiences with Aspire to STEM. For example, a headteacher could provide more high-level information on how Aspire to STEM influenced whole-school operations, while teachers have a better view of classroom-level changes.

Since the programme had started with some schools already assigned to Cohort 1 before the evaluation even began, a 'true' pre-intervention survey was impossible to administer. Nevertheless, the baseline survey asked headteachers about practices predating Aspire to STEM, particularly with respect to CPD and STEM engagement. This allowed the evaluation team to understand what usual practice looked like before the programme was delivered. However, given the turnover of school staff over the course of the project (spanning three academic years), follow-up surveys could not be conducted entirely with the same school staff. Despite this, and for purposes of comparison, follow-up surveys still asked if the respondent had completed the baseline survey.

Surveys were sent by STEM Learning on behalf of RAND Europe. While schools may have possibly been less inclined to candidly and transparently describe their experiences of the programme to the delivery team, the approach taken was the most practical one, given that the evaluation followed a quasi-experimental design, RAND Europe did not have a direct relationship with schools. This lack of relationship between RAND Europe and participating schools meant that there was a higher risk of schools not responding to surveys administered by the evaluation team.

In addition, educational leads only responded to a baseline survey. A second survey was planned with educational leads, but this was replaced with focus group discussions held from September to December 2019. This choice was motivated by the need to develop a deeper exploration of barriers and facilitators. While surveys provided an initial glimpse into barriers and facilitators, focus group discussions allowed for further probing into participants' experiences with the programme.

Interviews and focus group discussions with teachers, headteachers, STEM subject leaders, and educational leads across the case study schools allowed the evaluation team to elaborate on information reported by survey respondents. As such, interviews and focus group discussions prompted participants to describe barriers and facilitators in a detailed manner. The insights gathered from these conversations with stakeholders not only provided the evaluation team with a picture of how adherent Aspire to STEM was to its intended design, but it also surfaced factors that might have led schools to diverge from the said design.

Moreover, the content of these discussions differed based on the role of the participant within the school structure. For instance, teachers and STEM subject leaders were asked about the programme's contributions of Aspire to STEM towards improving teaching practice and any changes they had observed in their students due to their school's participation in the programme. On the other hand, discussions with headteachers focused on broader operational issues such as programme-related financing and availability of CPD opportunities for teachers. Educational leads, given their role of facilitating collaboration across schools, were primarily asked about partnership-level issues.

Taken together, the surveys, interviews, and focus group discussions provided a thorough view of schools' experiences with Aspire to STEM. Surveys offered the big-picture view of the programme's implementation, complemented by the more granular and fleshed out first-hand accounts that emerged from interviews and focus group discussions with school-level stakeholders.

Fidelity

As there is no single indicator for fidelity, we relied on surveys with educational leads and headteachers as well as the interviews in case study schools to understand the extent to which the intervention was implemented as intended by the developers (see Table 8). Similarly, when examining the barriers and facilitators of delivery, we explore the functioning of the school partnership as part of the Aspire to STEM model. For instance, we will look at the process of agreement on programme activities within the partnership and the influence of the differences and similarities in the priorities of individual schools.

Table 8: Fidelity indicators

Fidelity indicator	Data	Notes
Each school conducts needs assessment and action plan	STEM Learning records Educational leads surveys	In principle, this is a mandatory component that all schools will complete.
The schools are engaged in the partnership	Educational leads surveys Headteacher, STEM leaders, and teacher surveys	-

Regarding dosage, given the number and diversity of potential intervention activities, there is not one straightforward measure of how much intervention schools receive. However, the evaluation can draw on the monitoring data collected by STEM Learning to paint a picture of the dosage of intervention activities (see Table 9).

Table 9: Dosage indicators

Activity / resource	Data	Notes
Date of school needs assessment submission	Measured by STEM Learning records	Reflecting the effective duration of the intervention
Access to STEM Learning network and resources: STEM Ambassador volunteers, STEM Club infrastructure and community groups and resource eLibrary	Not measured	It was discussed that access to resources would be too difficult to measure, especially as many online materials are available without a login
Careers guidance and information for teachers and leaders	Number of career information events for teachers	Aggregated at school level by STEM Learning
Bespoke CPD for STEM teachers	Percentage of teachers attending CPD activities	Aggregated at school level by STEM Learning
Bespoke CPD for leaders	Percentage of school leaders attending CPD activities	Aggregated at school level by STEM Learning
Inspiration and enrichment activities	Number and type of community events	Aggregated at school level by STEM Learning
The schools remain in the partnership for the duration of the intervention (until March 2020)	STEM Learning records Educational leads surveys Headteacher, STEM leaders, and teacher surveys	Assessed based on self-report by headteachers and educational lead reports on whether all schools continue to attend meetings and respond to communication
Each school participates in at least some intervention activities (CPD, enrichment, etc.)	STEM Learning records Headteacher surveys	-
The partnership uses the available funding by the end of the project	STEM Learning records	-

Analysis

Data collected through the IPE was analysed through a combination of simple quantitative analysis and structured qualitative techniques. Using these approaches, quantitative findings served as a jumping point for more in-depth probing through qualitative analysis.

Baseline and endline survey data was used to generate descriptive statistics (e.g. range, frequencies). These descriptive statistics captured a broad picture of the range of beliefs and practices of stakeholders that participated in the intervention. Findings of the baseline survey were compared against those of the endline survey to surface changes in beliefs and practices between the start and the end of Aspire to STEM. It must be noted, however, that individual responses were treated as nominal variables, and as such, only frequency counts were generated. Therefore, comparisons between baseline and endline only noted increases (or decreases) in the proportion of respondents that responded a certain way to a given question. The quantitative analysis did not seek to explore whether observed changes in frequency counts were statistically significant. However, where there were notable differences in response frequency, the evaluation team explored the issue further through interviews and focus group discussions.

Interviews and focus group discussions were analysed using a framework analysis (see Srivastava and Thomson, 2009). By collating notes taken from discussions with staff members across the eight case study schools, the evaluation team developed detailed case study reports responding to the previously enumerated research questions. Afterwards, these case study reports were coded thematically using a predefined coding framework. The coding framework included both deductively and inductively determined themes.

An initial set of main themes was first identified deductively (see Fereday and Muir-Cochrane, 2006)—that is, based on the elements identified in the Theory of Change (e.g. CPD, STEM Learning resources, STEM careers education). This allowed the evaluation team to extract information that directly pertained to mechanisms that were anticipated to underlie implementation of Aspire to STEM. Other deductive themes included fidelity and dosage, which helped the evaluation team extract information regarding schools' adherence to the Aspire to STEM model.

However, while coding began deductively, the evaluation team allowed for inductive coding as well (see Fereday and Muir-Cochrane, 2006). This was especially important in surfacing factors that were not anticipated by the Theory of Change but nonetheless affected implementation of Aspire to STEM, as well as the unintended consequences of implementing Aspire to STEM. Examples of these themes included delays in programme set-up and partnership establishment, and some tasks needed to be carried out by school leadership, which were both found to be critical determinants of programme success.

In addition to these thematic codes, the evaluation team also coded each text extract as portraying either a barrier or facilitator to programme implementation. This allowed the evaluation team to sort through the data in a systematic manner, identifying barriers and facilitators that might have specifically influenced a given element of the Theory of Change.

The coding framework was also used to thematically categorise activities laid out in the action plans and other partnership monitoring documents provided by Aspire to STEM. For instance, activities could be coded into categories such as CPD, careers education, and community engagement, among others. The information extracted from these monitoring documents informed analysis of the extent to which partnership schools engaged with different programme components.

Given this analytical approach, the results of the IPE are also reported in a thematic fashion. This thematic approach allowed the evaluation team to drill down on the most crucial antecedents of programme effectiveness.

Cost evaluation

The cost evaluation was developed before the EEF had established current cost evaluation guidance. It was meant to provide an estimate on the cumulative cost of implementing Aspire to STEM at the school level and focused on the following categories:

- **Prerequisite costs.** Costs linked to setting up the intervention in schools (e.g. cost of consultancy process involved in conducting needs assessments and developing action plans).

- **Operational costs.** Costs associated with running the intervention (e.g. CPD and associated cover costs).

- **Any additional costs.** Costs associated with set-up and operation of the intervention that were not covered by the partnership fund.

The plan had been for cost data to come from administrative data collected by the delivery team at STEM Learning as well as through a survey question posed to the SLT in the endline survey, following the EEF guidance at the time. Administrative data was meant to give a relatively comprehensive overview of costs at the school level, while the endline survey focused on collecting additional costs incurred by schools' participation in the intervention.

However, data from STEM Learning was provided to the evaluation team at the partnership (operating across multiple schools) rather than individual school level. Data obtained outlines how much money was spent by each partnership in each cohort across a number of categories. It should also be noted that we did not have a 'business as usual' group as this evaluation was a quasi-experimental design and we had no actively involved comparator group.

We did not observe the flows of partnership funding to individual schools. Additionally, information is not available on which pupils were likely beneficiaries of the intervention and while Aspire to STEM is intended to be a whole-school intervention, not all pupils at participating schools were necessarily beneficiaries of the intervention. We were, therefore, not able to provide accurate figures for the average cost per school or pupil. However, we have calculated *indicative* averages. Using the average number of pupils per school at the January 2019 School Census, and assuming that all pupils in the 207 treatment schools (92 primary, 115 secondary) were beneficiaries of Aspire to STEM, we estimated the average cost per pupil to be £7.29, or £4,805 per school.

Timeline

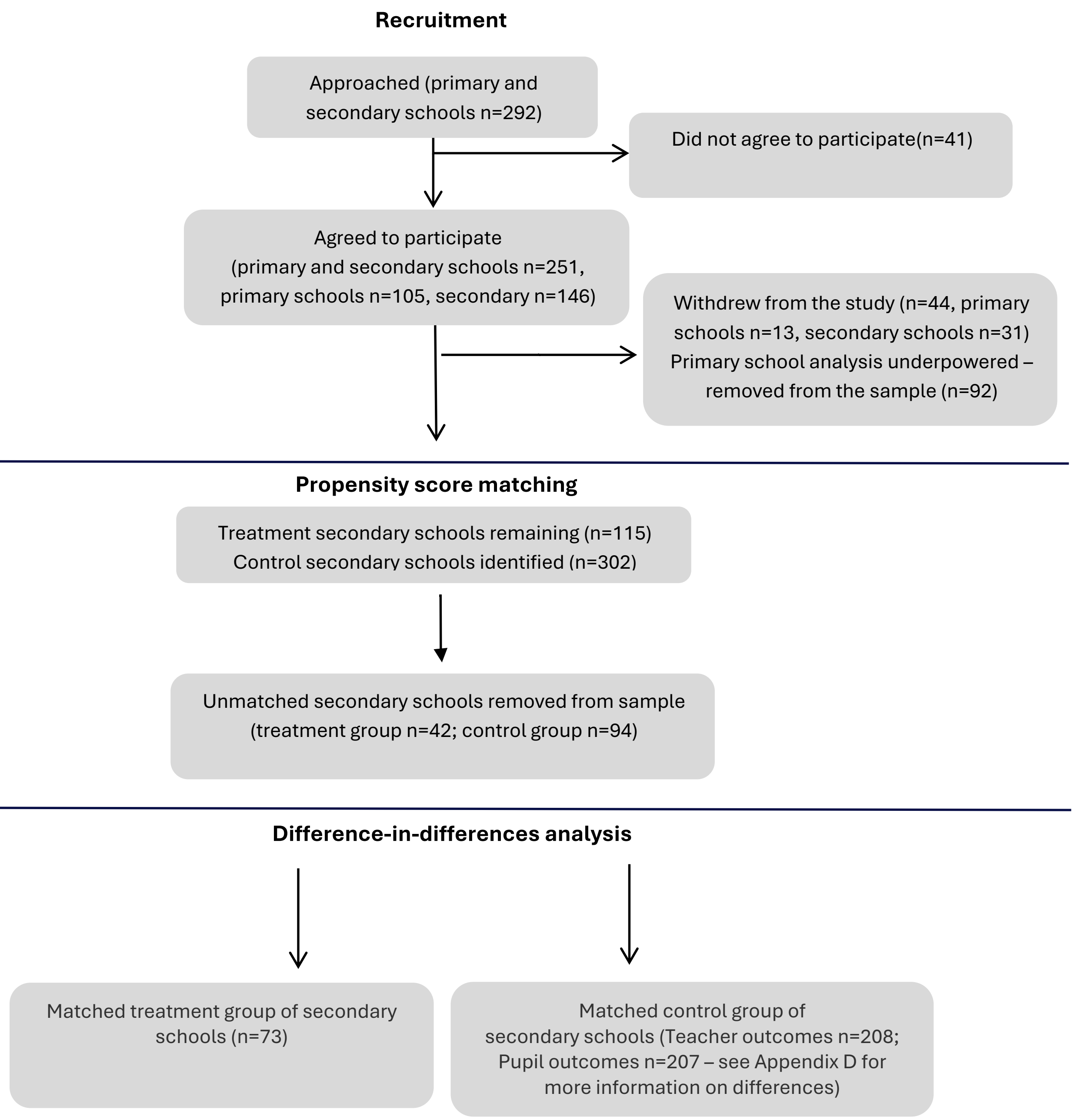
Table 10: Timeline

Dates	Activity	Staff responsible / leading
February 2018	Evaluation kick-off meetings	Bélanger (RAND)
February 2018	Intervention (needs assessment) begins in Cohort 1	Thirwell (STEM Learning); Jarvis (STEM Learning)
September 2018	Recruitment of schools completed	Thirwell (STEM Learning); Jarvis (STEM Learning)
September 2018	Intervention begins in Cohort 2	Thirwell (STEM Learning); Jarvis (STEM Learning)
October 2018 – November 2018	Initial surveys of headteachers and educational leads (Cohorts 1 and 2)	Shenderovich (RAND)
December 2018 – July 2019	Collation of weighting dataset for primary outcomes (pupil key stage 2 and key stage 4 outcomes)	Sutherland (RAND); Shenderovich (RAND)
November 2019 – December 2019	Teacher survey, educational leads survey 2, STEM subject leader survey, interviews in case study schools	Shenderovich (RAND)
February 2020	Headteacher survey 2	Shenderovich (RAND)
March 2020	Intervention completed in Cohorts 1 and 2	Thirwell (STEM Learning); Jarvis (STEM Learning)
May 2020	Sharing of STEM Learning monitoring data	Grey (STEM Learning); Jarvis (STEM Learning)
June 2021 – July 2021	Obtain teacher retention data from SWC data	Broeks (RAND)
September 2021	Updated study plan, post-Covid	Broeks (RAND)
January 2022 – June 2022	Obtain Key Stage 5 progression data (STEM subjects) for students sitting Key Stage 4 exams in 2019	Broeks (RAND)
July 2023 – May 2024	Analysis of parallel trends in SRS, exploring use of TWANG and alternatives to matching	Merewood (RAND)
June 2024	Submission and publication of revisions to the study plan	Merewood (RAND)
July 2024	Begin application for additional Key Stage 5 outcomes with the NPD/SRS teams	Merewood (RAND)
July 2024 – October 2024	Analysis of pupil-level outcomes (timeline dependent on additional data application as a prerequisite)	Merewood (RAND)
November 2024	Submission of draft report to the EEF	Speciani (RAND)
Summer 2025	Submission of the EEF final report	Speciani (RAND)

Impact evaluation results

Participant flow including losses and exclusions

Figure 2: Participant flow diagram



Primary outcome analysis

Under the original study plan,[34] the primary outcome was intended to be student attainment. However, due to the impact of COVID-19 on GCSE testing conditions and scores it was not possible to evaluate changes in student attainment. Therefore, our primary analysis focuses on answering the research question: Does Aspire to STEM lead to a higher proportion of students opting to take STEM subjects at key stage 5, compared to 'business as usual' in secondary schools? We focus on a primary outcome of progression to a STEM A level, which is defined as proportion of pupils who undertook a STEM subject at key stage 5, but associated with the key stage 4 setting they attended. For more information on the process used to match students from key stage 4 settings to their exam outcomes at key stage 5, please see Appendix D.

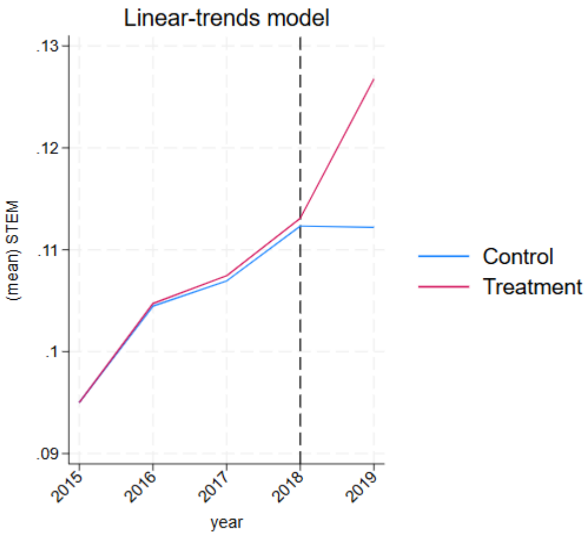
Our results, presented as Hedges' g effect sizes (Table 11), indicate that there was a statistically significant positive effect of Aspire to STEM on pupil progression into a STEM A level. The positive effect size (0.231) is significant at the 1% level, demonstrating a statistically significant effect of the programme. The untransformed regression coefficient (see Appendix B) suggests that pupils in our sample who attended a key stage 4 setting which implemented the Aspire to STEM programme were 1.42 percentage points more likely to sit a STEM subject at A level, two years later.

Table 11: Effect sizes for primary outcome analysis, by pupil-level outcomes

	Outcome	Unadjusted means				Effect size		
		Intervention group		Control group				
		Observations	Mean (SD)	Observations	Mean (SD)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Pupils	Progression to STEM A level	73	0.107 (0.062)	207	0.109 (0.061)	280 (73:207)	0.231 (0.111 – 0.351)	0.007

These results become even clearer when examined visually. Figure 3 visualises changes in the progression to STEM A level in the treatment and control group over time with means taken from the linear trends model (produced during the difference-in-differences estimation). There is a visible increase in progression to STEM A level for students who attended a treatment school immediately following the intervention, in contrast to very similar trends before the intervention began.

Figure 3: Longitudinal trends of progression to STEM A levels for treatment and control groups



[34] See: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/aspire-to-stem

Secondary outcome analysis

Our secondary outcome analysis focuses on answering the research questions related to science teacher outcomes, namely:

- Does Aspire to STEM lead to fewer (STEM) teachers leaving their school (teacher turnover), compared to 'business as usual' in secondary schools?
- Does Aspire to STEM lead to fewer (STEM) teachers leaving the profession (teacher wastage), compared to 'business as usual' in secondary schools?

We assess each outcome in turn, first teacher turnover and second teacher wastage.

Teacher turnover

We defined teacher turnover as the proportion of teachers in a school who are no longer teaching at the school in the following year. Using the weighted difference-in-differences design described above, no statistically significant impact of Aspire to STEM on teacher turnover was detected. The Hedges' g effect size was estimated to be -0.003 (p = 0.974) and the 95% confidence interval (CI) shows that the true result may sit on either side of zero. This implies that teacher turnover was not significantly affected by the introduction of Aspire to STEM.

Teacher wastage

We defined teacher wastage as the proportion of teachers in a school who are no longer in the teaching at any state-maintained school in the following years. Using the same design, again, we found no statistically significant impact of Aspire to STEM on teacher wastage. The Hedges' g effect size was estimated to be of 0.020 (p = 0.859). Again, the associated 95% CI shows that the true result may sit on either side of zero. This implies that teacher wastage was not significantly affected by the introduction of Aspire to STEM.

Table 12: Effect sizes for secondary outcome analysis, by teacher-level outcomes

	Outcome	Unadjusted means				Effect size		
		Intervention group		Control group				
		Observations	Mean (SD)	Observations	Mean (SD)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Teachers	STEM teacher turnover	73	0.226 (0.161)	208	0.234 (0.182)	281 (73:208)	-0.003 (-0.350 – 0.344)	0.974
	STEM teacher wastage	73	0.087 (0.090)	208	0.010 (0.105)	281 (73:208)	0.020 (-0.179 – 0.219)	0.859

Robustness tests

Including teacher outcomes as matching variables

Our first robustness test presented in this report aims to demonstrate the high quality of the matched sample, even when teacher-level outcomes are not included in the matching strategy. To test the robustness of the matched sample, we conducted the analysis of teacher outcomes (turnover and wastage) utilising pre-intervention teacher-level outcomes as matching covariates during the earlier matching stage.[35] It is important to note that this approach changes the composition of the matched sample slightly, as it technically constitutes a different matching strategy; this means that the treatment group now contains 74 settings (previously 73) while the control group now contains 205 settings (previously 208).

[35] These covariates are added to the logit model described in Equation 1.

Table 13 shows the results from this robustness test. Once again, p-values are very large (~0.5 for teacher turnover and ~0.9 for teacher wastage) indicating that there is no statistical identifiable effect of Aspire to STEM on either teacher outcome. As before 95% CIs sit on either side of zero, indicating effects are not distinguishable from null.

Overall, these results demonstrate the robustness of our favoured approach to generating a matched sample, as the inclusion of teacher-level outcomes in the matching strategy does not significantly alter results. The same conclusions that there is no statistically significant impact of Aspire to STEM on teacher turnover or teacher wastage, remain.

Table 13: Effect sizes for robustness test including teacher outcomes as matching variables

	Outcome	Unadjusted means				Effect size		
		Intervention group		Control group				
		Observations	Mean (SD)	Observations	Mean (SD)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Teachers	STEM teacher turnover	74	0.223 (0.160)	205	0.232 (0.180)	279 (74; 205)	-0.004 (-0.347 – 0.339)	0.497
	STEM teacher wastage	74	0.087 (0.090)	205	0.098 (0.100)	279 (74; 205)	-0.013 (-0.204 – 0.178)	0.889

Placebo tests

As a second robustness test, we run placebo tests, which artificially implement the policy one year earlier than happened in practise (i.e. in 2016/2017). Placebo tests are intended to demonstrate the robustness of our findings to the underlying parallel trends assumption, which requires that both the treatment and control groups were following similar paths before the intervention began. By artificially coding the intervention one year earlier, i.e. at a time where the intervention should not have had any effect on outcomes because it did not formally exist, we are able to check for anticipation effects in the treatment group.

The results in Table 14 demonstrate that there were no significant identifiable differences between the treatment and control groups across any of our outcomes in our placebo robustness tests (i.e. one year before the intervention began—2016/2017). The effect size for progression into STEM A level is close to zero and not statistically significant. The associated 95% CIs span zero. The effect sizes for the teacher-based outcomes are also small and not statistically significant, as demonstrated by the 95% CIs (which span zero in both cases).

This demonstrates that our analysis and results from the main model are robust, in line with the parallel trends assumption, and do not feature any anticipation effects. This is particularly relevant to our positive findings related to pupil progression to STEM A levels.

Table 14: Effect sizes for robustness placebo tests

	Outcome	Unadjusted means				Effect size		
		Intervention group		Control group				
		Observations	Mean (SD)	Observations	Mean (SD)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Pupils	Progression into STEM A level	73	(0.060)	207	(0.060)	280 (73:207)	0.021 (-0.095 – 0.138)	0.779
Teachers	STEM teacher turnover	73	0.226 (0.161)	208	0.234 (0.182)	281 (73:208)	0.039 (-0.308 – 0.386)	0.200
	STEM teacher wastage	73	0.087 (0.090)	208	0.010 (0.105)	281 (73:208)	0.011 (-0.188 – 0.210)	0.604

Single-level outcome regressions

Finally, we conducted single-level outcome regressions to test the robustness of our main specification. To undertake this robustness test, we ran regressions on the level of each outcome in the post-period to test for differences between the treated and untreated group. This regression featured the standard control set but also included a control for the pre-intervention level of the outcome (i.e. in 2016/2017), to control for differences between the treatment and control groups before Aspire to STEM began. The coefficients of interest from these regression models are shown in Table 15. This is interpreted as the impact of receiving Aspire to STEM on the outcome in the post-period.

Our results show consistency with the findings of the main analysis model presented earlier. When estimating the effect on progression to STEM A levels we again found a positive and significant effect of Aspire to STEM, and of a similar magnitude (0.016 compared to 0.014 in the main analysis). Similarly, for both teacher outcomes we again found small, negative effects of Aspire to STEM, that are not statistically significant from zero. This provides further evidence of positive effects on pupil progression to STEM A levels, but little to no effect of Aspire to STEM on either teacher retention outcome.

Table 15: Regression results for robustness test conducting single-level outcome regressions

Variable	Progression to STEM A level	Teacher turnover	Teacher wastage
Effect of Aspire to STEM	0.016** (0.004 – 0.027)	-0.023 (-0.061 – 0.015)	-0.016 (-0.036 – 0.005)

*Notes: Regression coefficients are shown for key variables within each regression. Results are presented as coefficients with 95 CIs shown in brackets underneath each. Statistical significance is shown in the following way: * p<0.05; ** p<0.01; *** p<0.001.*

Implementation and process evaluation results

The findings presented in this section are based on the qualitative data collected via visits to treatment case study schools, as well as focus groups with education leads and surveys of key stakeholders in treatment schools.

As mentioned on page 23, the IPE sought to answer the following questions:

1. What was the level of fidelity in the intervention schools?
2. What was the level of dosage in the intervention schools?
3. What factors and initial conditions appear to explain variation in the dosage and fidelity of implementation?
4. What appear to be the necessary conditions for success of the intervention?
5. What were the barriers to delivery?

Table 16 lists the data collection activities undertaken to answer the above research questions.

Table 16: Summary of IPE data collection activities

Data collection methods	Stage	Eligible respondents	N, individual respondents	Number of schools / partnerships represented	Total number of schools / partnerships	Non-response rate (at school / partnership level)
Educational leads survey	Baseline	Educational leads	18	18 partnerships	40 partnerships	55%
School survey	Baseline	SLT members, STEM subject leads, and STEM or primary school teachers	26	Majority of respondents did not indicate their school	189 schools	Majority of respondents did not indicate their school
Educational leads focus groups	Endline	Educational leads	18	18 partnerships	40 partnerships	Not applicable[a]
STEM subject leader survey	Endline	STEM subject leaders	44	39 schools	189 schools	79.37%
Headteacher survey	Endline	SLT members	40	38 schools	189 schools	79.89%
Teacher survey[b]	Endline	STEM or primary school teachers	80	50 schools	189 schools	73.54%
Case study interviews	Endline	SLT members, STEM subject leads, and STEM or primary school teachers	14	7 schools[c]		Not applicable[d]
Document review: action plans and monitoring documents provided by STEM Learning	Endline	Case study schools	7	7 schools		Not applicable[e]

[a] *Focus group discussions were purposely conducted only with educational leads who had responded to the baseline survey.*
[b] *It was possible for more than one respondent from each school respond to the teacher endline survey.*
[c] *Eight case study schools were intended, but only seven were recruited.*
[d] *Interviews and the document review were purposely done only with the seven case study schools.*
[e] *Ibid.*

It is important to note that the IPE was affected by several data collection challenges, including relatively high levels of survey non-response at the school and partnership level, lower than anticipated participation in some qualitative activities, and disruption to planned data collection during the onset of the Covid-19 pandemic. These factors reduce the generalisability of the survey findings and mean that results should be interpreted with appropriate caution. Nevertheless, the insights arising from the IPE still shed light on critical issues that affected the delivery of Aspire to STEM. Given the multi-layered analytical approach outlined earlier, the results of the IPE are reported in a thematic fashion with associated research questions referenced in the headlines. These headline findings are discussed in the text that follows.

Factors contributing to fidelity and dosage (IPE research questions 1, 2, and 3)

The intervention was delivered only partially as intended across treatment schools. While all participating schools completed the required needs analysis and action plan and engaged in at least some aspects of the programme, implementation was uneven. Schools engaged most consistently with CPD, which was widely perceived as relevant and well-delivered, and many schools reported meaningful collaboration within partnerships once these were established. Endline survey data reinforce this pattern, showing that 88% of STEM subject leaders and 76% of teachers participated in at least one CPD activity, indicating much stronger engagement with CPD than with other elements of the intervention. However, several other programme components, such as STEM Ambassadors, saw substantially lower levels of engagement, and their delivery diverged considerably from the programme design.

These issues with fidelity stemmed primarily from delays in early implementation, which reduced schools' overall time in the programme by up to six months. The slow completion of partnership milestones, limited clarity around partnership expectations, and heavy administrative requirements created early bottlenecks that impeded the timely roll-out of activities. In addition, school-level resource constraints, including staffing shortages, restricted cover availability, high workloads, and competing priorities, further limited schools' capacity to implement all elements of the intervention as intended. These structural barriers disproportionately affected smaller and disadvantaged schools, contributing to the observed variability in implementation fidelity across the treatment group.

Inefficiencies in meeting early-stage milestones

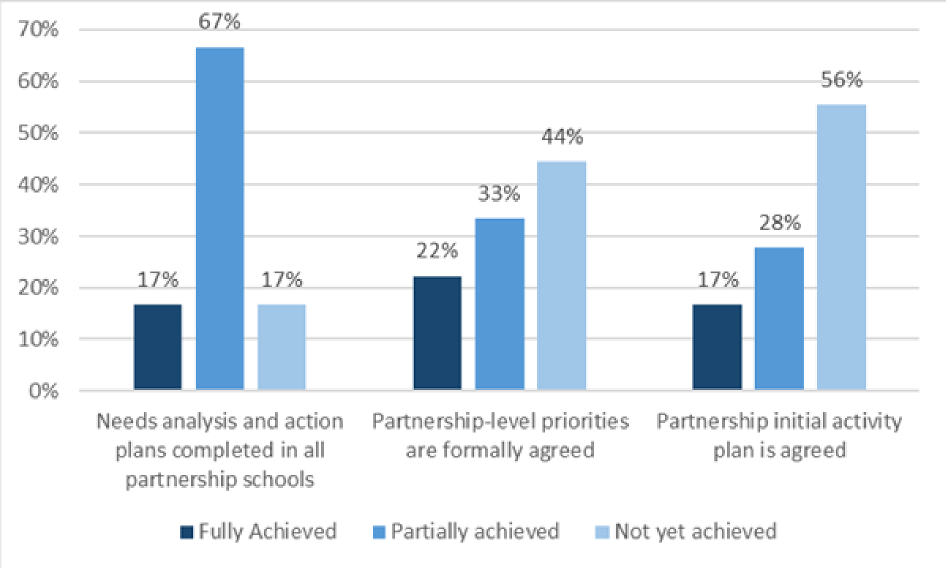
Participating schools unanimously recognised the importance of partnership working but struggled to fulfil the programme's early-stage partnership milestones. This recognition is evident in the observed incidence of partnership working among participating schools: according to endline surveys, 32 out of 38 headteachers reported having collaborated with other schools in their local partnership area. According to staff from case study schools, joint training sessions and regular partnership meetings were the most common platform for inter-school collaboration. However, despite the high incidence of partnership working reported by headteachers, surveys with educational leads revealed that, in the programme's early phases, a considerable number of partnerships came short of meeting the designated partnership milestones, like completing the need analysis and action plan.

Of the 18 educational leads surveyed at baseline, only four reported that the schools in their partnership area had managed to formally establish partnership-level priorities. Six of these respondents reported only partial attainment of the said milestone while eight observed that partnership-level priorities had not yet been agreed at all. Similarly, only three of these educational leads responded that schools in their area had completed their partnership activity plans; five described their partnership activity plans as having been only partially developed while ten reported there being no partnership activity plan in place at the time of data collection.

Schools faced similar delays in conducting a needs analysis and putting together an action plan. Only three out of the 18 educational leads surveyed at baseline reported full compliance with these prerequisites across all schools in their partnership area. In total, 12 of 18 responded that schools in their partnership area had only partially met this milestone, and three reported that the milestone had not been achieved. The document review corroborates these findings, with only two of the seven school action plans provided by Aspire to STEM including a thorough needs analysis that clearly articulated how proposed activities would address identified needs. Two of these action plans sparsely listed the schools' learning needs, whereas the needs analysis section was left completely blank in three action plans.

The low attainment of early-stage partnership milestones is illustrated in Figure 4 below. Despite delays, however, all participating schools eventually completed a needs analysis and developed an action plan since both were prerequisites to programme participation.

Figure 4: Early partnership milestones were, at best, only partially completed within the originally intended time frame



This backlog in meeting partnership milestones may be due to an insufficient time allotment for partnership-building. During focus group discussions, educational leads expressed that an additional six months given to partnership-building before starting programme activities would have helped them develop relationships across partnership schools. In turn, building these partnerships early in the programme's lifespan would have further facilitated inter-school collaboration in the longer term.

However, it must be noted that schools did not seem to have a unified understanding of what constituted partnership working in the context of Aspire to STEM. For instance, during interviews, staff from four case study schools alluded to partnerships that predated Aspire to STEM. While the said interviewees acknowledged how the programme helped strengthen these pre-existing partnerships, their erroneous understanding of partnership working within Aspire to STEM point to potential faults in the way the programme was communicated to participating schools. However, since this evaluation did not probe into this issue, these faults cannot be specified. Nonetheless, this finding highlights a need for Aspire to STEM and similar programmes to ensure that programme components are consistently communicated across different schools.

Highly uneven engagement in different programme components

The delays described above diminished the intended dosage of Aspire to STEM activities, with several schools participating in the programme for up to six months less than intended. For instance, in focus group discussions, two educational leads noted that, despite Aspire to STEM's two-year duration, schools in their partnerships did not receive two years' worth of CPD, due to delays in implementing the programme. Educational leads attributed these delays to heavy administrative procedures but did not specify where the bottlenecks lay. However, even if these discussions did not identify specific bottlenecks, the reduced duration of the programme due to implementation delays surfaces a need to streamline administrative processes.

Note, however, that while overall dosage was diminished, **certain components of Aspire to STEM saw higher levels of engagement than others, with CPD registering the highest engagement among STEM subject leaders and teachers.** Around 37 of 42 STEM subject leaders and 61 of 80 teachers surveyed at endline reported having participated in at least one CPD activity as part of their involvement with Aspire to STEM. Of the 19 teachers who were unable to participate in CPD as part of the programme, 11 teachers described time constraints as being the primary hindrance. Schools' propensity to engage with CPD at a comparatively higher level than other programme components is also reflected in the findings, with CPD activities dominating the action plans submitted by case study schools.

Teachers and STEM subject leaders consistently reported that **CPD was the most visible and coherent element of Aspire to STEM, and that it aligned closely with their perceived professional development needs**. Survey data indicate that teachers valued the tailored, practitioner-led nature of the CPD on offer, which enabled them to strengthen subject knowledge, develop new pedagogical strategies, and access high-quality resources. Case study schools echoed these perceptions, frequently describing CPD sessions as relevant, well-delivered, and immediately applicable to classroom practice. Although overall programme dosage was diminished by delays, CPD remained the area in which schools engaged most intensively, suggesting that this strand of the programme was both the most accessible and the most clearly communicated to practitioners.

On the other hand, there was greater variability in schools' level of engagement with STEM Learning's network and resources, which include: i) a resource eLibrary; ii) volunteer STEM Ambassadors; and iii) STEM Club infrastructure and community groups. While there was no data collected measuring the extent to which participants accessed STEM Learnings' e-resource library, interviewees from case study schools often described these resources as useful. On the other hand, the review of case study schools' action plans revealed generally limited engagement with other programme components.

Engagement with STEM Ambassadors was low across participating schools. Of the 18 educational leads surveyed, nine reported that the schools in their partnership area had not engaged with STEM Ambassadors while four were unsure. Only the remaining five reported engagements between partnership schools and STEM Ambassadors. Discussions with case study school staff point to possible reasons for this low engagement. Common reasons cited by case study schools include: i) 'onerous' administrative procedures for engaging with STEM Ambassadors; ii) non-responsiveness on the part of STEM Ambassadors; and iii) difficulty finding mutually amenable schedules. When the case study schools managed to engage with STEM Ambassadors, however, they described the experience positively. Therefore, the issue of low engagement has less to do with the quality of the STEM Ambassadors component of Aspire to STEM and is more likely due to administrative inefficiencies.

Similarly, schools' engagement with the programme's STEM careers education component seemed limited. At endline, less than half of both teachers (30 out of 80) and headteachers (18 out of 38) surveyed noted improvements in their respective school's engagement with STEM employers. Furthermore, despite all STEM subject leads surveyed at baseline indicating a desire to develop relationships with STEM employers, only 19 out of 39 surveyed at endline reported that their school's participation in Aspire to STEM had strengthened their relationship with STEM employers.

Despite STEM careers education being the top improvement area identified by case study schools, they rarely mentioned receiving guidance in this area. This could indicate either limited *engagement* with or limited *delivery* of Aspire to STEM's careers education component, but since this information was not obtained in this IPE, this cannot be ascertained. However, discussions with case study schools seem to point towards limited engagement due to various school-level factors, as described below.

School characteristics—such as the level of education taught and rurality—appeared to be associated with schools' degree of engagement with the programme's careers education component. For instance, case study schools that mainly taught primary school-aged children purposely limited their engagement with Aspire to STEM's careers education component because they felt that it was less relevant to their younger pupils. On the other hand, rural schools expressed a desire to engage more with the programme's careers education component, but their remote location and the relatively smaller number of employers in their locale prevented them from doing so. Case study schools serving disadvantaged populations also wished to engage more with the careers education component of Aspire to STEM, but their limited access to funding outside the allocated £25,000 for these projects meant that they had to channel resources towards other priorities.

Barriers to and facilitators of programme success (IPE research questions 4 and 5)

Resource availability

School-level resourcing capacity was an important determinant of broader implementation success, not only in relation to the programme's careers education component but across all areas of intervention. In this context, *resourcing* refers primarily to **staff time and capacity**, including time for senior leadership oversight, teacher release for CPD, co-ordination of partnerships, and administrative effort, as well as associated costs such as cover, travel, and scheduling. In the context of limited staffing capacity and competing priorities, some participating schools reported having to **reallocate staff time and operational resources away from Aspire to STEM**, constraining the extent to which they were able to engage fully with the programme. This issue was especially pronounced in smaller-sized schools—as remarked by staff in the smaller case study schools—which cited budget constraints and heavier teaching workloads as important implementation challenges. Survey responses corroborate this finding: according to endline surveys, 46 out of 76 teachers and 21 out of 37 headteachers found competing priorities to have presented, at least to some degree, difficulties in implementing Aspire to STEM in their schools. In addition, resourcing also encompassed schools' own discretionary budgets outside the programme, as some school leaders reported needing to draw on internal financial resources to support delivery, for example to fund staff cover, administrative support, or logistical costs not fully met by Aspire to STEM funding.

While engagement in CPD was comparatively high, teachers also reported substantial barriers to sustained participation. Time and workload pressures were the most frequently cited challenges, with several teachers noting that attending CPD required preparing cover work, catching up on missed teaching time, and balancing CPD commitments with the demands of teaching examination classes.

Human resource constraints were an equally important barrier to the successful implementation of Aspire to STEM. Challenges with staffing, absence, and turnover were commonly reported in midline and endline surveys with both headteachers and STEM subject leads. Problems around accessing cover were also common, with 43 out of 77 teachers surveyed at endline reporting that they had no cover at all for time allotted to Aspire to STEM and only 18 having cover for more than 75% of the time they spent towards the programme. These human resource constraints also commonly surfaced in interviews with case study school staff, many of whom cited difficulties finding cover for teachers to attend programme-related CPD. For example, while 45% of teacher respondents reported spending four or more hours on the programme in addition to their usual workload per month, over half of respondents (56%) reported having no cover at all, with less than a quarter (23%). Multiple staff from the case study schools also identified finding cover for teachers to attend CPD as a key challenge.

Staffing difficulties became even more pronounced given the additional administrative burden that accompanied programme implementation. Nearly half of surveyed teachers reported spending four or more hours on the programme, *in addition to* their usual workload per month. While these pressures partly reflect wider systemic issues, evidence from case study schools suggests that some challenges were compounded by Aspire to STEM's funding and delivery model. In particular, the reliance on educational leads as intermediaries for accessing funding and resources meant that staffing changes, limited administrative capacity, or misalignment between school priorities and partnership-level decisions could delay or constrain schools' ability to make full use of programme support. These issues are explored in more detail in the following section on operational features and leadership, which examines how funding mechanisms, decision-making structures, and school autonomy shaped implementation experiences.

Operational features of Aspire to STEM

In some instances, **Aspire to STEM's funding model compounded difficulties encountered by resource-constrained schools**. Since the programme was set-up such that the educational lead acted as a conduit for accessing resources, corresponding changes in staffing would make it difficult to obtain funding. For example, one case study school expressed that they would have been more engaged with Aspire to STEM had they been allowed access to funding and CPD resources without needing to go through a designated STEM Learning intermediary. Aligned with this sentiment, other case study schools suggested more flexibility for schools to choose areas in which they would invest programme resources. Bespoke provision is a core element of Aspire to STEM's implementation model, allowing schools more autonomy over their use of programme resources could have resulted in programme activities becoming even more responsive to participating schools' needs.

Indeed, **while staff in participating schools spoke positively about the bespoke nature of Aspire to STEM, some would have appreciated an even more tailored approach.** Both the educational leads surveyed at baseline and the interviewed case study school staff considered the programme's bespoke design to be a facilitator of programme success. However, headteachers surveyed at endline did not share the same sentiment, remarking how Aspire to STEM could benefit from greater tailoring to schools' needs. Teachers and headteachers surveyed at endline identified the following potential improvements:

- more concrete and tailored guidance for staff;
- better communication regarding the programme;
- increasing the locally available CPD provision; and
- giving more attention to other STEM subjects besides science.

Leadership and approach to decision-making

In the face of the challenges mentioned thus far, **supportive school leadership appeared to mitigate—at least slightly—against the adverse influence of resourcing constraints and administrative inefficiencies.** In endline surveys from nine out of 25 teachers and seven out of 27 headteachers, the SLT and subject leaders were described as having an enabling influence on their school's implementation of Aspire to STEM.

Case study interviews support these survey findings. According to interviews with case study school staff, a supportive SLT ensured that the implementation of Aspire to STEM would be prioritised. Prioritisation of Aspire to STEM meant that: i) there was a dedicated programme implementation budget; ii) staff were enabled to access a greater volume and range of Aspire to STEM-related CPD, for example, through leadership support to release staff time, approve participation, and align CPD choices with school priorities; and iii) staff involvement in Aspire to STEM was encouraged. Effective STEM subject leadership was described as playing a similar role, with supportive STEM subject leads ensuring adequate training and resource availability for school staff while also helping increase pupil engagement, in line with the project requirements.

In addition, focus group discussions with educational leads revealed that **school staff felt more inclined to support Aspire to STEM when they were actively involved in** shaping how the programme was implemented at school level. This included contributing to decisions about how programme resources were used, which priorities were addressed through CPD, and how activities were sequenced to align with school needs. School staff were reportedly less supportive when the programme was implemented by multi-academy trust executives without the input of the former. Since surveys with multiple teachers, headteachers, and educational leads identified staff engagement as an important facilitator in implementing Aspire to STEM, and school leadership was described as influencing staff engagement, school leadership can therefore, be seen as one mediator of programme effectiveness.

Subjective gains and ways forward (IPE research questions 3 and 4)

Teachers attributed many of the positive subjective outcomes observed during the programme, including increased confidence, motivation, and instructional competence, to their participation in CPD. Case study interviews suggest that CPD often served as a catalyst for renewed enthusiasm for STEM teaching, providing staff with fresh ideas, clearer curricular understanding, and greater assurance in delivering practical work and real-world applications. These improvements were also perceived to have a tangible impact on students, with teachers reporting increased engagement, curiosity, and aspiration in STEM subjects following the adoption of new approaches learned through CPD. Even where engagement with other programme components was limited, the CPD strand appeared to exert a meaningful influence on classroom practice and, by extension, on pupil STEM experiences.

Despite the challenges faced in programme implementation and the reduced dosage due to delays, **there still appeared to be subjective gains in STEM teachers' confidence, motivation, and feelings of competence.** Around 70 of the 79 teachers surveyed at endline reported feeling, at least to some extent, more confident in their ability to teach STEM-related subjects. In addition, 74 of these 79 teachers felt at least somewhat more motivated to teach STEM-related subjects; 53 of the 74 indicated that they were, 'to a great extent', more motivated. These gains in confidence and motivation seem to have also positively influenced teachers' job satisfaction, with 59 of the 79 surveyed teachers reportedly feeling happier with their jobs after participating in Aspire to STEM.

Interviews with case study school staff corroborate the subjective improvements identified in the survey analysis. Case study schoolteachers reported improvements in their STEM subject knowledge, a greater awareness of available resources, and more ideas on how to approach STEM teaching. During interviews, teachers also felt more knowledgeable about what they need to teach for each year group and therefore, more prepared for Ofsted inspections. Moreover, interviewees reported feeling that they and their subject were more valued.

Teachers also perceived gains in their students' interest in STEM subjects, with 63 of 79 surveyed teachers responding that their students were, at least to some extent, more engaged in STEM-related subjects, following their school's participation in the programme. Around 31 of the 37 headteachers surveyed at endline agreed with this sentiment. Accounts from case study school staff agree with this sentiment, with school staff observing improvements in pupils' confidence and STEM-related aspirations. Some interviewees from case study schools attributed these pupil-level improvements to the previously described improvements in teachers' confidence in and enthusiasm for STEM instruction.

However, while stakeholders perceived Aspire to STEM to have a positive impact across different levels of the school structure, it is worth reiterating that adherence to the intended programme design was modest. Such challenges in programme adherence need to be addressed if Aspire to STEM and similar programmes are to attain their more objective target outcomes.

All in all, the IPE found that the programme was only partially delivered as intended, with significant variation in fidelity and dosage across schools. While most schools completed core requirements and engaged strongly with CPD, which was widely valued and led to improvements in teacher confidence, motivation, and classroom practice, other components such as partnerships, STEM Ambassadors, and careers education saw limited or inconsistent uptake. Early implementation delays, administrative burdens, and lack of clarity around partnership expectations reduced programme duration and created bottlenecks, while school-level constraints (e.g. staffing shortages, workload pressures, limited resources) further hindered delivery, particularly in smaller and disadvantaged schools. Despite these challenges, supportive leadership, staff engagement, and the programme's tailored approach acted as key facilitators, and there is evidence of positive observed impacts on both teachers and pupils.

Cost

Prerequisite and operational costs

Data from STEM Learning was collected at the partnership level and outlines how much money was spent by each partnership in each cohort across a number of categories. These categories were: consultancy; CPD; community event; cover costs; educational lead consultation; and other. While prerequisite (or set-up) costs included consultancy activities (e.g. conducting needs assessment and completing action plans), operational (or running) costs included: i) CPD sessions; ii) community events; iii) providing staff cover; iv) consultation with education leads; and v) other miscellaneous activities.

As part of Aspire to STEM, each of the 40 partnerships were given £25,000 of funding for the following activities: i) bespoke CPD for teachers and leaders; ii) career guidance for pupils; iii) access to STEM enrichment and inspiration activities; and iv) access to online resources via STEM Learning's STEM Club infrastructure and eLibrary.[36]

As seen in Tables 17 and 18, partnerships in both cohorts spent most of their funding on CPD. This corroborates survey and interview data, which found reportedly high levels of engagement with Aspire to STEM's CPD component. Cohort 1 and Cohort 2 had similar average spending on CPD over the duration of the intervention. Cohort 1 spent an average of £16,305.76, compared to an average of £16,701.40 across Cohort 2. The smallest spend on CPD was a Cohort 1 partnership in the East Midlands which spent £11,226.00. This partnership also had the lowest total spending, £17,396.00.

Table 17: Summary of Cohort 1 expenditures[a]

	Consultancy (NA, AP, bespoke)	CPD (individual / group)	Community events	Staff cover	Consultations with educational leads (mentoring reviews)	Other
Spending range	£900.00–£3,376.00	£14,462.21–£17,682.00	£1,500.00–£2,100.00	£500.00	£4,278.00–£5,400.00	£0.00–271.20
Mean expenditure	£1,665.12	£16,305.76	£1,625.33	£500.00	£4,864.87	£20.75
Total expenditure	£24,976.80	£244,586.35	£24,380.00	£7,500.00	£72,973.07	£311.20

[a] *Spending across two years of implementation.*
AP=Alternative Provision; NA=National Assessment

Table 18: Summary of Cohort 2 expenditures[a]

	Consultancy (NA, AP, bespoke)	CPD (individual / group)	Community events	Staff cover	Consultations with educational leads (mentoring reviews)	Other
Spending range	£900.00–£2,305.81	£11,226–19,115.40	£1,500.00–£2,846.14	£500.00–£750.00	£2,520.00–£5,750.00	£0.00–£1,516.80
Mean expenditure	£1,384.90	£16,701.40	£1,652.37	£520.00	£4,451.21	£86.76
Total expenditure	£34,622.46	£417,534.91	£41,309.14	£13,000.00	£111,280.35	£2,169.01

[a] *Spending across two years of implementation.*
AP=Alternative Provision; NA=National Assessment

For both cohorts, consultations with educational leads incurred the second largest cost, with Cohort 1 having a slightly higher average spending than Cohort 2. Across both cohorts, the average costs of other expenses tended to be similar. In line with our findings from surveys and interviews, staff cover costs were found to be low across partnerships.

The spending patterns observed suggest that partnerships used Aspire to STEM primarily to facilitate school staff's access to CPD. While CPD was an important element of the programme's Theory of Change, other core programme elements appear to have been relatively under-utilised by partnership schools. This does raise some questions about whether the programme was implemented entirely as originally designed.

---

[36] Access to online resources was free of charge for participating schools.

Additional costs

In the endline survey, headteachers from participating schools were asked to estimate the sum of any additional implementation costs incurred by their school since the start of Aspire to STEM. This additional spending could include costs of organising school visits to STEM employers, resources purchased to support a STEM week or fair, costs to collaborate with other schools in their partnerships, and other such ongoing expenditures.

Around 34 of the 40 surveyed headteachers indicated that they did not know how much additional spending their school had incurred in implementing the programme. Of the six respondents who were aware of additional costs incurred by the intervention at their school, only three respondents could give indicative amounts. Qualitative responses ranged from one respondent mentioning that additional costs were 'in excess of allocation by far', to another respondent describing additional spending as being 'very little'. Another reported that 'all costs [were] covered' by programme funding. Where headteachers gave indicative amounts, the provided estimates were £1,000, £1,200, and £1,500. While there was only a very small number of respondents, it appears additional spending were minimal and it is possible that the programme in its entirety incurred little, if any, additional costs for schools. Furthermore, it is reasonable to assume that if additional costs incurred by schools had been excessive, this would have come to the attention of headteachers.

Indicative average costs

We did not observe the flows of partnership funding to individual schools. Additionally, information is not available on which pupils were beneficiaries of the intervention and while Aspire to STEM is intended to be a whole-school intervention, not all pupils at participating schools were necessarily beneficiaries of the intervention.

However, while granular data on spending at the school was not available, we provide indicative average costs of Aspire to STEM. Our calculations are shown in Table 19. We do not estimate separate average costs for primary and secondary schools, we instead assume each Aspire to STEM school received an equal amount of funding through the programme. We know 207 schools (92 primary, 115 secondary) received Aspire to STEM. Dividing the total programme cost (£994,643) between these treatment schools suggests an average cost per school of £4,805. Furthermore, using the average number of pupils per school at the January 2019 School Census, we estimate the number of pupils the Aspire to STEM programme reached was 136,482. Again, dividing the total programme cost by this number of pupils gives an estimated average cost per pupil of £7.29.

Table 19: Estimated average cost per school and per pupil, by Cohorts 1 and 2

	Primary	Secondary	Total
Number of Aspire to STEM schools	92	115	**207**
Average number of pupils per school[a]	276	966	**–**
Estimated number of pupils	25,392	111,090	**136,482**
Total Aspire to STEM cost			**£994,643.29**
Estimated average cost per school			**£4,805.04**
Estimated average cost per pupil			**£7.29**

[a] Mean averages calculated using the total numbers of pupils reported as attending state-funded primary and secondary school in the January 2019 School Census. See: ***https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2019****.*

Conclusion

Table 20: Key conclusions

Key conclusions
1. Due to COVID-19, the intended student attainment outcomes for this evaluation (key stage 2 and key stage 4 attainment in mathematics and STEM subjects) were not collected. Key conclusions are therefore based on progression to A level, supported by secondary outcomes and qualitative data from the IPE.
2. The evaluation found that attending an Aspire to STEM secondary school slightly increased the likelihood of key stage 4 pupils progressing to taking a STEM subject at A level two years later, by 1.42 percentage points. This is equivalent to around an additional 1 in 70 year 11 pupils taking up a STEM A level.
3. There was no difference in science teacher turnover between secondary schools which received Aspire to STEM and those that did not. This implies bespoke CPD provision does not influence science teachers' retention.
4. Participating schools recognised the importance of partnership working but struggled to fulfil the programme's early-stage partnership milestones, including conducting the needs analysis and action planning.
5. CPD was the most visible and coherent element of Aspire to STEM and was seen as aligning closely with teachers' perceived professional development needs. While engagement in CPD was comparatively high, teachers also reported substantial barriers to sustained participation, including time and workload pressures.

Impact evaluation and IPE integration

Evidence to support the logic model

The findings presented throughout this study provide partial but meaningful support for the Aspire to STEM Theory of Change. Several components appear to have operated as intended. The strongest evidence concerns the pathway between teacher-level changes (increased confidence, motivation, and competence) and pupil-level outcomes. Surveyed teachers reported greater confidence in teaching STEM, increased motivation, and a stronger sense of professional competence following participation in programme CPD. These perceived gains were mirrored in schools where interviewees described improved STEM subject knowledge, greater awareness of available teaching resources, and more ideas for approaching STEM instruction. Both teachers and headteachers also reported noticing improvements in pupils' confidence, engagement, and interest in STEM subjects. These findings align with the positive impact observed in the quantitative analysis on STEM A level uptake, suggesting that the teacher-level mechanisms outlined in the logic model may have contributed to pupil-level changes. However, caution is warranted in interpreting these pathways, as the evaluation was not able to measure or control for variation in schools' access to, or engagement with, wider STEM Learning network resources and materials. As a result, it is not possible to isolate the relative contribution of specific programme components, and some observed effects may reflect differing levels of exposure across schools.

There is limited evidence to support other components of the programme, particularly those relating to leadership development, careers education, and wider enrichment activities. While some school leaders reported supportive SLTs and active subject leadership facilitating implementation, engagement with careers guidance and employer links was limited across most partnerships. Fewer than half of surveyed teachers and headteachers reported improvements in employer engagement, and case study schools often noted receiving little guidance in this area. Similarly, use of STEM Ambassadors and other enrichment components varied widely due to administrative burdens, non-responsiveness, and logistical challenges.

Finally, the evidence for the component linking improved teaching and school environment to teacher retention was limited. Although most surveyed teachers reported increased confidence and motivation, these subjective improvements did not translate into observable differences in retention or turnover outcomes in the impact analysis. Survey and interview data suggest that broader structural factors, including workload, staffing shortages, and competing priorities, were more influential on retention decisions than programme-specific CPD. This indicates that the Theory of Change assumptions about the influence of CPD on teacher retention may require refinement, particularly in contexts facing wider recruitment and workload pressures.

Overall, the evaluation supports several of the teacher- and pupil-level pathways, but suggests that the leadership, community engagement, and retention pathways were limited, mainly due to implementation barriers and programme delays. Areas requiring further exploration include leadership development, careers education, engagement with STEM employers, and the role of CPD dosage in shaping teacher outcomes, none of which could be robustly assessed due to missing participation data and uneven delivery.

Interpretation

The quantitative and qualitative results converge. The impact evaluation suggests that Aspire to STEM induced changes in pupils' likelihood of progressing to a STEM A level, but not for teachers' retention outcomes. This difference appears to be at least partially explained by the degree to which different parts of the programme were implemented.

The impact evaluation found a positive and statistically significant effect on STEM A level uptake. This aligns with previous quantitative studies (Fletcher-Wood and Zuccollo, 2020; Gore *et al.,* 2017) which show that sustained, subject-specific CPD can lead to improvements in teaching quality and pupil engagement, particularly in STEM. This is further complemented by evidence that suggests that GSCE qualifications, alongside preferences, and pathways, have an important role to play in progression of young people into STEM subjects, particularly for those from underrepresented backgrounds (Hodge *et al.,* 2024). Unfortunately, due to the impacts of Covid-19, we were unable to explore the role qualifications played in driving our findings. However, regardless of mechanism, the positive effect we estimate on STEM A level uptake aligns with this evidence and provides new quasi-experimental support for STEM CPD.

It is also well aligned to our IPE evidence which showed that CPD was the most consistently delivered and well-engaged component of the programme: 37 of 42 STEM subject leaders and 61 of 80 teachers surveyed reported participating in at least one CPD activity, and CPD activities dominated the action plans reviewed in the document analysis. Surveyed teachers also reported increased confidence, motivation, and competence in teaching STEM, which case study schools described as boosting pupil engagement and aspirations. These teacher-level gains map directly onto the Theory of Change and may explain the observed pupil outcome.

In contrast, the programme showed no measurable impact on teacher retention or turnover, despite substantial investment in teacher development. While CPD was the most consistently delivered and positively received component of Aspire to STEM, IPE data suggest that engagement with CPD alone was not sufficient to influence teachers' retention decisions. Teachers frequently cited workload pressures, lack of cover, staffing shortages, and competing priorities as dominant factors shaping their working conditions, limiting not only the extent to which CPD could be fully embedded in practice but also its potential to affect longer-term career intentions. In addition, administrative delays reduced programme exposure in some schools by up to six months, further constraining dosage. Taken together, these findings suggest that even high-quality CPD has limited capacity to offset wider systemic pressures on teacher retention, particularly in the context of Covid-19-related disruption and longstanding workforce challenges.

This is also consistent with the existing CPD literature (Allen and Sims, 2017; Bélanger and Broeks, 2016; DfE, 2019; Foster, 2018). Previous studies have found that CPD alone rarely shifts retention behaviours when structural drivers such as workload, behaviour, salary, and local labour market conditions exert stronger influence. Surveyed teachers and headteachers in this evaluation described precisely these constraints, reinforcing findings from wider retention research.

Other components of the intervention, particularly STEM Ambassadors, employer engagement, and enrichment activities, were delivered unevenly and at lower dosage. Schools described administrative 'onerous' processes, difficulty securing STEM Ambassador responsiveness, or limited local employer availability (especially in rural settings). These implementation gaps weakened the intended whole-school and community-engagement elements of the programme's Theory of Change, which were designed to reinforce staff motivation, professional identity, and longer-term career commitment. By contrast, pupil impacts could still emerge through improvements in classroom practice driven by CPD, even when wider engagement components were only partially realised. This asymmetry helps explain why positive effects were observed for pupils, while staff-level outcomes, which are more sensitive to systemic conditions and sustained organisational support, did not change measurably.

Limitations and lessons learned

The findings presented in this evaluation should be interpreted with caution due to several limitations in the study design, data availability, and implementation context.

The first limitation is that the evaluation design predates updated guidance by the EEF on implementation and usual practice. As a result, no business as usual data was collected at baseline or endline. This restricts the ability to assess whether Aspire to STEM represented a substantial shift from schools' pre-existing STEM CPD and careers provision. Without a clear counterfactual description of usual provision, attribution of observed differences to the programme is less secure.

Second, the evaluation was constrained by missing compliance and dosage data. Partnership-level spending information was available, but school-level participation in CPD, careers activities, STEM Ambassador sessions, or enrichment activities was not systematically recorded. Programme delays, administrative inefficiencies, and Covid-19 disruption contributed to these gaps. The absence of detailed implementation records limits the precision of dosage–outcome analysis and prevents disaggregated impact estimates for schools with higher versus lower participation.

The evaluation also faced low response rates across several surveys, especially among teachers and headteachers at endline, raising the possibility of non-response bias. Teachers who responded may have been more engaged or more positive about the programme, meaning that self-reported gains in confidence and motivation could be overestimated. Similarly, the small sample of surveyed educational leads offers only a partial picture of partnership functioning. Although qualitative case studies mitigated this to some extent, findings that rely on self-reported perceptions should be interpreted with care.

The impact evaluation also faced some challenges. It was not possible to evaluate any attainment outcomes, as national assessments were disrupted by Covid-19, restricting the analysis to participation outcomes. Although Aspire to STEM included primary and secondary schools, the matched comparison sample for primary schools was too small to support a feasible impact analysis, meaning the findings apply only to secondary schools. In addition, while the evaluation expanded beyond the original two-period design to use the full available time series, it was still not possible to account for the staggered roll-out of the programme across partnerships. Finally, the analysis took place during a period of significant Covid-19 disruption, which may have affected both implementation and outcomes in ways not fully captured by the available data.

Implementation constraints may have also introduced potential bias. As summarised in Box 1, more than half of the schools experienced administrative delays, leading to up to six months less programme exposure than intended. These delays represent a clear threat to fidelity and dosage, both of which affect internal validity. Other barriers, including staff shortages, difficulties accessing cover, budget constraints, low engagement with STEM Ambassadors, and limited employer access in rural areas, also reduced adherence to programme design. In several schools, staff interpreted 'partnership working' as extensions of pre-existing collaborations rather than the formal Aspire to STEM partnership model, suggesting variation in how programme components were understood and delivered.

The Covid-19 pandemic represents an additional and substantial limitation. The programme and evaluation were designed prior to the Covid-19 pandemic, and the ensuing school closures, remote learning, staffing shortages, and altered workload patterns likely disrupted both implementation and outcome measures. Teacher decisions to move schools or leave the profession during this period may also have been influenced by these external factors complicating the interpretation of the null findings for turnover and wastage.

Future research and publications

Several lessons emerge from the evaluation that could inform future research and programme design.

First, clearer and earlier communication about programme expectations, particularly around partnership working, could have improved fidelity. Case study schools frequently highlighted misunderstandings about the nature of Aspire to STEM partnerships. Establishing shared expectations and allocating more time for early partnership-building, as recommended by educational leads, would likely have strengthened coherence across participating schools. Further research could help unpick this approach to working.

Second, future evaluations of partnership-based or bespoke interventions would benefit from systematic participation and compliance tracking. This includes recording CPD attendance, the number and nature of careers activities delivered, STEM Ambassador engagements, and the extent of action plan completion at school level. Without these data, it is difficult to understand variation in outcomes or refine programme design.

Third, the challenges summarised in Box 1, administrative inefficiencies, financial and human resource constraints, limited careers engagement, and the inflexibility of the funding model, suggest that successful implementation depends heavily on enabling conditions. Future research could explore the importance of streamlining Aspire to STEM's offer, strengthening support for careers components, and considering more flexible funding mechanisms that would allow schools greater autonomy.

Finally, the evaluation highlights the importance of leadership engagement. Schools with supportive SLTs and strong subject leadership were more likely to prioritise Aspire to STEM, allocate staff time, and participate in CPD. Securing senior leader buy-in should therefore be an explicit condition of participation in similar initiatives.

Box 1: Main challenges and ways forward

1. Long, drawn-out administrative procedures caused delays in the commencement of Aspire to STEM, which resulted in diminished dosage. Some schools reported participating in the programme for up to six months less than intended.

   Recommendation: Administrative processes should be streamlined, and more time should be dedicated towards partnership-building. This builds on educational leads' suggestion to allow more time to build relationships with partnership schools in the lead-up to implementing Aspire to STEM.

2. Aspire to STEM's careers education component saw limited engagement, especially among primary schools and schools located in rural areas.

   Recommendation: To improve primary schools' engagement in careers education, it is important to build awareness of the importance of careers education in primary education. On the other hand, since rural schools described difficulties engaging with employers in their community, subsequent programmes could explore factors that prevent employer engagement in rural communities and tailor their support, so it bridges this gap.

3. Financial and human resource constraints hampered the implementation of Aspire to STEM, with budgetary limitations, heavier teacher workloads, and difficulties accessing staff cover for programme involvement being the most notable issues.

   Recommendation: Since supportive school leadership was found to mitigate, at least partly, against resourcing constraints, senior leaders buy-in should be a priority throughout programme implementation. In schools with supportive leadership, Aspire to STEM was more actively prioritised, which in practice enabled greater teacher participation in CPD through actions such as approving release time, allocating internal budgets for cover, and aligning CPD choices with school priorities, rather than through an expansion in the overall supply of CPD. To further support equitable access to CPD across schools, consideration should also be given to reconfiguring the programme's funding model to increase flexibility and reduce administrative barriers, particularly for resource-constrained schools.

4. The relative inflexibility of Aspire to STEM's funding model undermined the programme's bespoke design.

   Recommendation: Schools could be given more autonomy in the use of programme funding. Giving schools a greater ability to determine how they would use programme funding would make Aspire to STEM more responsive to schools' needs.

References

Allen, R. and Sims, S. (2017) *'Improving Science Teacher Retention: Do National STEM Learning Network Professional Development Courses Keep Science Teachers in the Classroom?'* London: Wellcome Trust. Available at: **https://wellcome.ac.uk/sites/default/files/science-teacher-retention.pdf** (accessed 16 06 2026).

Anders, J., Brown, C., Ehren, M., Greany, T. and Nelson, R. (2017) *'Evaluation of Complex Whole-School Interventions: Methodological and Practical Considerations'.* London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-evaluation-reports-and-research-papers/methodological-research-and-innovations/evaluation-of-complex-whole-school-interventions (accessed 16 06 2026).

Anders, J. and Jerrim, J. (2021) *'Statistical Analysis Plan for Schools Partnership Programme'.* London: Education Endowment Foundation. Available at: **https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/SPP_SAP_v3.pdf?v=17 07805077** (accessed 16 06 2026).

Atteberry, A., Loeb, S. and Wyckoff, J. (2017) 'Teacher Churning: Reassignment Rates and Implications for Student Achievement'. *Educational Evaluation and Policy Analysis,* 39: 1, 3–30. **https://doi.org/10.3102/0162373716659929**

Bélanger, J. and Broeks, M. (2016) *'Attracting and Retaining Teachers in Cambridgeshire: Working Conditions and Teacher Flows From a School Workforce Census Data Perspective'.* RAND Corporation. Santa Monica, CA. Available at: **https://www.rand.org/pubs/research_reports/RR1850.html** (accessed 16 06 2026).

Data Protection Act (2018) c.29 Available at: **https://www.legislation.gov.uk/ukpga/1998/29** (accessed 16 06 2026)

Dawson, S., Tang, S. and Worth, J. (2023) 'The Impact of Training Bursaries on Teacher Recruitment and Retention: An Evaluation of Impact and Value for Money'. Slough: NFER. Available at: **https://www.nfer.ac.uk/publications/the-impact-of-training-bursaries-on-teacher-recruitment-and-retention/** (accessed 16 06 2026).

Department for Education (DfE). (2016) '*Defining Achieving Excellence Areas Methodology Guidance Note'.* London: Department for Education. Available at: **https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/508392/Methodology_guidance_note_-_defining_achieving_excellence_areas.pdf** (accessed 16 06 2026).

Department for Education (DfE). (2017) *'Opportunity Areas Selection Methodology'.* London: Department for Education. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650036/Opportunity_areas_selection_methodology.pdf (accessed 16 06 2026).

Department for Education (DfE). (2018) '*Initial Teacher Training: Trainee Number Census – 2018 to 2019'.* GOV.UK. Available at: https://www.gov.uk/government/statistics/initial-teacher-training-trainee-number-census-2018-to-2019 (accessed 16 06 2026).

Department for Education (DfE). (2019) 'School Workforce in England: November 2018'. GOV.UK. Available at: https://www.gov.uk/government/statistics/school-workforce-in-england-november-2018 (accessed 16 06 2026).

Department for Education (DfE). (2025a) 'Initial Teacher Training Census. Academic Year 2025/26'. GOV.UK. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/initial-teacher-training-census/2025-26 (accessed 16 06 2026).

Department for Education (DfE). (2025b) 'School Workforce in England. Reporting Year 2024'. GOV.UK. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england/2024 (accessed 16 06 2026).

Department for Education (DfE). (2026) *'Every Child Achieving and Thriving'*. GOV.UK. Available at: https://www.gov.uk/government/publications/every-child-achieving-and-thriving/every-child-achieving-and-thriving-html-version (accessed 16 06 2026).

Fereday, J. and Muir-Cochrane, E. (2006) 'Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development'. *International Journal of Qualitative Methods*, *5*: 1, 80–92. https://doi.org/10.1177/160940690600500107

Fletcher-Wood, H. and Zuccollo, J. (2020) '*Evidence Review: The Effects of High-Quality Professional Development on Teachers and Students*'. London: Education Policy Institute. https://epi.org.uk/publications-and-research/effects-high-quality-professional-development/ (accessed 16 06 2026).

Foster, D. (2018) *'Teacher Recruitment and Retention in England'*. House of Commons Library. Commons Briefing Papers CBP-7222. UK Parliament. Available at: **https://dera.ioe.ac.uk/id/eprint/32668/1/CBP-7222%20%281%29.pdf** (accessed 16 06 2026).

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., Doolittle, F. and Warner, E. (2011) *'Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation (NCEE 2011-4024)'*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M.S., Heppen, J.B., Walters, K., Parkinson, J., Smith, T.M., Song, M., Garrett, R., Yang, R., Borman, G. D. and Wel, T.E. (2016) *'Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher Professional Development (NCEE 2016-4010)'*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gore, J., Lloyd, A., Smith., M., Bowe, J., Ellis, H. and Lubans. D. (2017) 'Effects of Professional Development on the Quality of Teaching: Results From a Randomised Controlled Trial of Quality Teaching Rounds'. *Teaching and Teacher Education,* 68, 99–113. **https://doi.org/10.1016/j.tate.2017.08.007**

Greaves, J. and Brawley, S. (2025) '*UK STEM Skills Pipeline* (POSTnote No. 746)'. Parliamentary Office of Science and Technology. UK Parliament. https://post.parliament.uk/research-briefings/post-pn-0746/ (accessed 16 06 2026).

Griffin, B.A., Ridgeway, G., Morral, A.R., Burgette, L.F., Martin, C., Almirall, D., Ramchand, R., Jaycox, L. H. and McCaffrey, D.F. (2014) 'Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG)'. RAND Corporation. Available at: **http://www.rand.org/statistics/twang** (accessed 16 06 2026).

Groom, M., Merewood, J., Oades, F., Singh, B., Ilie, S., Shenderovich, Y., Sutherland, A., Mihaly, K., Dimova, S. and Broeks, M. (2025) 'Study Plan for Aspire to STEM'. London: Education Endowment Foundation. Available at: **https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/aspire_to_stem_-_study_plan_-_amended_v2.pdf?v=1763677044** (accessed 16 06 2026).

Hedges, L.V. (1981) 'Distribution theory for Glass' estimator of effect size and related estimators'. *Journal of Educational Statistics,* 6: 2, 107–128. **https://doi.org/10.3102/10769986006002107**

Hedges, L.V. (2017) 'Challenges in Building Usable Knowledge in Education'. *Journal of Research on Educational Effectiveness*, 11: 1, 1–21. **https://doi.org/10.1080/19345747.2017.1375583**

Hodge, L., Stevenson, L., Griffiths, J. and Ramaiah, B. (2024) '*Progression at Age 16 of Young People From Underrepresented Backgrounds Towards Careers in STEM'*. London: Education Policy Institute. Available at: **https://epi.org.uk/publications-and-research/progression-at-age-16-of-young-people-from-underrepresented-backgrounds-towards-careers-in-stem/** (accessed 16 06 2026).

Hodgen, J., Taylor, B., Anders, J., Tereshchenko, A. and Francis, B. (2019) *'Study Plan for the Student Grouping Study: Investigating the Effects of Setting and Mixed Attainment Grouping'*. London: Education Endowment Foundation.

https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Students_Grouping_Study_Study_Plan_2019.11.19.pdf?v=1707829329 (accessed 16 06 2026).

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2015) 'Implementation and Process Evaluation (IPE) for Interventions in Education Settings: An Introductory Handbook'. London: Education Endowment Foundation. Available at: **https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/IPE_Handbook.pdf** (accessed 16 06 2026).

Hutchison, D. and Styles, B. (2010) *'A Guide to Running Randomised Controlled Trials for Educational Researchers'*. Slough: NFER. Available at: **https://www.nfer.ac.uk/media/qdjknw3i/a_guide_to_running_randomised_controlled_trials_for_educational_researchers.pdf** (accessed 16 06 2026).

Maynard, R.A. and Dong, N. (2013) '*PowerUp!:* A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies'. *Journal of Research on Educational Effectiveness,* 6: 1, 24–67. **https://doi.org/10.1080/19345747.2012.673143**

McKenzie, D. (2012) 'Beyond Baseline and Follow-Up: The Case for More T in Experiments'. *Journal of Development Economics*, 99: 2, 210–221. **https://doi.org/10.1016/j.jdeveco.2012.01.002**

McNeill, J., Butt, G. and Armstrong, A. (2014) 'Developing Collaborative Approaches to Enhance the Professional Development of Primary Mathematics Teachers'. *Education 3–13, 44*: 1, 426–441. https://doi.org/10.1080/03004279.2014.973896

Nunes, T., Bryant, P., Strand, S., Hillier, J., Barros, R. and Miller-Friedmann, J. (2017*) 'Review of SES and Science Learning in Formal Educational Settings. A Report Prepared for the EEF and the Royal Society'.* London: Education Endowment Foundation. Available at: **https://www.royalsociety.org/-/media/policy/topics/education-skills/education-research/evidence-review-eef-royalsociety-22-09-2017.pdf** (accessed 16 06 2026).

Opfer, D. (2016) 'Conditions and Practices Associated with Teacher Professional Development and its Impact on Instruction in TALIS 2013'. *OECD Education Working Papers*, No. 138. Paris: OECD Publishing. **https://dx.doi.org/10.1787/5jlss4r0lrg5-en**?

Pahlke, E., Hyde, J.S. and Allison, C.M. (2014) 'The Effects of Single-Sex Compared with Coeducational Schooling on Students' Performance and Attitudes: A Meta-Analysis'. *Psychological Bulletin*, 140: 4, 1042–1072. **https://doi.org/10.1037/a0035740**

Sims, S. and Fletcher-Wood, H. (2018) *'Characteristics of Effective Teacher Professional Development: What We Know, What We Don't, How We Can Find Out'.* London: UCL Institute of Education. Available at: https://improvingteaching.co.uk/wp-content/uploads/2018/09/Characteristics-of-Effective-Teacher-Professional-Development.pdf (accessed 16 06 2026).

Snell, R. J. S., Klassen, R. M., Thompson-Lee, S., & Wang, H. (2024). *Back to basics with teacher recruitment: What do STEM undergraduates want? European Journal of Education*, *60*(1). https://doi.org/10.1111/ejed.12840

Srivastava, A. and Thomson, S.B. (2009) 'Framework Analysis: A Qualitative Methodology for Applied Policy Research'. *Journal of Administration and Governance, 4*: 2, 72–79. **https://ssrn.com/abstract=2760705**

Straw, S. and Macleod, S. (2013) *'Improving Young People's Engagement with Science, Technology, Engineering and Mathematics (STEM)'.* London: NFER. Available at: https://www.nfer.ac.uk/policies/improving-young-peoples-engagement-with-science-technology-engineering-and-mathematics-stem/ (accessed 16 06 2026).

Straw, S., Nelson, J., Maxwell, B. and Coldwell, M. (2022) '*Teaching and Leadership Innovation Fund (TLIF): Fund-level Evaluation: Final Report'*. London: Department for Education. https://assets.publishing.service.gov.uk/media/631f678ee90e077dc154d3a3/Teaching_and_Leadership_Innovation_Fund__TLIF__Fund-level_evaluation.pdf (accessed 16 06 2026).

Stuart, E.A., Huskamp, H.A., Duckworth, K., Simmons, J., Song, Z., Chernew, M.E. and Barry, C.L. (2014) 'Using Propensity Scores in Difference-in-Differences Models to Estimate the Effects of a Policy Change'. *Health Services and Outcomes Research Methodology*, 14: 4, 166–182. **https://doi.org/10.1007/s10742-014-0123-z**

Thomson, D. and Plaister, N. (2022) *'Stimulating Physics Network: Phase 4 and Phase 5 Evaluation'*. London: FFT Education Datalab. Available at: **https://ffteducationdatalab.org.uk/wp-content/uploads/2022/01/SPN-evaluation-final-Jan-22.pdf** (accessed 16 06 2026).

Thompson-Lee, S., See, B.H. and Klassen, R.M. (2025) 'A Systematic Review of STEM Teacher Recruitment and Retention Interventions'. *International Journal of STEM Education,* 12, 33. **https://doi.org/10.1186/s40594-025-00550-6**

Williamson, E., Morley, R., Lucas, R. and Carpenter, J. (2011) 'Propensity Scores: From Naive Enthusiasm to Intuitive Understanding'. *Statistical Methods in Medical Research,* 21: 3, 273–283. **https://doi.org/10.1177/0962280210394483**

Wing, C., Simon, K. and Bello-Gomez, R.A. (2018) 'Designing Difference in Difference Studies: Best Practices for Public Health Policy Research'. *Annual Review of Public Health,* 39, 453–469. **https://doi.org/10.1146/annurev-publhealth-040617-013507**

What Works Clearinghouse (2022) *'What Works Clearinghouse procedures and standards handbook'*. , National Center for Education Evaluation (NCEE), Institute of Education Sciences, U.S. Department of Education. Available at: **https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf** (accessed 16 06 2026)

Woolridge, J.M. (2012) *'Introductory Econometrics: A Modern Approach'*. Mason, OH: South-Western Cengage Learning.

Appendix A: EEF cost rating

Appendix A Table 1: Cost rating

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Effect size estimation

Appendix B Table 1: Effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance	Population variance (if applicable)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
Progression to STEM A level	0.0090	0.0142	73 (0)	0.0039	207 (0)	0.0037	0.0038	–
STEM teacher turnover	-0.0070	-0.0005	73 (0)	0.0259	208 (0)	0.0331	0.0313	–
STEM teacher wastage	-0.0120	0.0020	73 (0)	0.0081	208 (0)	0.0110	0.0103	–

Note: Numbers rounded to four decimal places.

Appendix C: Propensity score matching and assessing balance

Unmatched sample

Initially the balance in the unmatched sample between the treated and control groups was assessed. We used the standardised percentage of bias to provide a standardised measure of imbalance between treated and untreated subjects. Standardised percentage of bias is calculated as the percentage difference of the sample means between the treatment and control groups, as a percentage of the square root of the average of the sample variance between the two groups.[37] Since this percentage of bias is standardised, it helps address any large differences in means, making it easier to interpret the magnitude of the imbalance across variables with different scales and units of measurement. A positive standardised percentage of bias indicates a higher mean in the treatment group compared to the control group, while a negative standardised percentage of bias demonstrates a lower mean. A large, standardised percentage of bias in either direction is indicative of an imbalance between the treatment group and control group.

Other metrics, such as the standardised difference, could also have been used and would present results with a similar inference. Standardised differences compare the means in units of pooled SD. However, a limitation of this measure to assess balance is the lack of consensus regarding the threshold value of standardised difference that indicates imbalance between treated and untreated subjects, which can make it difficult to interpret the results.

Appendix C Table 1 shows there is imbalance across the range of covariates used for matching, between the secondary schools that received Aspire to STEM (treatment) and those that did not (control) settings. For most covariates, the absolute percentage bias indicates considerable imbalance. In some cases, there is a higher mean in the treatment group and in others there was a higher mean in the control group. Overall, the secondary school sample is shown to have an absolute average bias of 13.3%, highlighting the need for propensity score matching to improve compatibility between groups.

Appendix C Table 1: Balance in the unmatched sample

Variable	*Treatment*	*Control*	*% of bias*
Average maths component of Attainment 8 in 2015/2016	8.9027	8.7445	16.0
Average maths component of Attainment 8 in 2016/2017	7.8514	7.8483	0.3
% achieving EBacc 2 sciences component[a] in 2015/2016	0.48203	0.49584	-9.2
% achieving EBacc 2 sciences component in 2016/2017	0.47149	0.4823	-8.0
% eligible for FSM in 2017/2018	36.649	34.871	12.6
% with EAL status in 2017/2018	10.635	13.809	-19.4
% local authority-maintained schools	0.27027	0.37799	-23.1
% mixed sex schools	0.95946	0.96651	-3.7
Average number of pupils per school in 2017/2018	911.07	876.01	11.0
Exposure to other interventions	0.51351	0.36842	29.4
Mean absolute % of bias			13.3
Treated N			115
Control N			302

Notes: Means of covariates of secondary school settings are presented for the treatment group in column one, and for the control group in column two. '% of bias' is generated using the post-estimation commands from the 'psmatch2' package in Stata 17,[38] and is defined as: the % difference of the sample means in the treated and control sub-samples as a percentage of the square root of the average of the sample variances in the treated and control groups. 'Mean absolute % of bias' is the mean of the absolute values of the % of Bias presented in the table across all covariates.

[a] *This requires students to attain a GCSE pass (grades 4–9) in at least two of the following science disciplines: biology; chemistry; physics; or computer science.*

[37] Leuven, E. and Sianesi, B. (2018) *'PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing'*. Statistical Software Components S432001. Boston, MA: Boston College Department of Economics.

[38] Leuven, E. and Sianesi, B. (2018) '*PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*'. Statistical Software Components S432001. Boston, MA: Boston College Department of Economics.

To explore which propensity score matching strategy produces the best sample size and the greatest reduction in bias (i.e. generates the smallest difference between the treatment and control groups across the covariates included), we tested several strategies. The standardised percentage of bias for each covariate, and the absolute mean of this standardised percentage of bias, for each of these strategies is presented in Appendix C Table 2.

Appendix C Table 2: Percentage bias with other specifications for secondary school sample

Variable	**1 nearest neighbour No replacement average exam outcomes**	**1 nearest neighbour No replacement percentage outcomes**	**2 nearest neighbour average exam outcomes**	**Calliper of 1/4 of SD**	**Kernel**	**Coarsened exact matching**	**Unmatched sample**
Average maths component of Attainment 8 in 2015/2016	-1.4	7.4	-7.4	-3.7	1.9	N/A	16.0
Average maths component of Attainment 8 in 2016/2017	3.4	5.9	-1.6	0.6	-0.5	N/A	0.3
% achieving EBacc 2 sciences component[a] in 2015/2016	9.1	6.2	-0.7	8.8	1.7	N/A	-9.2
% achieving EBacc 2 sciences component in 2016/2017	14.2	9.9	9.6	12.8	2.1	N/A	-8
% eligible for FSM in 2017/2018	6.1	-12.0	10.1	8.0	2.5	N/A	12.6
% with EAL status in 2017/2018	-3.9	-1.1	4.8	-3.3	-0.7	N/A	-19.4
% Local authority-maintained schools	2.9	-5.9	2.9	3.0	-2	N/A	-23.1
% mixed sex schools	7.2	-7.2	10.8	7.4	6.4	N/A	-3.7
Average number of pupils per school in 2017/2018	9.7	-9.1	-0.5	14.7	-3.4	N/A	11
Exposure to other interventions	-8.3	8.3	11.1	-8.6	-0.3	N/A	29.4
Mean absolute % of bias	6.62	7.3	6.0	7.09	2.15	N/A	13.3
Treated N	73	73	73	71	73	42	115
Control N	73	73	94	71	208	62	302
Effective sample size	146	146	148	142	199.69	75	N/A

Notes: Different specifications of propensity score matching strategies are presented for secondary schools. One nearest neighbour with no replacement and the average outcome in settings shown in column one; 1 nearest neighbour with no replacement and the percentage outcomes in settings shown in column two; 2 nearest neighbours with average outcomes in column three; 1 nearest neighbour with no replacement and a calliper equal to 1/4 of SD shown in column four; kernel matching shown in column five; coarsened exact matching in column six; and the unmatched sample for reference in column seven. '% of bias' is generated using the post-estimation commands from the 'psmatch2' package in Stata 17,[39] *and is defined as: the % difference of the sample means in the treated and control sub-samples as a percentage of the square root of the average of the sample variances in the treated and control groups. 'Mean absolute % of bias' is the mean of the absolute values of the % of bias presented in the table across all covariates.*

[a] *This requires students to attain a GCSE pass (grades 4–9) in at least two of the following science disciplines: biology; chemistry; physics; or computer science.*

EBacc=English Baccalaureate; N/A=not applicable.

As is clear from Appendix C Table 2, the sample size generated from the coarsened exact matching process was so small (n=75) that it was not considered a valid strategy in this context. When including all of the covariates used in other matching strategies to conduct coarsened exact matching, the settings included in the matched sample are only those which have missing values among the covariates. Therefore, we additionally tested the impact of removing covariates from the matching coarsened exact matching strategy, including the removal of the average attainment scores in 2016 (i.e. only

[39] Leuven, E. and Sianesi, B. (2018) '*PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*'. Statistical Software Components S432001. Boston, MA: Boston College Department of Economics.

including attainment in 2017), and separately the removal of the school gender mix and whether the school is independent or maintained. When testing these additional coarsened exact matching strategies, the sample size did not drastically improve. In addition, we tested several different cutoff strategies including Freedman-Diaconis, Shimazaki-Shinomoto, Sturges, and Scott cut-offs, none of which significantly increased the sample size. Finally, we tested our own coarse cut-offs which placed observations into groups based on two equally spaced cut-offs, the effective sample size generated is considerably smaller than for any other method of matching even when such a crude coarsened exact matching strategy is implemented. Therefore, while we present the sample size generated through this method, we do not present an assessment of the reduction in bias, as it is not possible to assess a reduction in bias for observations which are largely matched on missing values.

As is shown in Appendix C Table 2 above, both 1 nearest neighbour with no replacement strategies result in smaller mean absolute standardised percentage of bias than the unmatched sample, but do not perform as well as the kernel matching strategy. The 1 nearest neighbour strategy which uses the average outcomes in each setting results in a greater reduction of bias than the 1 nearest neighbour strategy which uses the percentage of students achieving minimum outcomes. As a result, all other strategies reported use average outcomes rather than percentage outcomes. We additionally tested a 1:2 nearest neighbour matching strategy, which resulted in only a nominally larger effective sample size and a largely similar level of reduction in bias as the 1 nearest neighbour strategies.

Using a calliper of 1/4 of SD has very little impact on the matching process. Further callipers were tested up to 1/10 of SD but altering the calliper widths did not substantially change these results and a calliper of 1/4 of SD was used as the final specification included in Appendix C Table 2. The sample size remains very similar, with only two settings lost from the treatment and two from the control, and there is no reduction in the mean absolute percentage of bias.

The kernel[40] matching strategy results in the smallest mean absolute standardised percentage of bias at 2.15%, in contrast to the 13.3% in the unmatched sample. Of particular note is the reduction in the average maths attainment in 2015/2016 (reduced from 16% bias to 1.9% bias), the percentage eligible for FSM (reduced from 12.5% bias to 2.5% bias) and the total number of pupils (reduced from 29.4% bias to -0.3% bias). However, kernel matching does result in a slight increase of bias in the school gender mix (from -3.7% to 6.4%) and a very small increase in bias for the average maths attainment in 2016/2017 (from 0.3% to -0.5%). In addition to representing the greatest reduction in the mean absolute standardised percentage of bias, kernel matching also generates the largest sample size.

The matched sample following the kernel matching process includes 208 settings in the control and 71 settings in the treatment. However, kernel matching assigns weights (between 0.01 and 0.99) to the settings in the control and so a count of the number of control and treatment settings does not reflect the weighting assigned to each unit. We therefore, calculate the effective sample size using the following equation:[41]

$$Effective\ Sample\ Size = \frac{(\sum Weights)^2}{\sum(Weights^2)}$$

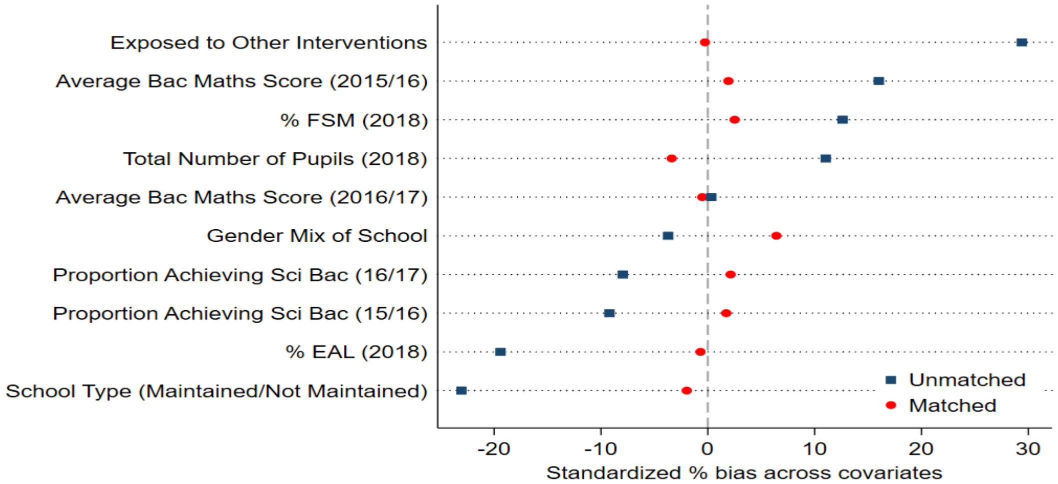
By adjusting for the weights generated during the kernel matching, the effective sample size gives a more realistic indication of the relative sample size gains made in using kernel matching, when compared to other non-weighting approaches. The effective sample size for kernel matching is calculated at approximately 200 observations. For comparison, the total sample size for 1 nearest neighbour matching with no replacement, calculated by adding the treatment and control settings together, is 146 observations. Therefore, even after accounting for the weighting, there are substantial gains to sample size under kernel matching.

[40] Using a Gaussian kernel function.
[41] Golinelli, D., Ridgeway, G., Rhoades, H., Tucker, J. and Wenzel, S. (2012) 'Bias and Variance Trade-Offs When Combining Propensity Score Weighting and Regression: With an Application to HIV Status and Homeless Men'. *Health Services and Outcomes Research Methodology,* 12, 104–118. https://doi.org/10.1007/s10742-012-0090-1

Given the kernel approach resulted in the smallest mean absolute standardised percentage of bias, and the largest effective sample size, we use kernel matching in our main analysis. The balance between the treatment and matched counterfactual groups after kernel matching is presented in greater detail in Appendix C Table 3. The mean of the treatment and weighted control group for each of the covariates used in the matching process, as well as the standardised percentage of bias and mean absolute standardised percentage of bias are presented.

The standardised percentage of bias is also shown in Appendix C Figure 1, where it is compared with the standardised percentage of bias for the unmatched sample (from Appendix C Table 1). It demonstrates the reduction in bias as a result of the kernel matching process, as for all covariates other than the gender mix of schools the percentage of bias is considerably lower in the matched sample.

Appendix C Figure 1: Standardised percentage of bias for the kernel matched and unmatched samples



Note: The standardised percentage of bias across the covariates used during the kernel propensity scoring matching process for the unmatched sample (blue) and the matched sample (red).

Appendix C Table 3: Balance in the kernel matched secondary school sample

Variable	*Treatment*	*Control*	*% of bias*
Average maths component of Attainment 8 in 2015/2016	8.9137	8.8947	1.9
Average maths component of Attainment 8 in 2016/2017	7.8575	7.8625	-0.5
% achieving EBacc 2 sciences component[a] in 2015/16	0.48411	0.48152	1.7
% achieving EBacc 2 sciences component in 2016/2017	0.47479	0.4719	2.1
% eligible for FSM in 2017/2018	36.11	35.755	2.5
% with EAL status in 2017/2018	10.767	10.88	-0.7
% Local authority-maintained schools	0.27397	0.28321	-2
% mixed sex schools	0.9589	0.94675	6.4
Average number of pupils per school in 2017/2018	911.12	921.88	-3.4
Exposure to other interventions	0.50685	0.50817	-0.3
Mean absolute % of bias			2.15
Treated N			73
Control N			208
Effective sample size			199.69

Notes: Means of covariates of secondary school settings are presented for the matched sample following kernel matching. The treatment group is shown in column one, and the control group in column two. '% of bias' is generated using the post-estimation commands from the 'psmatch2' package in Stata 17,[42] and is defined as: the % difference of the sample means in the treated and control sub-samples as a percentage of the square root of the average of the sample variances in the treated and control groups. 'Mean absolute % of bias' is the mean of the absolute values of the % of bias presented in the table across all covariates.

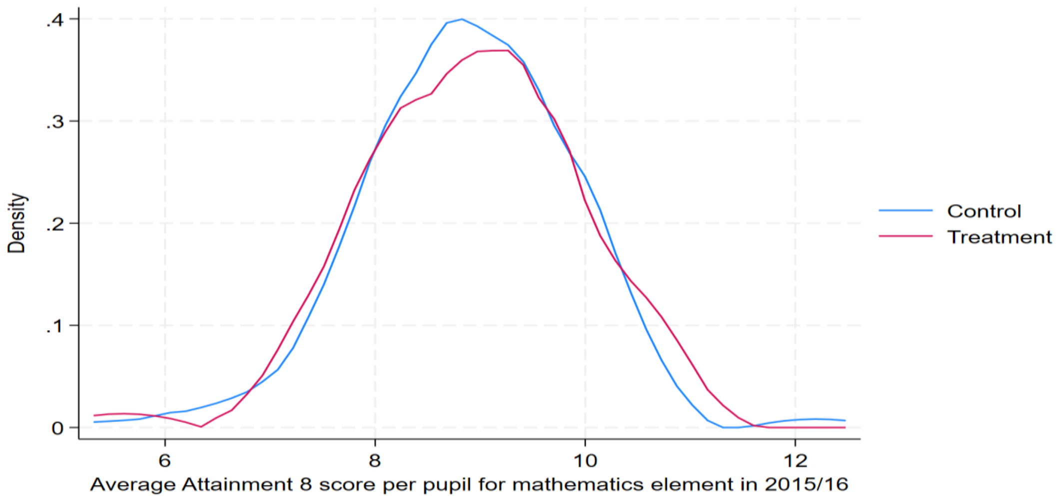
[a] This requires students to attain a GCSE pass (grades 4–9) in at least two of the following science disciplines: biology; chemistry; physics; or computer science.

[42] Leuven, E. and Sianesi, B. (2018) *'PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing'*. Statistical Software Components S432001. Boston, MA: Boston College Department of Economics.

In addition to examining the reduction in bias in the means of variables used during the matching process, we also examine the full distribution of continuous variables in both the treatment and control settings. In order to demonstrate the similarities in the distribution of continuous matching variables between the treatment and control settings, we plot kernel density plots. In the figures below, we present a comparison of plots for the matched treatment group (i.e. only treated units which are in the matched sample) and the matched control group (weighted as appropriate from the kernel matching methods utilised). This allows us to compare the distribution of each matching variable in the matched treatment and control groups. We do this for the maths component of Attainment 8 in 2015/2016 and 2016/2017, percentage achieving EBacc 2 sciences component in 2015/2016 and 2016/2017, for the percentage of pupils who are eligible for FSM and who have EAL status in 2017/2018, and the total number of pupils in the setting in 2017/2018.

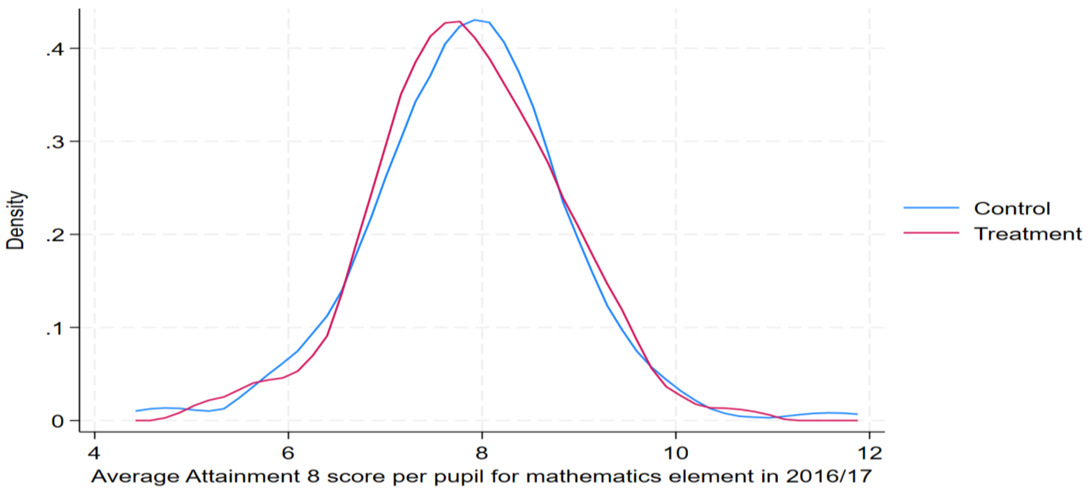
All the density plots show clear commonalities in the distribution of the treatment and control groups. However, it is important to note that Appendix C Figure 2 reveals that the peak of the distribution in the percentage achieving the EBacc 2 sciences component in 2015/2016 for the treatment group is lower than that of the control group. Despite the difference in the level of the distribution, the two distributions follow a similar pattern across the entire outcome and there are no statistically significant differences in means between the treatment and matched counterfactual. Therefore, we conclude that the kernel matching process produces an accurate match between the treatment and control settings with regard to covariate distribution as well as means.

Appendix C Figure 2: Distribution of average Maths attainment 8 score in 2015/2016 for treated and control settings



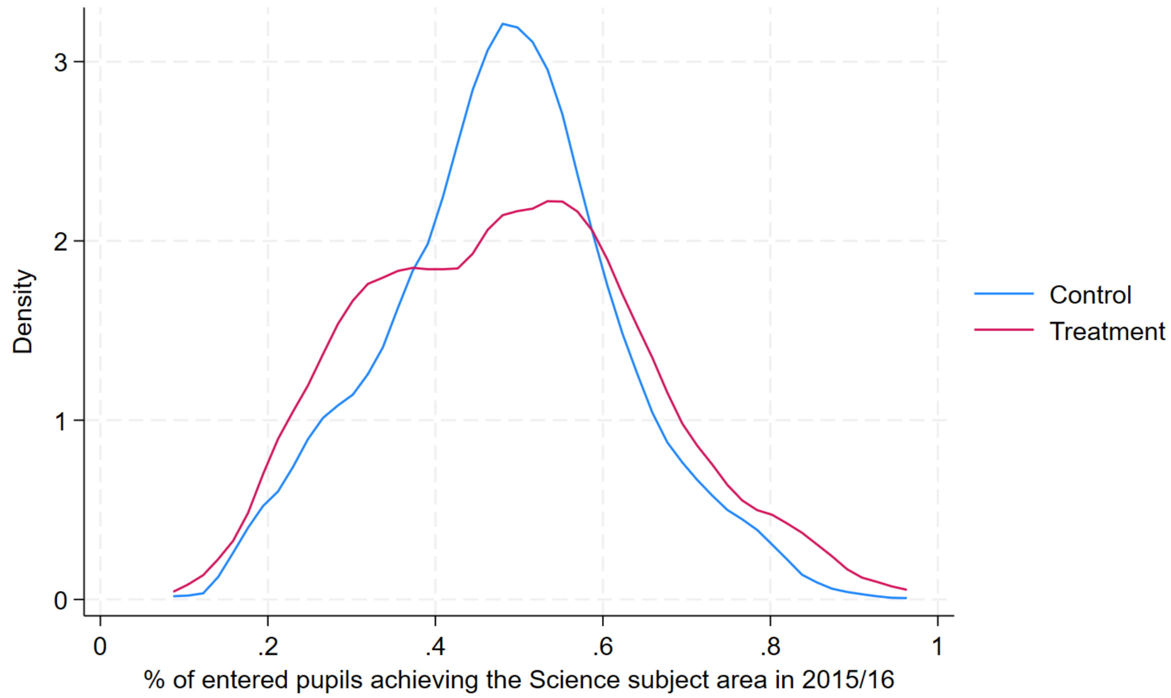
Notes: Kernel density plot of the average maths component of Attainment 8 in 2015/2016 for the treatment (red) and the control settings (blue).

Appendix C Figure 3: Distribution of average Maths attainment 8 score in 2016/2017 for treated and control settings



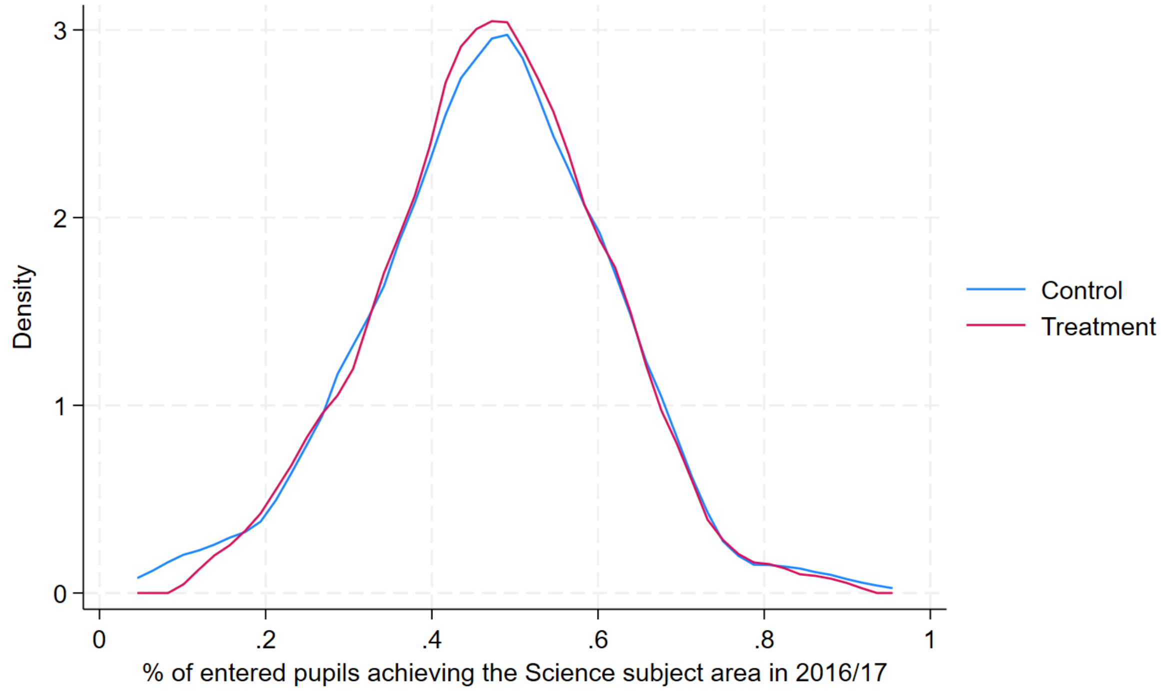
Notes: Kernel density plot of the average maths component of Attainment 8 in 2016/2017 for the treatment (red) and the control settings (blue).

Appendix C Figure 4: Distribution of % achieving the science attainment score in 2015/2016 for treated and control settings



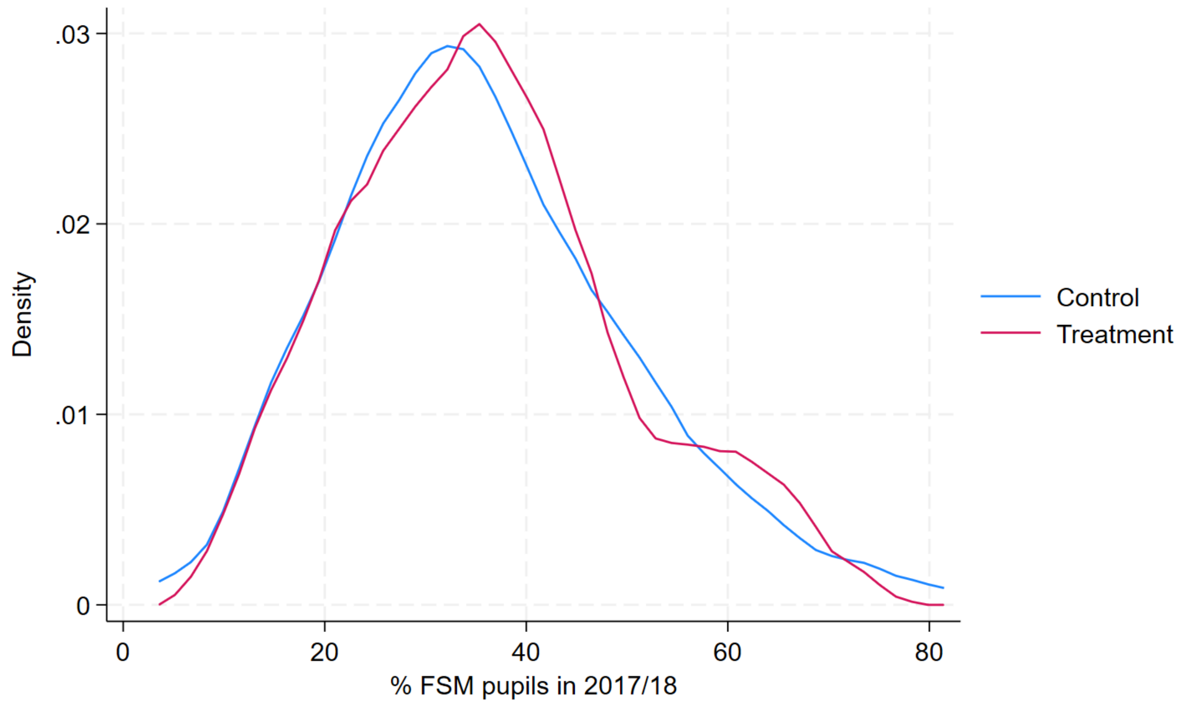
Notes: Kernel density plot of the % of pupils who achieved the EBacc 2 sciences component in 2015/2016 (this requires students to attain a GCSE pass (grades A–C/9–4) in at least two of the following science disciplines: biology; chemistry; physics; or computer science). Treatment (red) and the control settings (blue).*

Appendix C Figure 5: Distribution of % achieving the science attainment score in 2016/2017 for treated and control settings



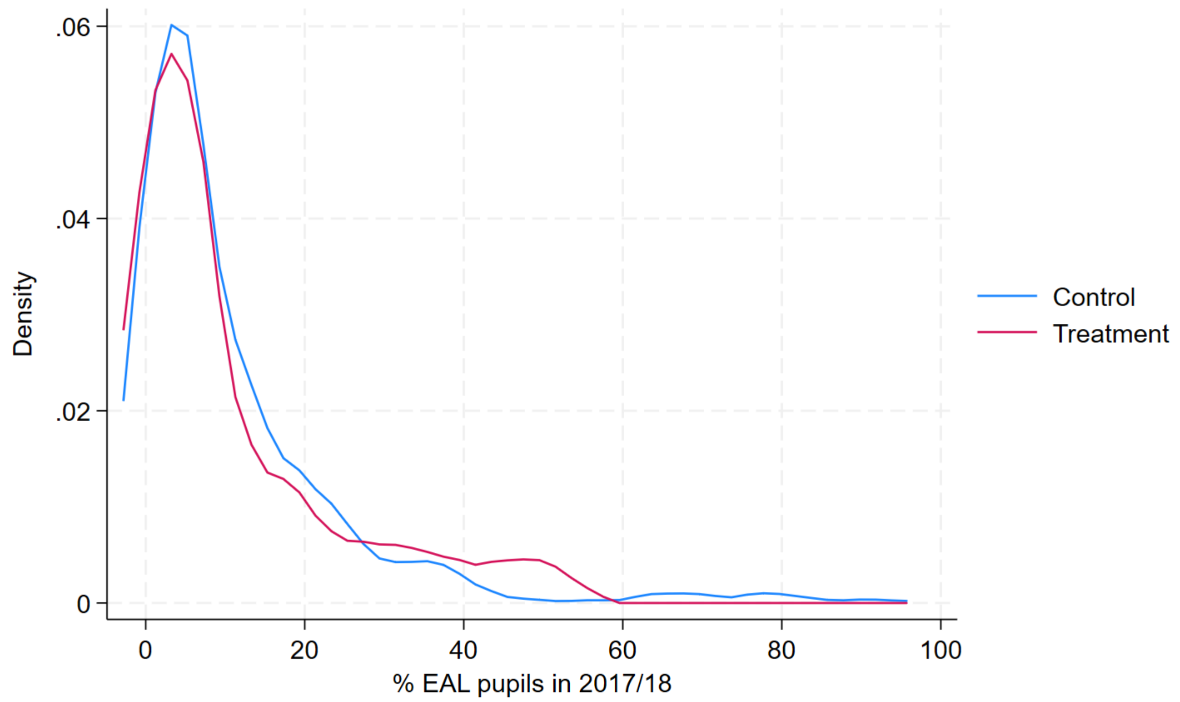
Notes: Kernel density plot of the % of pupils achieving the EBacc 2 sciences component in 2016/2017 (this requires students to attain a GCSE pass (grades A–C/9–4) in at least two of the following science disciplines: biology; chemistry; physics; or computer science). Treatment (red) and the control settings (blue).*

Appendix C Figure 6: Distribution of percentage eligible for FSM in 2017/2018 for treated and control settings



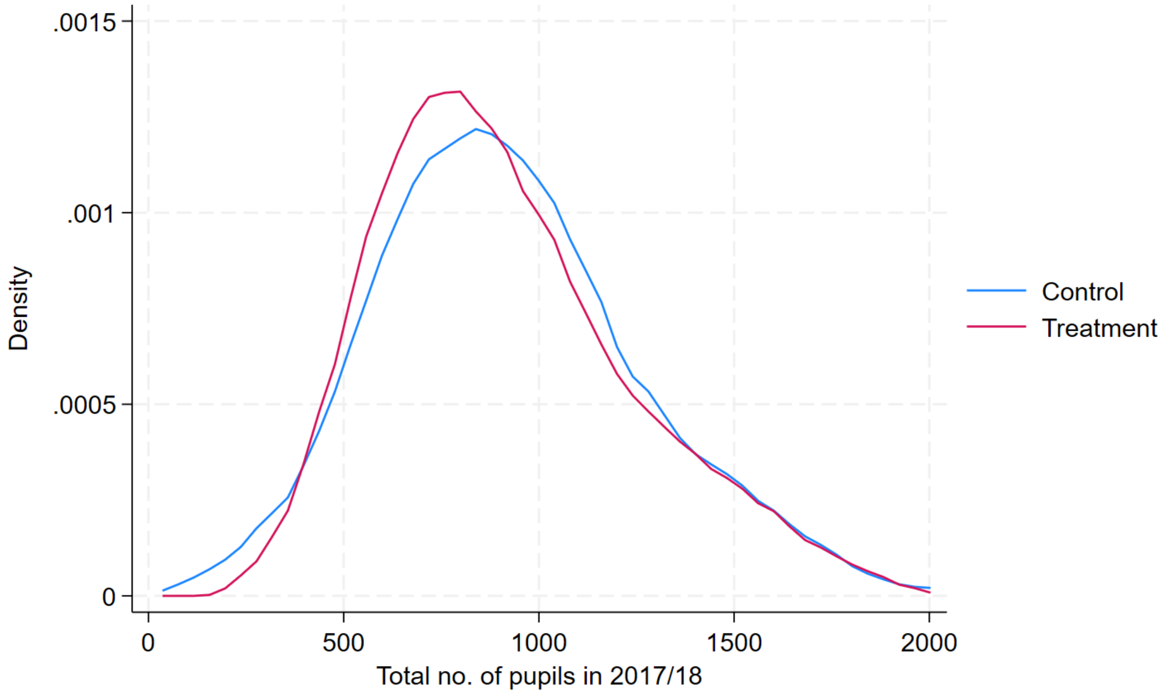
Notes: Kernel density plot of the % of pupils who are eligible for FSM in 2017/2018 for the treatment (red) and the control settings (blue).

Appendix C Figure 7: Distribution of percentage who have EAL status in 2017/2018 for treated and control settings



Notes: Kernel density plot of the % of pupils who have EAL status in 2017/2018 for the treatment (red) and the control settings (blue).

Appendix C Figure 8: Distribution of the number of pupils in 2017/2018 for treated and control settings



Notes: Kernel density plot of the total number of pupils in the setting in 2017/2018 for the treatment (red) and the control settings (blue).

Appendix D: Pupil-level outcome matching process within the NPD

Over the five academic years 2014/2015 to 2018/2019 a total of 295,963 pupils were identified in the School Census as completing key stage 4 at schools enrolled in Aspire to STEM and comparator schools. Of those, 253,576 (85.7%) were successfully matched to key stage 5 Exam records two years after completing key stage 4. Appendix D Table 1 illustrates that the match rate is broadly consistent across years.

Appendix D Table 1: Matching to key stage 5 Exam records across treatment and counterfactual schools

Academic year completed key stage 4	Counterfactual		Treatment		Total	
	Schools N	Pupils N	Schools N	Pupils N	Schools N	Pupils N
2014/2015	249	45,104	86	15,719	**335**	**60,823**
2015/2016	273	46,073	91	15,300	**364**	**61,373**
2016/2017	282	45,343	93	14,546	**375**	**59,889**
2017/2018	283	43,233	93	14,471	**376**	**57,704**
2018/2019	274	42,248	90	13,926	**364**	**56,174**

As shown in Appendix D Table 1, not all pupils were matched. We would not expect match rates to be perfect for a number of reasons. Some pupils will not continue in education and training (according to official statistics, in 2024, 6.2% of 16–17 years old were Not in Education, Employment, or Training[43]), others will complete qualifications not captured in the key stage 5 Exam table, and finally while two years is a typical length of study for A levels we know some young people take three or more years to sit exams.

Appendix D Table 2: Matching to key stage 5 Exam records in Aspire to STEM schools

Academic year completed key stage 4	Not matched to key stage 5 Exam table two academic years later	Matched to key stage 5 Exam table two academic years later	Total pupils
2014/2015	9,417	51,406	60,823
2015/2016	9,905	51,468	61,373
2016/2017	8,817	51,072	59,889
2017/2018	7,197	50,507	57,704
2018/2019	7,051	49,123	56,174
Total	**42,387**	**253,576**	**295,963**

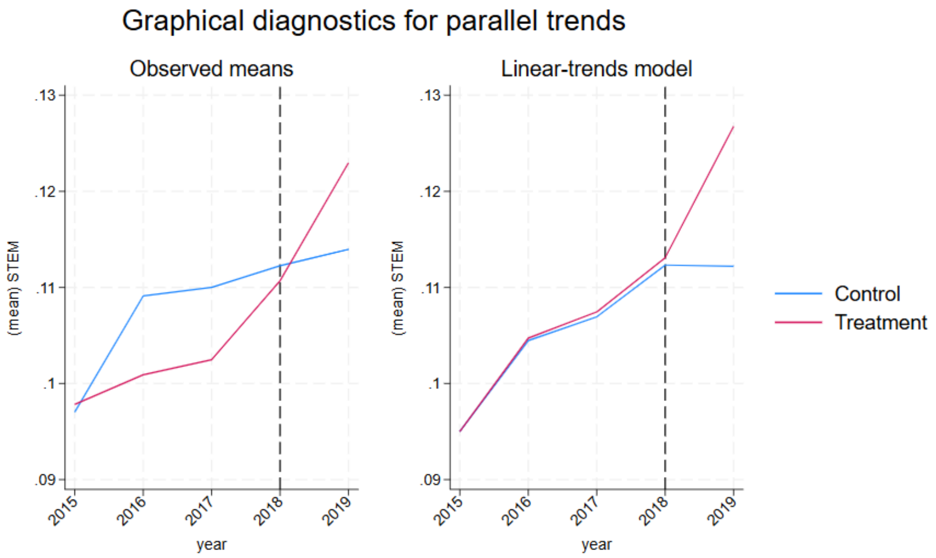
[43] See: **https://explore-education-statistics.service.gov.uk/find-statistics/participation-in-education-and-training-and-employment/2024**

Appendix E: Parallel trends assessment

Once the sample of pupils at the Aspire to STEM schools had been identified, we conducted a test of the parallel trends assumption, to establish if the average outcome would have evolved in the same (parallel) way in both the treatment and the counterfactual group had treatment not occurred. Appendix E Figure 1 plots the evolution of the progression to STEM A level outcome over time. The left panel plots the observed outcome means for each group and shows similar increases in the fraction of pupils entering STEM A levels when taken over the whole four pre-treatment time periods. The right panel, plots the estimates of a linear trends model. This is a fitted model estimated during the difference-in-differences process. It shows even more clearly that prior to treatment, the fraction of pupils entering STEM A levels in treatment and control schools evolved in a very similar way.

However, as a result of the slight difference in observed mean pre-intervention trends, we tested parallel trends statistically using a chi-squared test. This test failed to reject the null hypothesis of parallel trends (p = 0.906), supporting the position that pupil progression rates to STEM A levels in the treatment and control group evolved in similar ways prior to treatment, and the key assumption underpinning our identification strategy holds.

Appendix E Figure 1: Graphical diagnostic for parallel trends



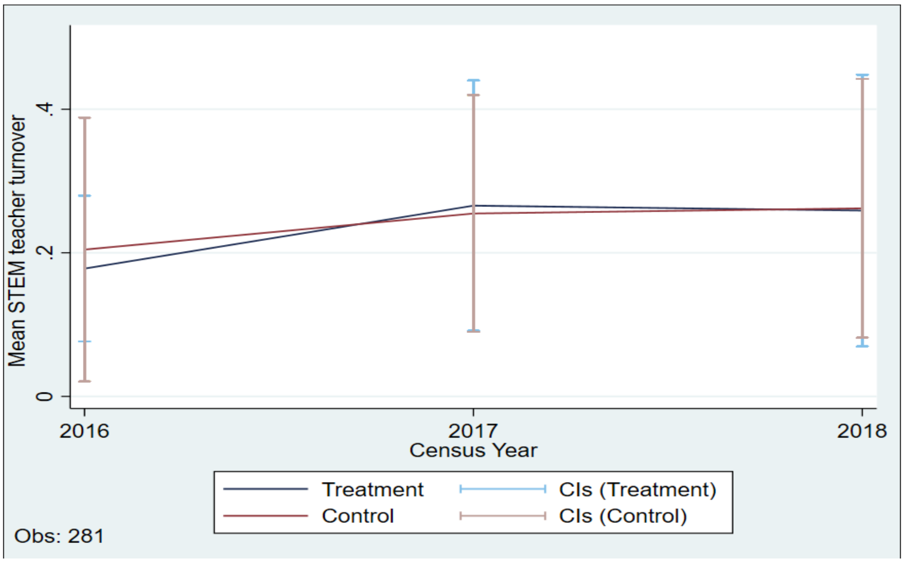
Notes: 'year' refers to the year pupils completed Key Stage 4, so year = 2019 refers to the 2018/2019 academic year. The outcome is the fraction of these pupils who entered at least one STEM A level two years later, so for year = 2019, the outcome is measured in the academic year 2020/2021.

Teacher-level parallel trends analysis

As with the pupil analysis, the parallel trends assumption is required to hold for valid inference to be drawn from the teacher outcome estimates. The teacher analysis uses a shorter pre-intervention period than the pupil analysis which reduces the confidence in the satisfaction of the parallel trend's assumption.

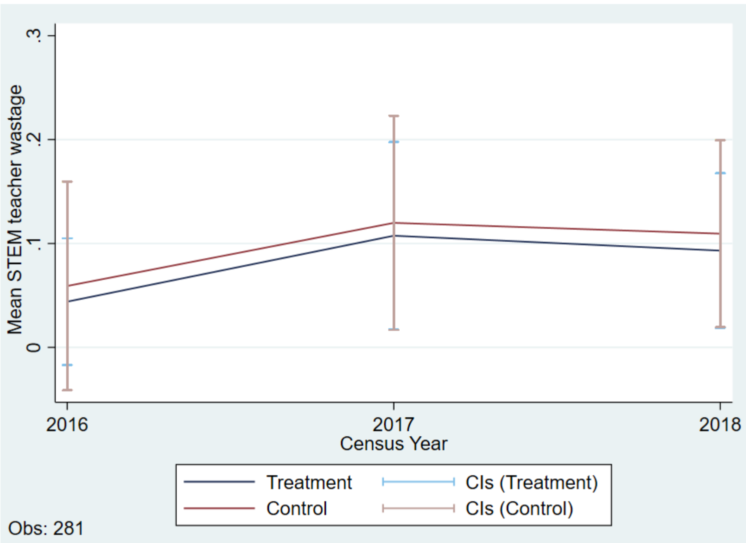
Appendix E Figures 2 and 3 plot the means of the teacher outcome variables. A visual inspection provides some evidence that both treatment and control schools likely shared parallel trends in the pre-intervention period for both the teacher turnover and teacher wastage outcomes. Given the smaller time window compared to the pupil analysis we do not present linear trend models. To confirm this, we tested parallel trend statistically using chi-squared tests. These tests failed to reject the null hypothesis of parallel trends (teacher turnover p = 0.117; teacher wastage p = 0.577). This supports the position that teacher outcomes in the treatment and counterfactual settings were similar prior to the intervention.

Appendix E Figure 2: Trends in the rate of teacher turnover for matched secondary schools; true means



Note: Trends are plotted as the true (observed) means of the outcome (rate of teacher turnover) for the treatment group and the control group. CIs are plotted for each time point.

Appendix E Figure 3: Trends in the rate of teacher wastage for matched secondary schools; true means



Note: Trends are plotted as the true (observed) means of the outcome (rate of teacher wastage) for the treatment group and the control group. CIs are plotted for each time point.

Education Endowment Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn