Focus for Teacher Assessment of Primary Science (Focus4TAPS)

Evaluation report

September 2026

Sarah Sharrock, Liza Benny, Sehaj Bhatti, Rebecca Parker, Kate Wadsworth, Joe Crowley, Kate Baird, and Alina Haque

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

Table of contents

About the evaluator

The project was independently evaluated by a team from the National Centre for Social Research: Sarah Sharrock; Dr. Hayley Leonard; Dr. Liza Benny; Dr. Georges Poquillon; Sehaj Bhatti; Rebecca Parker; Kate Wadsworth; Kate Baird; Alina Haque; Charlotte Tomlinson; Dr. Enes Duysak; Dr. Terry Ng-Knight; and Dr. Gayle Munro.

The lead evaluator was Dr. Hayley Leonard up until endline testing. Sarah Sharrock was the principal investigator who led the evaluation through to completion of endline testing, analysis, and reporting.

Contact details:

Sarah Sharrock
National Centre for Social Research
250 City Road,
London,
United Kingdom,
EC1V 2AB

Email: Sarah.Sharrock@natcen.ac.uk

Acknowledgements

We are grateful to the many people whose contributions made this evaluation possible. Our sincere thanks go to the schools who participated in the evaluation, including pupils, teachers, and science leads who contributed their time, views, and experiences. Thanks too to the Focus for Teacher Assessment of Primary Science (Focus4TAPS) trainers who took part, and to Sarah Earle, the programme developer at Bath Spa University, whose contribution and collaboration has been invaluable to the success of this research. For their work to coordinate and collect testing data, we also thank the AlphaPlus team led by Mibéka Tan Panza, Andy Lewis, and Serena Idris, and the AlphaPlus testers who travelled nationwide to collect outcomes data in schools. Finally, we would like to thank our colleagues at the Education Endowment Foundation, especially Flavy Sen Sharma and Pamela Shaw, for their steady support throughout.

Executive summary

The project

Focus for Teacher Assessment of Primary Science (Focus4TAPS) is a professional development programme designed to support primary school teachers to improve formative assessment in science and strengthen pupils' Working Scientifically skills. These skills focus specifically on how scientific knowledge is built, how investigations are designed and carried out, and how evidence is evaluated and communicated. The programme is based on the Teacher Assessment in Primary Science (TAPS) model developed at Bath Spa University, which emphasises using rich, classroom-based evidence to inform teaching and learning. A previous **efficacy trial** of the programme (Mujtaba *et al.,* 2022) showed promising impacts on pupils' progress in science, which motivated further research to assess its effectiveness with a larger group of schools. As a part of the programme, teachers attended three full-day training sessions across the 2024/2025 school year and were asked to deliver at least five Focused Assessment lessons periodically across the year. Science leads were also invited to participate in the training to support in-school implementation and encourage wider adoption of the programme.

The evaluation was a two-arm cluster randomised effectiveness trial that involved 240 schools, with one Year 5 class per school selected to participate, across 13 regions in England. In total, the final sample for analysis included 5,142 Year 5 pupils. Schools were randomly allocated either to deliver Focus4TAPS or continue with their usual science teaching over the trial period, which ran across the 2024/2025 school academic year. The evaluation was coupled with an implementation and process evaluation (IPE) that included teacher surveys, interviews, focus groups, observations of teacher training and of lesson delivery, and cost analysis.

Table 1: Key conclusions

Key conclusions	
1.	Pupils in Focus4TAPS schools made no additional months' progress in science attainment, on average, compared to pupils in other schools. This result has a moderate to high security rating.
2.	Among pupils eligible for free school meals (FSM), those in Focus4TAPS schools made one month's additional progress in science attainment, on average, compared to those in other schools. However, there is uncertainty around these findings as the possible range of impact estimates include both a reduction and increase in pupil science attainment, and the results may have lower security than the overall findings because of the smaller number of pupils.
3.	Pupils in Focus4TAPS schools reported more positive attitudes and beliefs about science teaching and learning than pupils in other schools.
4.	Focus4TAPS training was implemented largely as intended, with consistent delivery of programme content. However, the number of lessons teachers delivered varied considerably, and, given the limited availability of data on this, it is not possible to determine whether more consistent lesson delivery would have led to improved attainment outcomes.
5.	Teachers viewed the programme positively and reported increased confidence in teaching science, particularly in relation to using formative assessment and developing pupils' Working Scientifically skills. Pupils were observed taking a more active role in scientific enquiry. However, the extent to which these practices were consistently embedded varied across schools.

EEF security rating

These findings have a moderate to high security rating. This was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was well powered. However, the following factors reduced the security of the trial. Around 21.1% of the pupils who started the trial were not included in the final analysis because they were absent at the time of endline testing or their Key Stage 1 attainment data could not be used, either due to non-consent or records being unavailable. There were some differences in prior attainment and FSM status between pupils in schools that delivered Focus4TAPS and those in comparison schools.

Additional findings

Pupils in Focus4TAPS schools made no additional months' progress in science attainment compared to pupils in the teaching as usual group. This is our best estimate of impact, which has a moderate to high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme ranges from two months' less progress, to up to two months' additional progress. A robustness check, which adjusted for baseline differences between groups and used an alternative specification of Key Stage 1 reading and maths attainment consistent with the efficacy trial, produced a larger effect estimate equivalent to approximately one month's additional progress. However, the findings remained inconclusive and included both negative and positive impacts on pupil attainment.

Pupils eligible for FSM in Focus4TAPS schools made, on average, one additional month's progress in science attainment, compared to pupils in the teaching as usual group. The analysis was based on a smaller subgroup of pupils, and the possible impact of the programme ranged from two months' less progress to three months' additional progress, meaning that the overall finding was inconclusive. Across secondary outcomes, small positive effects were observed in pupils' attitudes towards science learning and teaching in Focus4TAPS schools compared to those in the teaching as usual group.

The IPE found that the Focus4TAPS training was delivered largely as intended. Training workshops were delivered consistently across regions. Teachers who attended generally engaged positively with the content and structure of the programme by using Focused Assessment lesson plans, focusing on individual elements of Working Scientifically, and drawing on formative assessment to adapt teaching and support pupil learning. Where adaptations occurred, these were aligned with programme guidance. Among teachers who responded to the endline survey, just over half reported delivering the intended minimum of five lessons, though given the low survey response rate this should be interpreted with caution. Teachers reported that some scientific enquiry elements like focused recording and self-assessment were easier to include, and that activities such as peer assessment were harder to apply consistently across their lessons. Further, some elements of the Focus4TAPS approach closely resembled practices already used in schools and some schools were already using Primary Science Teaching Trust resources (which includes TAPS resources) prior to the trial. This overlap may have reduced the distinctiveness of Focus4TAPS relative to teaching as usual, particularly where fewer Focused Assessment lessons were delivered.

Lesson observations in Focus4TAPS group schools showed pupils taking ownership of their learning, including planning elements of investigations, selecting variables, and recording their own results. Teachers highlighted that the approach supported a positive classroom culture and engagement with scientific investigation, and that practical activities were particularly inclusive for pupils who were less confident in science. Pupils in focus groups also reported being more aware of what they were learning and why, with many able to articulate the specific skill they were practising. Our findings suggest that while Focus4TAPS led to valuable shifts in classroom practice, variability in the number of lessons delivered is one potential explanation that shaped the extent to which these changes translated into measurable improvements for pupils.

Cost

The average cost of delivering Focus4TAPS in the trial for one school/class was £4,085.36, or £59.21 per pupil per year when averaged over three years, making it a very low-cost programme.

Impact

Table 2: Summary of impact on primary outcome

Outcome / group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P-value	EEF cost rating
Science attainment	0.03 (-0.11 to 0.17)	0		5,142	0.69	£ £ £ £ £
Science attainment – FSM pupils	0.07 (-0.12 to 0.25)	1	n/a	1,633	0.48	n/a

Introduction

Background

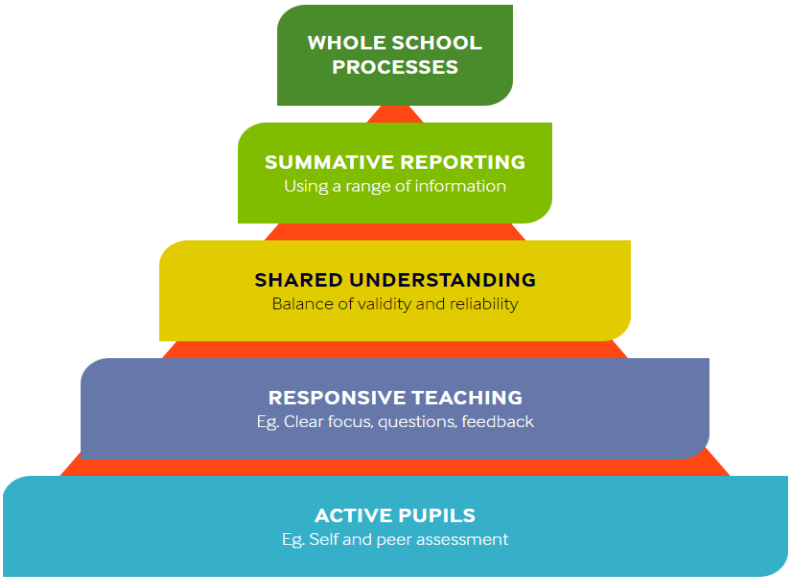
The Focus for Teacher Assessment of Primary Science (Focus4TAPS) programme aims to improve science attainment among pupils in primary schools by improving teaching approaches and assessment in science. It is a primary school science intervention based on the 'Teacher Assessment in Primary Science (TAPS)' model developed by Bath Spa University with funding from the Primary Science Teaching Trust (PSTT). In the TAPS model, teachers routinely collect rich formative assessment data during everyday classroom work to make informed judgements about pupils' progress and adapt their teaching accordingly. The programme model also encourages teachers to adopt a 'Focused Assessment' approach, focusing each lesson on a single aspect of scientific enquiry (Davies and McMahon, 2003; Earle, 2021).

The programme centres on the teaching and assessment of 'Working Scientifically' in the English National Curriculum, which relates to the set of scientific enquiry skills pupils use in their practical science lessons to investigate questions, collect evidence, and draw conclusions—such as observing, measuring, predicting, and interpreting results. For Year 5 pupils, Working Scientifically involves: planning different types of investigations and controlling variables; taking accurate and precise measurements using a range of equipment; recording data using increasingly complex graphs and diagrams; interpreting and reporting results; and linking them to scientific evidence (PSTT, 2022). Figure 1 shows the six Working Scientifically elements covered by the Focus4TAPS approach: 'ask questions and plan enquiry' and 'set up and predict' under **Plan**; 'observe and measure' and 'record' under **Do**; and 'interpret and report' and 'evaluate' under **Review**.

Figure 1: Working Scientifically wheel

Over the course of an academic year, Focus4TAPS comprises a series of three professional development training sessions and an online bank of 60+ resources to support implementation of at least five Focused Assessment lessons, each concentrating on a specific element of scientific enquiry. These lesson resources include an identified focus for assessment, learning objectives and criteria for success, and guidance on how to interpret pupils' responses. The programme's wider guidance aims to foster and support assessment for learning across the science curriculum and the wider school context. In this trial, the programme was delivered in Year 5 classes.

Figure 2: TAPS pyramid model (Davies *et al.,* 2017)



Policy background

While science is considered a core subject in the English primary curriculum, evidence suggests improvements are needed to enhance the quality of science teaching in primary schools. For example, the Office for Standards in Education, Children's Services and Skills (Ofsted) 2021 science research review found a downward trend in pupils' performance in national Key Stage 2 science sampling tests after the 2014 curriculum was introduced.[1] Specifically, the percentage of pupils reaching the expected standard in science fell from 28% in 2014 to 21% in 2018 (Ofsted, 2021). Research has also demonstrated that science attainment in England compares negatively with other countries. In the 2019 Trends in International Mathematics and Science Study (TIMSS), for example, attainment among pupils aged 10–11 in England ranked outside that of the top ten participating countries (Richardson *et al.,* 2020).

Primary school teachers report relatively low levels of science-specific qualifications or training, a key influence on pupil outcomes. For example, more than two-thirds of primary teachers surveyed as part of the Wellcome Trust research had not studied science subjects beyond GCSE level or equivalent; almost half of respondents felt their initial teacher training had not adequately prepared them to assess pupils' attainment in science. (Wellcome Trust, 2017). Building on this, the State of Primary Science in the UK report has highlighted a decline in teachers' confidence, support, and access to subject-specific training in the period since 2020 (McCormick, 2025). Drawing on national survey data from teachers and science leaders, this research concluded that science was often allocated less curriculum time and fewer resources than other core subjects. Teachers reported difficulties in teaching scientific enquiry and progression, reflected in variability in the quality of science provision and pupils' learning experiences (McCormick, 2025).

Research suggests that a formative assessment approach can be used to effectively engage pupils in their own learning, contributing to improving pupil learning outcomes (Estyn, 2022). Formative assessment (also known as 'assessment for learning') involves teachers gathering evidence during lessons to understand pupils' progress and adapt their teaching accordingly (Wiliam, 2011). By providing timely feedback and insight into pupils' understanding, formative assessment enables teachers to tailor instruction and support learning in real time. This may be especially valuable in science, where effective teaching depends on identifying and addressing pupils' misconceptions (Ofsted, 2021) and given teachers' comparatively low confidence in being able to assess pupils' progress in science compared to other curriculum subjects (Wellcome Trust, 2017).

[1] Science sampling tests were biennial assessments administered to a representative sample of approximately 9,500 pupils across 1,900 schools to monitor national standards over time. This testing programme has since been discontinued.

Implementation of high-quality, systematic formative assessments can, however, be challenging for teachers, as this requires both subject expertise and time for lesson preparation and delivery. It may be particularly challenging in science, where teachers are required to assess both conceptual understanding and enquiry skills while lessons are ongoing (Hondrich *et al.,* 2016). Tailored training and professional development opportunities have been shown to help teachers overcome these challenges (Estyn, 2022).

Evidence for the Focus4TAPS programme

An efficacy randomised controlled trial of the Focus4TAPS programme was conducted by a research team from University College London (UCL) between 2018 and 2021. This was co-funded by the Wellcome Trust as part of the Education Endowment Foundation (EEF) themed round on improving science education. The evaluation involved 121 schools and 2,822 pupils in England. It tested whether the intervention worked under developer-led conditions.

The efficacy trial found that the programme ran as expected, and that higher science test scores among pupils were associated with greater compliance of teachers with the programme requirements (such as attending more training sessions and conducting more lessons). Teachers in Focus4TAPS schools reported higher confidence in some aspects of teaching and assessing science than their peers in other schools and reported a range of benefits resulting from the programme.

On average, Year 5 pupils in the intervention group made the equivalent of two additional months' improvement in science attainment compared to pupils in other schools (Mutjaba *et al.,* 2022). These results were awarded a 4-padlock high security score under the EEF's 'Classification of the Security of Findings' (EEF, 2019). The evaluation also found a similar positive effect for children eligible for free school meals (FSM), though this was based on smaller numbers of pupils and may therefore need to be interpreted with caution (Mutjaba *et al.,* 2022). No effect was evident for the efficacy trial's secondary outcome, which assessed pupil attitudes towards science: attitudes and beliefs towards science were generally positive and were similar across both Focus4TAPS and control group schools.

The efficacy trial demonstrated that Focus4TAPS could have a positive impact on pupils' science attainment when delivered under controlled conditions. As such, development to support and evaluate delivery at scale was recommended.

Changes to the delivery model since the efficacy trial

Train-the-trainer

The most significant change from the efficacy trial was adaptation of the programme's training model to enable delivery at scale. Bath Spa University adapted the programme to a 'train-the-trainer' model and appointed a cohort of trainers who were trained on how to deliver the teacher training in the pre-trial period. To support trainers' preparation for delivery in the intervention period, and pilot processes to gather data on teachers' views and experiences of the training, Bath Spa University also carried out a pre-trial training practice during the 2023/2024 academic year.

- **'Train-the-trainer'.** Trainers appointed by Bath Spa University attended a four day, in-person Continuing Professional Development (CPD) course delivered by the programme developer in early 2023. Their training focused on the content of the teacher training workshops, including taught sessions and practical activities, with group discussions to address questions and share ideas and best practice.
- **Pre-trial practice.** Bath Spa University ran a one-off **pre-trial training practice** designed to support trainer preparation. Trainers piloted delivering Focus4TAPS training to 100 schools (two teachers per school) that would not be involved in the effectiveness trial. Training feedback questionnaires were also piloted at this stage to inform the content and process to capture evaluation data.

During the intervention period, the Bath Spa University programme developer provided resources and support to trainers to facilitate their practice, minimise burden, and standardise delivery across the regions. These included:

- workshop materials such as presentation slides and speaker notes, handouts, kit boxes, and a kit list;

- templates for correspondence with training groups;
- trainer guidance via regular email—including reminders about when and how to communicate with training groups;
- drop-in trainer webinars before each teacher training workshop, to answer any questions, confirm logistical arrangements, and refresh trainers' knowledge; and
- post-workshop feedback meetings for trainers to reflect on delivery, share tips and tricks with each other, and discuss any improvements that could be made.

Teacher training

In the efficacy trial, teacher training was delivered by the programme developer. Due to Covid-19 restrictions, it took place remotely in six online sessions, each lasting 60 minutes. Training sessions did not include demonstrations or opportunities to try out practical exercises. For the effectiveness trial, by contrast, teacher training comprised three full-day in-person workshops.

Teacher training workshops were delivered across 13 regions, by pairs of trainers in each region. A total of 25 trainers delivered training; all were appointed during the pre-trial phase. Some had delivered training in the pre-trial practice. Others reported that they had either attended as a training participant and seen how teacher training workshops were facilitated from that perspective, or that they had worked with their training partner to fill any gaps in their understanding.

The training was structured as three separate workshops interspersed with classroom delivery. This staged approach was intended to support teachers' development over time, by providing opportunities to trial and reflect on their application of Focus4TAPS in their classrooms, as well as revisiting, reinforcing, and building on key concepts from earlier training sessions each time. (This structure contrasts with other CPD models like Primary Science Quality Mark [PSQM] and Ogden Trust where theory and practice are more clearly separated.)

The content of the Focus4TAPS training workshops remained largely consistent between efficacy and effectiveness trials. Small changes were made to ensure discussions highlighted the programme's alignment with the most recent guidance for primary science teaching practice, however. This included, for example, government guidance such as the EEF's 'Improving Primary Science: Guidance Report' (Luxton and Pritchard, 2023).[2]

Intervention

This section presents a description of Focus4TAPS using the Template for Intervention Description and Replication (TIDieR) checklist.

Who: Implements

Professor Sarah Earle and Bath Spa University developed Focus4TAPS and supported this evaluation as the delivery partner. Following UCL's efficacy trial, Dr Earle led programme development work to implement a 'train-the-trainer' model for delivery at scale; in the 2023/2024 academic year prior to this evaluation, Bath Spa University carried out trainer recruitment, training, and piloting.

As the delivery partner for this evaluation, Bath Spa University held responsibility for recruiting school settings and supporting trainers, schools, and teachers with ongoing delivery during the intervention period.

The programme's trainers were 25 individuals appointed and trained by the programme developer as part of the programme development phase prior to the evaluation. This differed from the efficacy trial, where teacher training was provided by the programme developer directly.

Who: Participants

Focus4TAPS participants included schools, science leads, Year 5 teachers, and Year 5 pupils in classes whose teachers attended the training. Teacher training was delivered by the trainers appointed and trained by Bath Spa University. They

[2] See: **https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/primary-science-ks1-ks2**

delivered training to participating schoolteachers and science leads, who themselves delivered Focus4TAPS science lessons to their Year 5 pupils. Schools were encouraged to use the programme approach in other classes, but only one Year 5 class was included in the evaluation.

Participating schools were located in England; had pupils on roll who would be entering Year 5 during the 2024/2025 school year; and had not taken part in other recent primary science trials or the EEF Focus4TAPS subsidised programme.

To participate, teachers/science leads had to teach Year 5 and/or be a science lead, and be able to attend training in: Newcastle; Sunderland; Sheffield; Liverpool; Bury; Nottingham; Derby; Wolverhampton; Suffolk; Essex; Southampton; South-east London; or West London.

Pupil participation in the intervention was based on their being part of a single class of Year 5 pupils taught by these science teachers. All pupils in each schools' nominated class were eligible to take part in Focus4TAPS lessons, which were intended to be delivered as a whole-class intervention.

What: Materials and procedures

Recruitment of schools, teachers, and pupils

Bath Spa University was responsible for recruiting schools to the effectiveness trial. Bath Spa University invited schools to take part, collected completed Memorandum of Understanding (MoU) from those who opted in to the trial, and shared sample information securely with the research team. The research team held responsibility for randomising schools to either the intervention or control group for the evaluation.

Bath Spa University asked schools in the intervention group to nominate one Year 5 science teacher and one science lead to attend the Focus4TAPS training. In instances where both roles were held by a single member of staff, schools had the option of sending just that person to the training or inviting a second colleague to attend alongside them. Science leads of participating schools were invited to take part in the programme with the aim of promoting a whole-school approach to teaching science and facilitating in-school support for participating teachers.

Staff in participating schools identified their eligible Year 5 class and sent a letter (provided in Appendix L in Further Appendices) to their parents/carers inviting them to take part. These letters contained information about the programme delivery and the evaluation. Parents/carers were given the opportunity to opt their child out of the evaluation by letting their class teacher know that they did not give their permission for their child to participate in the research. This was done using an 'opt-out' form provided in the information letter.

Teacher training and support

The Focus4TAPS programme took place over the course of the 2024/2025 academic year. The teacher training comprised three full-day professional development workshops (one per school term). Each workshop was delivered by a pair of Focus4TAPS trainers.

In the first workshop, trainers introduced teachers to key elements of the Focus4TAPS approach, including focusing lessons on a single element of Working Scientifically, focused recording, and formative assessment.[3] Trainers modelled how these elements could be implemented in classrooms. They provided teachers with opportunities to reflect on their current practice. This content aligned with the second layer of the 'TAPS pyramid model', which focuses on 'responsive teaching'.

In the second workshop, trainers focused on supporting teachers to identify opportunities for pupils to take a more active role in practical science lessons. This included self-assessment and peer assessment, and pupils using scientific enquiry. Trainers also guided teachers to review examples of pupil work and to reflect on when and how to use the five types of

[3] 'Focused recording' refers to a streamlined approach to documenting information in science lessons. Instead of documenting every step of a full scientific enquiry, teachers select one specific element of the investigation (e.g. planning, measuring, concluding) for the pupils to record.

enquiry: comparative testing; pattern seeking; grouping and classifying; observing over time; and research. This reflected the first layer of the TAPS pyramid model, which describes what 'active pupils' look like in practice.

In the final workshop, trainers built on content from earlier sessions to introduce approaches to developing a shared understanding of formative assessment and pupil progression. Trainers linked the Focused Assessment approach to the National Curriculum and existing primary science pedagogy. Trainers also gave teachers time to discuss their plans for continuing with Focus4TAPS in their schools. This workshop aligned with the third layer of the TAPS pyramid model, which focuses on building 'shared understanding'.

Each training session was delivered by two trainers, in person, to teachers who were grouped by region. This was different to the efficacy trial, where training was delivered by the programme developer. Where teachers were unable to attend a live session, they could also view a summary video to catch up.

Additional sources of support for teachers to deliver the programme comprised: email contact with their trainers between workshops; and in-school support via discussions with colleagues, especially their science lead who also attended training.

Delivery of Focus4TAPS lessons

Teachers were expected to deliver a minimum of five lessons using the Focus4TAPS approach to teaching science, focusing on a different element of Working Scientifically and using Focused Assessment in each lesson. The aim of the intervention was for the strategies to be integrated into the broader science curriculum, with a clear focus for teaching, learning, and assessment developed in all science lessons.

The Focus4TAPS programme provided teachers with example lesson plans and activities for use in their classrooms. Teachers could select a minimum of five TAPS plans to carry out during the year from an online bank of 40+ lesson plans on the PSTT website. Additional resources required for delivery of these lessons comprised of standard school materials (e.g. pencils or coloured paper).

The Focus4TAPS lesson resources, including the lesson plans and examples, intended to support delivery across a range of contexts. Teachers were expected to tailor their lessons to the specific needs of pupils in their classes and could choose from the lesson plans accordingly.

How: Mode of delivery

Focus4TAPS lessons were to be delivered in person.

Teacher training was intended to be delivered in person. Training resources were also made available online to allow teachers to catch up on missed sessions and/or refer to them to refresh their memory.

This differed from the efficacy trial, where teacher training was provided online due to Covid-19 disruptions.

How much and when: Dosage

Teachers were required to attend three training workshops, one in each term of the 2024/2025 academic year. The workshops were held in September/October 2024, January/February 2025, and April/May 2025.

Teachers were expected to deliver a minimum of five Focused Assessment lessons between October 2024 and June 2025. Trainers also encouraged teachers to apply the Focus4TAPS principles to other science lessons.

The intervention was not intended to increase or otherwise alter the amount of science learning delivered to participating classes, unless the school was currently teaching less than the recommended two hours per week.

Where: Location

Teacher training took place in 13 regions across England: Newcastle; Sunderland; Sheffield; Liverpool; Bury; Nottingham; Derby; Wolverhampton; Suffolk; Essex; Southampton; South-east London; and West London.

Delivery of Focus4TAPS lessons took place in the participating schools in England. Lessons were to be held at the participating school in the Year 5 science class taught by teachers who attended the training. Developers specified that the lessons should take place in their usual classroom.

Tailoring

Teacher training was expected to be consistent across regions, in that trainers used the same core materials (slides, resources, and links) in line with their pre-trial training. Trainers were, however, given flexibility as to the specific examples they used.

The number of staff who attended training from each school could vary from one to two, depending on whether science lead and Year 5 teacher roles were held by the same individual in their school.

The Focus4TAPS training programme had two core components that schools were expected to follow:

- Schools were expected to attend all three training workshops. These were expected to be full-day sessions and take place each term of the academic year. However, schools were permitted to catch-up online if unable to attend a session in person.
- The same school staff were expected to attend each workshop.

Delivery of Focus4TAPS lessons in schools was expected to be consistent in four key ways:

- At least five lessons delivered over the course of the academic year.
- Lessons used Focused Assessment and focused on a specific element of Working Scientifically.
- Delivered by the same teacher who attended Focus4TAPS training to the nominated Year 5 class.
- Delivered as a whole-class intervention.

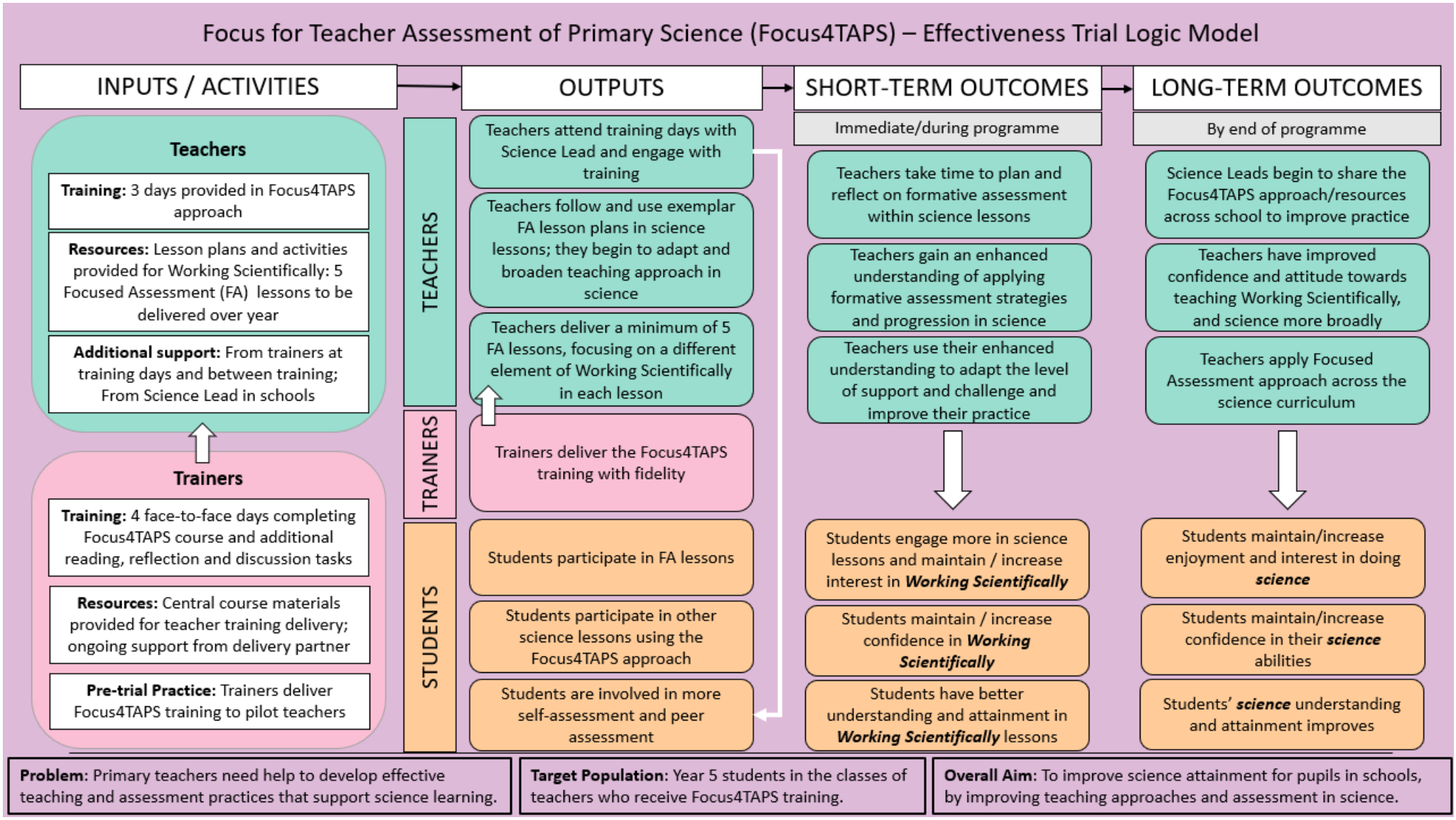
The delivery model encouraged tailoring in the following ways:

- Teachers were expected to tailor their teaching to the needs and relative progression of the pupils in their class—a fundamental aspect of the formative assessment and learning approach. While teachers were expected to follow lesson plans and examples closely in the first instance, they could select from a wide range of lesson plans and activity examples. The programme expectation was that they would adapt and broaden their approach as they built familiarity and confidence through the programme.
- Duration of science lessons was not required to be consistent across schools. A range of lesson lengths were reported in the baseline data on schools' standard practice. The programme was not expected to influence any change in how much schools taught science or how long lessons were, unless schools were delivering less than the government-recommended two hours of science on a weekly basis.
- Where schools worked to an existing scheme of work, they were encouraged to adapt Focused Assessment lesson plans to align with their existing structures.

Logic model

The logic model for this intervention is shown in Figure 3. It was developed by the National Centre for Social Research (NatCen) research team in collaboration with Bath Spa University. This built on the previous model for the efficacy trial, which was reviewed and revised in December 2023 in the EEF-led workshop.

Figure 3: Focus4TAPS logic model



Evaluation objectives

This evaluation of Focus4TAPS is designed as a two-arm cluster randomised effectiveness trial, with schools as the unit of randomisation and pupils as the unit of analysis. Our integrated evaluation design included both an impact evaluation and an implementation and process evaluation (IPE). The evaluation protocol and Statistical Analysis Plan (SAP) were published on the **EEF's website** (Leonard *et al.,* 2024; Leonard *et al.,* 2025).

The primary research question (RQ) for the impact evaluation was:

RQ1: What is the impact of the Focus4TAPS programme on the science attainment of Year 5 pupils (as measured by a science assessment), compared with usual practice?

It also aimed to answer five secondary RQs.

RQ2: What is the impact of the Focus4TAPS programme on Year 5 pupils' attitudes towards science compared to usual practice?

RQ3: Is any impact (RQ1 and RQ2) moderated by eligibility for FSM?

RQ4: Is any impact (RQ1 and RQ2) moderated by any of the following characteristics? (a) gender, (b) exposure to science, (c) Key Stage 1 maths category, and (d) pupils' pre-intervention interest in science.

RQ5: What is the Complier Average Causal Effect (CACE) on the science attainment of Year 5 pupils?

RQ6: What is the per pupil cost of Focus4TAPS?

The IPE aimed to answer the following broad RQs:

IPE RQ1: How is the Focus4TAPS programme delivered, and what are the key facilitators and barriers for delivery?

a. What is 'Teaching as Usual' and how does Focus4TAPS differ?

b. Was Focus4TAPS implemented by teachers as intended?

    i. Did teachers focus on one element of Working Scientifically in each lesson?

    ii. Did teachers plan and reflect on formative assessment within science lessons?

    iii. To what extent were teachers able to adapt the level of support and challenge in lessons based on their formative assessment?

    iv. To what extent were students involved in more self-assessment and peer assessment compared to teaching-as-usual?

c. Was Focus4TAPS training implemented as intended by trainers?

d. What were the drivers and obstacles to implementation of the intervention/training, and was this affected by region/school context?

e. How well was the Focus4TAPS developer able to monitor fidelity and quality of the training programme?

f. How scalable is the Focus4TAPS developer's approach to quality assurance of the training programme?

g. How many Focused Assessment lessons were delivered in intervention schools?

h. What was the amount of time spent delivering the Focused Assessment lessons?

IPE RQ2: How engaged are teachers and pupils in Focus4TAPS, and how is engagement affected by context, teacher characteristics, and pupil characteristics?

a. Did the intervention reach its intended population?

b. Are there patterns in terms of pupil characteristics for those who were/were not reached by the intervention?

c. How engaged were teachers with the training programme and how useful did they find it?

d. How did teachers feel about the intervention lesson resources?

e. What were the perceived costs and benefits of being involved in the intervention?

f. What were the drivers/challenges to teachers' engagement with the training and the intervention?

g. How engaged were pupils in the intervention Focused Assessment lessons and did this differ from their engagement in other science lessons?

h. How did pupils feel about science lessons, and Working Scientifically in particular?

i. What were the drivers/challenges to pupils' engagement?

j. Did patterns of engagement and drivers/challenges differ by pupil characteristic, i.e., by pupil disadvantage?

IPE RQ3: What are the longer-term implications of Focus4TAPS for teaching and learning science?

a. Could teachers in the intervention group continue to implement Focus4TAPS without any further support or training?

b. What are the drivers and obstacles to scaling up the training model to increase the number of trainers and teachers delivering Focus4TAPS?

c. What are the perceived effects of Focus4TAPS on pupils' engagement with learning / school more broadly, and does this differ by pupil characteristics?

d. What are the perceived effects of Focus4TAPS on schools and teachers?

    i. To what extent have teachers gained an enhanced understanding of formative assessment in science?

    ii. To what extent have teachers improved confidence and attitudes towards Working Scientifically and science more broadly?

    iii. To what extent have teachers begun to apply the Focused Assessment approach across the science curriculum?

e. How do Science Leads in schools intend to support other teachers in their schools to implement the approach?

The key implementation dimensions assessed in the IPE were: implementation fidelity and quality (including dosage); reach; teachers' responsiveness; pupils' responsiveness; differentiation; and wider outcomes.

Ethics and trial registration

NatCen's Research Ethics Committee (REC) reviewed and approved the proposed research design for this project on 11 January 2024. The REC process provides ethical guidance and oversight, ensuring that all NatCen research meets the ethical standards required by government and other funders. NatCen's ethics procedure is aligned with the 2021 Government Social Research (GSR) professional guidance and the 2021 Social Research Association (SRA) ethics guidance.

The trial was **registered** in an Open Science Framework registry of the Centre for Open Science in June 2024 (registration number: an3yq).

School recruitment

Bath Spa University recruited 249 eligible primary schools between February and May 2024. NatCen drafted the recruitment documents for the evaluation, including a school information sheet, an MoU, a parent/carer information sheet, a dedicated **study web page**, and a **privacy notice**.

Recruitment materials included a school information sheet, which set out a brief description of the Focus4TAPS programme, what taking part in the programme and evaluation involved, the benefits of taking part, and eligibility criteria. The MoU explained in more detail the expectations and responsibilities for schools in the intervention and teaching as usual groups. By signing the MoU, schools formally agreed to take part in the evaluation and to the conditions associated with taking part.

Once the schools had signed the MoU, teachers handed out information sheets to the parents/carers of all pupils who would be entering Year 5 in the 2024/2025 academic year at the school. The parent/carer information sheets provided information about the research's purpose and explained that their child's school was going to securely share information about their child with NatCen. This information included the child's first name, surname, gender, date of birth, Unique Pupil Number (UPN), eligibility for FSM, pupil's class, and pupil's teacher. The school also provided school-level average Key Stage 1 maths and reading scores from the 2021/2022 academic year. The information sheet also informed parents/carers of their right to opt their child out of the evaluation, and that doing so meant their children would not take part in any assessments, surveys, or interviews, nor have their information shared. Schools were asked to allow two weeks for parents/carers to opt-out their child from participating in the evaluation activities before sharing any pupil information.

Data protection

For the duration of the evaluation, NatCen was the data controller, responsible for deciding the purpose and legal basis for processing data. NatCen stored and handled all data securely and confidentially in line with the **UK General Data Protection Regulation** (GDPR; GDPR, 2016). School- and pupil-level information was collected and processed on the 'legitimate interest' basis.[4] A legitimate interest assessment (LIA) was conducted for this project. The lawful basis of legitimate interest was selected because the processing of personal data was necessary to conduct an independent evaluation of the Focus4TAPS programme. NatCen's legitimate interests included generating robust evidence that contributes to the evidence base on improving primary science education. No special category data was collected as part of the evaluation.

Consent was obtained on an opt-out basis, namely, participants had an option to withdraw themselves or their child from the study. Parents/carers and staff received a hardcopy or e-version of the relevant privacy notice, and the same was available on the study website. Participants taking part in IPE activities (interviews, observations, surveys) were asked to 'opt in' and were reminded that participation was voluntary.

[4] Under the UK GDPR, legitimate interest allows personal data to be processed for a clearly defined research purpose where the processing is necessary, proportionate, and does not override the rights and freedoms of the individuals concerned (GDPR, 2016).

Bath Spa University was a data processor during recruitment, and **AlphaPlus** Consultancy was a data processor for the pupil assessments—their roles are detailed further in the 'Methods' section below.

NatCen received personal data from the delivery team at Bath Spa University, which included names and contact details for staff working at the delivery schools and contact information for the schools. Data sharing between NatCen and Bath Spa University was governed by a data sharing agreement (DSA). NatCen also received personal data relating to the children directly from settings through a secure website.

NatCen stored and handled all data securely and confidentially in line with UK GDPR (GDPR, 2016). Only named individuals at NatCen had access to data collected as part of the evaluation, monitored through a data security plan set-up by the principal investigator. In addition, some third parties (e.g. transcribers, printers, and consultants working for AlphaPlus) had limited access to personal data. NatCen's privacy notice provided further information on who had access to data collected as part of the evaluation.

After the evaluation is completed, data from the impact evaluation will be stored in the EEF archive. At that point, the EEF will become the data controller for the archived trial data. In the future, researchers might link the data deposited in the EEF's archive to information held in the Department for Education's National Pupil Database (NPD) to undertake further research. For further information, see the **EEF's archive privacy notice**.

Project team

The Focus4TAPS training programme was delivered by Bath Spa University. Table 3 lists the delivery team at Bath Spa University.

Table 3: Delivery team

Team member	Institution	Role
Professor Sarah Earle	Bath Spa University	Delivery partner project lead
Focus4TAPS trainers	Range of institutions including independent consultants, school-based, university-based, and science charity organisations	Recruitment and leading the three-day training

A team of education and evaluation specialists at NatCen carried out the evaluation (Table 4).

Table 4: Evaluation team

Name	Job title	Project role
Sarah Sharrock	Research director, Centre for Children and Families	Principal investigator and IPE lead
Dr Liza Benny	Research director, Centre for Evaluation	Impact evaluation lead until April 2026
Dr Enes Duysak	Research director, Centre for Evaluation	Impact evaluation lead
Sehaj Bhatti	Senior researcher, Centre for Children and Families	IPE support and day-to-day project manager
Rebecca Parker	Senior researcher, Centre for Evaluation	Impact evaluation support
Joe Crowley	Senior researcher, Analytics	Impact evaluation support
Kate Wadsworth	Researcher, Centre for Children and Families	IPE fieldwork and support
Kate Baird	Researcher, Centre for Children and Families	IPE fieldwork and support
Alina Haque	Researcher, Centre for Children and Families	IPE fieldwork and support
Dr Gayle Munro	Director, Centre for Children and Families	IPE quality assurance

Name	Job title	Project role
Dr Terry Ng-Knight	Research director, Centre for Evaluation	Impact evaluation quality assurance
Dr Hayley Leonard	Research director, Centre for Children and Families	Principal investigator until April 2025
Dr Georges Poquillon	Research director, Centre for Evaluation	Impact evaluation lead until February 2025

Methods

Trial design

The impact evaluation was designed as a two-arm cluster randomised effectiveness trial of the effect of Focus4TAPS on Year 5 science attainment. The unit of randomisation was at the school level, and unit of analysis was at the pupil level.

Schools were randomly assigned to either the Focus4TAPS intervention group or the teaching as usual group. In the Focus4TAPS intervention group, Year 5 pupils received the Focus4TAPS programme during their science lessons, while those in the teaching as usual group continued with their standard science teaching approach.

Schools assigned to the Focus4TAPS intervention group were offered a three-day training programme during the 2024/2025 academic year at a reduced fee of £150. To incentivise participation and completion of all trial requirements, Focus4TAPS schools received £200, and teaching as usual schools received £500 upon trial completion. Additionally, teaching as usual schools were offered access to the Focus4TAPS programme at a significantly reduced fee after the trial ended.

The primary and secondary outcome measures in this trial were the same measures used in the Focus4TAPS efficacy trial. The primary outcome measure for the trial was science attainment among Year 5 pupils, assessed using a science attainment test developed by the University of York (Joshi *et al.,* 2021; provided in Appendix H in Further Appendices). This test comprises 15 questions with 38 individual items, covering a range of subject areas, challenge levels, and includes the Working Scientifically subscale. Secondary outcomes included pupils' attitudes and beliefs towards science, measured using a questionnaire developed for the efficacy trial (see Appendix I in Further Appendices). This questionnaire consists of seven subscales: interest and enjoyment in science; confidence in science; perceptions of science teachers' teaching; self-regulation of learning in science; self-efficacy for Working Scientifically; Working Scientifically beliefs; and wider benefits of science.

The baseline measure for the primary outcome analysis was Key Stage 1 reading and mathematics classifications, obtained from the NPD. These measures were selected due to their standardisation across schools and availability for Year 5 pupils at the start of the Focus4TAPS intervention. Baseline data for the secondary outcome measure were collected using the attitudes and beliefs towards science questionnaire, which was administered to pupils at the beginning of the trial and again alongside the science attainment assessment at endline. Further details on the outcome and baseline measures are provided in the 'Outcome measures' section. Table 5 summarises the overall trial design.

Table 5: Trial design

Trial design, including number of arms		Two-arm cluster randomised effectiveness trial
Unit of randomisation		School
Stratification variable (s) (if applicable)		Geographical region of the intervention training delivery School level prior Key Stage 1 reading and maths attainment
Primary outcome	Variable	Science attainment
	Measure (instrument, scale, source)	Mark achieved on a science attainment test developed by University of York (Joshi *et al.,* 2021) for the EEF with a continuous single scale (0–45)
Secondary outcome(s)	Variable(s)	Attitudes to science

	Measure(s) (instrument, scale, source)	Questionnaire developed by Focus4TAPS efficacy trial evaluation team, and used in efficacy trial, assessing: • interest/enjoyment in science • confidence in science • perceptions of science teachers' teaching • self-efficacy within Working Scientifically • wider views concerning Working Scientifically • wider views concerning science Scores for each factor are scaled from 1 (Disagree a lot) to 4 (Agree a lot)
Baseline for primary outcome	Variable(s)	(1) Key Stage 1 reading attainment (2) Key stage 1 maths attainment
	Measure (instrument, scale, source)	(1) Key Stage 1 reading classification: KS1_READ_OUTCOME (2) Key Stage 1 maths classification: KS1_MATH_OUTCOME (NPD)
Baseline for secondary outcome(s)	Variable	Attitudes and beliefs towards science
	Measure (instrument, scale, source)	Baseline pupil survey data

There were important changes to the trial design compared with the earlier efficacy trial. In the present trial, delivery was adapted to a train-the-trainers approach, introducing an intermediary between the programme developer and teachers, whereas the efficacy trial used a direct delivery model in which programme developers trained teachers directly. In addition, delivery was scaled up to a larger sample of primary school pupils across a wider range of regions in England.

Participant selection

Bath Spa University were responsible for recruiting schools to take part in the evaluation. As part of the recruitment campaign, Bath Spa University approached schools directly and disseminated information via multiple channels, including holding information events for practitioners in each local authority. In total, 260 schools were recruited across the following regions: Newcastle; Sunderland; Sheffield; Liverpool; Bury; Nottingham; Derby; Wolverhampton; Suffolk; Essex; Southampton; South-east London; and West London. Schools were asked to sign an MoU confirming their commitment to delivering the programme as required and taking part in evaluation activities.

Schools were recruited conditional on signing the MoU and meeting the following eligibility criteria:

- the school had children in Year 4 during the 2023/2024 school year who would then be in Year 5 during the 2024/2025 school year; and
- schoolteachers were able to attend training in one of the regions included in the trial.

Schools were not eligible for the trial if any of the following applied:

- the school had taken part as an intervention school in any other recent primary science trial such as the EEF 'Thinking Doing Talking Science' trial;
- the school was currently involved in the EEF-funded 'PSQM' trial or the EEF Focus4TAPS subsidised programme.

Recruited schools designated a Year 5 teacher who would attend the training (or, the nominated Year 5 teacher for teaching as usual schools) and identified a Year 4 class that would be in Year 5 during the 2024/2025 academic year for participation

in the trial. All pupils in the nominated class were eligible to take part in the evaluation and no pre-tests or scores was used as an eligibility criterion.

Teachers distributed information sheets to the parents/carers of all Year 4 pupils in the nominated class. These explained that the school would securely share information about their child with NatCen and included withdrawal forms through which parents/carers could request that their child not take part in the evaluation. This opt-out process was completed before the start of the 2024/2025 academic year. The information sheets provided to schools and parents/carers are available in the trial protocol (Leonard *et al.,* 2024) on the EEF's website.

Between March and June 2024, schools provided pupil information (UPN, forenames, surname, date of birth, and gender) for the pupils who had not opted out of the evaluation. This information was collected prior to randomisation to minimise the risk of school-level attrition post-randomisation.

In autumn 2024, schools were requested to update their Year 5 pupil lists to account for changes in pupil composition from the previous academic year, including any new pupils that may have joined in September 2024. Pupil information was securely transferred to NatCen through a secure upload platform, which has been used in previous trials funded by the EEF. These pupil lists were used to link to the NPD to obtain baseline Key Stage 1 attainment data, as well as information on pupils' eligibility for FSM.

Outcome measures

Baseline measures

The baseline measures used for the primary outcome analysis were pupils' Key Stage 1 maths and reading attainment classifications, used as indicators of prior attainment, as science attainment was not assessed at baseline. These measures were chosen because Key Stage 1 was the most recent standardised examination completed by Year 5 pupils prior to the start of the Focus4TAPS intervention. The measure is standardised across schools, and evaluators are encouraged to source measures of prior attainment from the NPD (EEF, 2022).

Key Stage 1 attainment data were obtained from the NPD and were categorical with the following categories:

- A = Absent.
- D = Disapplied.
- BLW = Below.
- PKF = Pre-key stage standards.
- WTS = Working towards the expected standard.
- EXS = Working at the expected standard.
- GDS = Working at a greater depth within the expected standard.

To control for prior Key Stage 1 attainment in the primary analysis, binary variables were created for Key Stage 1 maths and reading attainment. The categories 'EXS' and 'GDS' were combined and coded as one, indicating that the pupil was working at the expected standard or above, with all other categories coded as zero. The categories 'A' and 'D' were classed as missing, in alignment with the approach used in the efficacy trial, as this reflected prior attainment being unavailable.

In the efficacy trial, the correlation coefficient between Key Stage 1 measures and science attainment (the primary outcome) was 0.529. In the present trial, the observed correlation between Key Stage 1 measures and science attainment was 0.585 (see the 'Sample size' section below for further details). Based on findings from the efficacy trial, it was anticipated that baseline scores would explain around 28% of variance in science attainment at endline, as this was the percentage of the variance that Key Stage 1 measures explained in science attainment in the efficacy trial.

The secondary outcome analysis included baseline measures for each of the beliefs and attitudes to science subscales. Each subscale was treated as a separate outcome and analysed individually. This data was collected at baseline (completed between 11 September and 14 October 2024).

Primary outcome

The primary outcome for this evaluation was science attainment among Year 5 pupils, measured by a science attainment test developed by the University of York (Joshi *et al.,* 2021, updated 2022). The test measured pupils' knowledge and skills in biology, chemistry, physics, and Working Scientifically, reflecting both subject-specific content and scientific inquiry skills. The logic model for Focus4TAPS hypothesised that ensuring a shared understanding between teachers and pupils of learning aims and success criteria, as well as the teaching of Working Scientifically alongside practical work, would support improved science attainment.

The science attainment test consists of 15 questions with 38 parts, yielding a maximum score of 45 marks. It includes multiple choice, free-text, drawing, and graphical responses, ensuring a comprehensive assessment of pupils' scientific abilities. No scaling or transformation procedures were applied to the raw scores of the science attainment test. The scores were analysed directly as collected, with psychometric properties supporting their use in raw form.

Considering data from 1,510 pupils who completed all items on the science attainment test, the assessment demonstrated acceptable reliability (38 items, Cronbach's alpha = 0.82). The standardised alpha was also 0.82, with Guttman's Lamda 6, G6(smc) = 0.83. The 'alpha if item deleted' values remained close to 0.80, demonstrating that no item substantially affected the overall internal consistency of the test. These results show a comparable level of reliability to the efficacy trial, which found acceptable internal consistency for the science attainment test (Cronbach's alpha = 0.87 across 38 items).

The science attainment test was administered to pupils at endline in June and July 2025 using a paper booklet format, with pupils having 45 minutes to complete it under the supervision of AlphaPlus, an external organisation subcontracted by the NatCen evaluation, who oversaw invigilation to ensure pupils worked independently. Invigilators were blinded to the school's randomly allocated group (Focus4TAPS or teaching as usual) to prevent bias. The completed tests were processed by AlphaPlus and double marked by two independent markers using the provided mark scheme. Any discrepancies between the initial two marks were reviewed by a third independent marker. All data processors and markers were blind to pupil intervention allocation to prevent any biases during the marking process. The mark scheme used in the efficacy trial was retained for this trial and accounted for the pupils potentially selecting multiple options, and for blank responses.

Secondary outcomes

The secondary outcome focused on pupils' attitudes and beliefs towards science. The Focus4TAPS intervention was expected to increase pupils' interest and confidence in science in the long term. Previous research has shown that attitudes towards science had been proven to be an important driver of a pupil's trajectory towards or away from science (Archer *et al.,* 2017; Mujtaba and Reiss, 2014; Mujtaba *et al.,* 2018). Such attitudes influence pupils' engagement and performance in scientific subjects and align with the trial's logic model, which hypothesised that increased interest and confidence in science would contribute to improved educational outcomes.

Pupils' attitudes and beliefs were measured using a questionnaire developed for the efficacy trial. The questionnaire assesses seven subscales of pupils' attitudes towards science:

1. Interest and enjoyment in science.
2. Confidence in science.
3. Perceptions of science teachers' teaching.
4. Self-regulation of learning in science.
5. Self-efficacy for Working Scientifically.
6. Working Scientifically beliefs.
7. Wider benefits of science.

The questionnaire comprises 47 questions across the seven subscales, with responses recorded on a 4-point Likert scale: 1 ('Disagree a lot'); 2 ('Disagree a little'); 3 ('Agree a little'); and 4 ('Agree a lot'). The subscales had demonstrated strong internal consistency in the efficacy trial, with Cronbach's alpha values ranging from 0.70 to 0.90, indicating reliable measurement. The secondary outcomes were each calculated as the average of the relevant questionnaire items where pupils provided answers. Table 6 below summarises the secondary outcome measures alongside the Cronbach's alpha estimates for this trial.

Questionnaires were administered as paper booklets for pupils to complete. The data collection occurred at baseline in September to October 2024 and at endline alongside the science attainment assessment in June and July 2025. The NatCen data team were responsible for collecting and scoring the questionnaire data. To maintain the integrity of the data collection process, the team was blinded to whether pupils had received the Focus4TAPS intervention. This method reduced potential bias and ensured objective scoring of the questionnaire responses.

A total of 5,075 pupils (2,556 Focus4TAPS; 2,519 teaching as usual) completed baseline and 5,352 (2,596 Focus4TAPS; 2,756 teaching as usual) completed endline questionnaires.

Table 6: Secondary outcome measures

Outcome	Questionnaire items	Cronbach's alpha
Interest and enjoyment in science	• I enjoy learning science • I wish I did not have to study science (reversed scores) • Science is boring (reversed scores) • I learn many interesting things in science • I like science • I look forward to learning science in school • Science teaches me how things in the world work • I like to do science experiments • Science is one of my favourite subjects	Baseline: 0.87 Endline: 0.88
Confidence in science	• I usually do well in science • Science is harder for me than for many of my classmates (reversed scores) • I am just not good at science (reversed scores) • I learn things quickly in science • My teacher tells me I am good at science • Science is harder for me than any other subject (reversed scores) • Science makes me confused (reversed scores)	Baseline: 0.78 Endline: 0.79
Perceptions of science teachers' teaching	• I know what my teacher expects me to do in science • My teacher is easy to understand when teaching science • I am interested in what my teacher says about science • My teacher gives me interesting things to do in science • My teacher has clear answers to my questions about science • My teacher is good at explaining science • My teacher lets me show what I have learned in science • My teacher does a variety of things to help us learn science • My teacher tells me how to do better when I make a mistake in science • My teacher listens to what I have to say about science	Baseline: 0.85 Endline: 0.87
Self-regulation of learning in science	• I reappraise my experiences in science so I can learn from them • I try to think about my strengths and weaknesses in science • I think about my actions in science to see whether I can improve them • I think about my past experiences in science to understand new ideas • I try to think about how I can do things better next time in science	Baseline: 0.75 Endline: 0.80
Self-efficacy for Working Scientifically	• I can plan different science investigations	Baseline: 0.81

Outcome	Questionnaire items	Cronbach's alpha
	• I can take measurements using different equipment • I can record and present results in tables, charts, and graphs • I can use results to make predictions for more science investigations • I can write or present explanations of results • I can understand scientific evidence and arguments	Endline: 0.83
Working Scientifically beliefs	• Doing a science investigation is a good way to find out if something is true • Ideas in science sometimes change • Good answers are based on evidence from many different investigations • It is good to try an investigation more than once to make sure of your findings • Sometimes scientists change their minds about what is true in science • The ideas in science books sometimes change	Baseline: 0.77 Endline: 0.80
Wider benefits of science	• Science can help people make things • Science can help the environment • Science can help animals • Science can help people be healthy	Baseline: 0.72 Endline: 0.72

Sample size

At the protocol stage, the trial was powered to detect an effect of the size found in post-hoc analysis of the efficacy trial on the primary outcome (science attainment among Year 5 pupils) and on the secondary outcomes (science attainment for FSM-eligible Year 5 pupils). After accounting for expected attrition at both school and pupil levels, we anticipated a school-level attrition rate of approximately 15%, resulting in 221 schools remaining in the trial. Pupil-level attrition was estimated at around 16%, leading to an average of 21 pupils per class by the end of the trial. These attrition rates were based on experiences from the efficacy trial, which faced disruptions due to Covid-19. Given that similar disruptions were unlikely to occur during the effectiveness trial, these estimates were expected to be conservative.

Our power calculations, informed by the post-hoc analysis of the efficacy trial, assumed a Type I error rate of 0.05 and a Type II error rate of 0.20 (power of 0.80). The intracluster correlation coefficients (ICCs) used in the calculations (0.204 overall, 0.248 FSM) were based on the final analysis of the efficacy trial, where observed clustering effects were higher than initially assumed at protocol and randomisation stage (Mutjaba *et al.,* 2022). The planned recruitment of 260 schools was expected to yield a minimum detectable effect size (MDES) of 0.182 (after attrition) for the primary analysis of science attainment among Year 5 pupils. With five pupils per school assumed to be eligible for FSM (after attrition), the trial was also powered to detect an MDES of 0.227 for the FSM subgroup analysis. Details are provided in the trial protocol (Leonard *et al.,* 2024) under the 'Sample size' section.[5] We conducted this and subsequent power calculations using the PowerUp! tool (Dong and Maynard, 2013).

At the randomisation stage, we updated the power calculations using the actual number of schools (n=240) recruited for the trial. The average cluster size was estimated at 25 pupils per year group per school, with approximately six pupils per class eligible for FSM, based on the pupil upload data available at that time. Other core assumptions were the same as used at the protocol stage, including around ICCs and pre-test/post-test correlations. With these updated sample sizes, the overall MDES at randomisation stage for the primary analysis was 0.173 standard deviation (SD), and 0.212 SD for the FSM subgroup analysis.

At the analysis stage, we again updated power calculations using the final number of schools and pupils included in the final analysis. These calculations used updated correlations between baseline Key Stage 1 reading and maths attainment

[5] Focus4Taps evaluation protocol can be retrieved from: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/focus4taps_-_evaluation_protocol.pdf?v=1767696174

and science attainment, along with updated ICCs estimated using the primary outcome analysis sample and FSM subgroup analysis sample.

Following the approach of the efficacy trial, Key Stage 1 reading and maths categories were transformed into numeric 'working scales' for the purposes of MDES calculations (1–5 scales; 1 = 'BLW'; 2 = 'PKF'; 3 = 'WTS'; 4 = 'EXS'; 5 = 'GDS'). Across all Year 5 pupils with available Key Stage 1 reading and maths attainment and science attainment scores, the observed correlations between Key Stage 1 working-scale attainment and science attainment were moderate ($r = 0.582$, $p<0.001$ for reading and $r = 0.588$, $p<0.001$ for maths). The average of the two correlation coefficients (0.582+0.588 = 1.170/2 = 0.585) was used in the MDES calculations to provide a single summary measure of the association between baseline attainment and science attainment. The same approach was applied for the subsample of pupils eligible for FSM (Key Stage 1 reading: $r = 0.528$, $p<0.001$; Key Stage 1 maths: $r = 0.527$, $p<0.001$; average used for MDES calculation = 0.541). With these updated assumptions, the overall MDES at endline is 0.149 SD for the primary analysis and 0.196 SD for the FSM subgroup analysis sample.

Table 7: MDES at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.182	0.227	0.173	0.212	0.149	0.196
Pre-test/post-test correlations	Level 1 (pupil)	0.529	0.501	0.529	0.501	0.585	0.541
	Level 2 (school)	0	0	0	0	0	0
ICCs	Level 2 (school)	0.204	0.248	0.204	0.248	0.129	0.115
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.80	0.80	0.80	0.80	0.80	0.80
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size[a]		21	5	25	6	19	4
Number of schools	Intervention	111	111	120	120	113	113
	Control	110	110	120	120	115	110
	Total:	221	221	240	240	228	223
Number of pupils	Intervention	2,331	555	3,000	720	2,496	864
	Control	2,310	550	3,000	720	2,646	769
	Total:	4,641	1,105	6,000	1,440	5,142	1,633

[a] *The harmonic mean of the number of pupils per school was used in the power calculations, following PowerUp! guidance (Dong and Maynard, 2013).*

Randomisation

Randomisation was carried out at the school level, stratified by region and by school's prior level of Key Stage 1 attainment (below or above the sample median), using data from the 2021/2022 academic year, when Year 5 pupils in the trial completed their Key Stage 1 assessments. This stratification ensured a balanced distribution of schools in the Focus4TAPS and teaching as usual groups, accounting for regional training delivery and variation in prior attainment, which may proxy for teacher quality at the school level.

By July 2024, 240 schools had been recruited to the trial and signed an MoU. Two statisticians within NatCen's evaluation team randomised schools on 03 July 2024 using Stata version 17 (StataCorp LLC, College Station, Texas, USA) applying a 1:1 allocation ratio between the Focus4TAPS and teaching as usual groups. We used the *randtreat command* with *misfits(global)* to manage cases where school numbers were uneven within strata. The distribution of schools across randomisation strata by trial arm is available in the trial SAP (Leonard *et al.,* 2025).[6]
School identities were removed from the data used by NatCen researchers during randomisation, with meaningful identifiers such as school names removed and linked back only after randomisation was completed. The Focus4TAPS protocol describes the rationale for the randomisation approach, and the trial's SAP (Leonard *et al.,* 2025) provides further details, including the Stata syntax used.

The randomisation process allocated 120 schools to the Focus4TAPS intervention group and 120 schools to the teaching as usual group. NatCen recorded the allocation in Excel and shared it with the delivery team, who then communicated the allocation to schools on 04 July 2024.[7]

Statistical analysis

The impact analysis followed the SAP (Leonard *et al.,* 2025), published on the **EEF website**, and outlined the analysis in line with the EEF statistical guidance (EEF, 2022). The impact evaluation analysis was conducted in the Office for National Statistics (ONS) Secure Research Service (SRS),[8] with analysis undertaken using R version 4.4.0, RStudio 2023.06.1, and Stata version 18.

Primary analysis

The primary outcome analysis addressed the following question:

RQ1: What is the impact of the Focus4TAPS programme on the science attainment of Year 5 pupils (as measured by a science assessment), compared with teaching as usual?

The primary outcome analysis followed an intention-to-treat (ITT) approach to estimate the impact of the Focus4TAPS programme on the science attainment of Year 5 pupils, with schools analysed as per their randomised allocation to either Focus4TAPS or the teaching as usual group. A two-level linear regression model was used with pupils at level 1 and schools at level 2, including school-level random effects to account for within-school clustering effects.

The primary outcome was pupils' science attainment score (marks ranging from 0 to 45) at endline. Pupils' Key Stage 1 maths and reading attainment classifications were included as baseline measures of prior attainment for the primary analysis. In addition, a binary indicator of treatment allocation, the randomisation strata, and age were also included as covariates in the primary outcome model. The model included pupils' age in months to account for the developmental effect

[6] Focus4TAPS evaluation SAP can be retrieved from:
https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/focus4taps_-_statistical_analysis_plan.pdf?v=1767696174
[7] NatCen analysts did not conduct the analysis blind to randomisation; however, they did not have access to pupil or school names and worked only with pseudonymised identifiers.
[8] Work undertaken in the ONS SRS does not imply the endorsement of the ONS or other data owners.

reported in the materials for the attainment test as well as to improve the efficiency of estimation. This implied the following basic model:

$$ScienceAttainment_{ij} = \beta_0 + \beta_1 KS1Maths_{ij} + \beta_2 KS1Reading_{ij} + \beta_3 Intervention_j + Strata_j'\beta_4 + \beta_5 age_{ij} + u_j + e_{ij}$$

where pupils (i) are nested within schools (j). β_0 is an overall intercept, $KS1Maths_{ij}$ and $KS1Reading_{ij}$ are baseline Key Stage 1 maths and reading attainment categories, and the intervention effect is captured by β_3 The $Strata_j'$ variables represent the stratification blocks issued from the stratification criteria used during randomisation,[9] while age_{ij} is age in months. The term u_j is a school-level random effect and e_{ij} is the error term, both assumed to be normally distributed and uncorrelated with covariates included in the model.

The impact of the Focus4TAPS programme is expressed using unadjusted and adjusted differences in means, and a standardised effect size calculated using Hedges' g formula for standardising effect sizes.

Secondary analysis

The secondary outcome analysis addressed the following question:

RQ2: What is the impact of the Focus4TAPS programme on Year 5 pupils' attitudes towards science compared to usual practice?

The secondary outcomes were pupils' attitudes towards science measured by the seven subscales from the attitudes to science questionnaire. An ITT approach was used with separate multilevel linear regression models (with school-level random intercepts and other covariates included in the analysis of the primary outcome) fitted for each subscale score including, as an additional covariate, the corresponding science subscale score at baseline. As outlined in the 'Secondary outcomes' section above, the attitudes to science questionnaire was administered at both baseline and endline. A baseline measure was included in the model to improve statistical power as the baseline and endline subscale scores were expected to be correlated. The basic form of the model for each subscale was as follows:

$$SubscaleScore_{ij,k} = \beta_0 + \beta_1 BaselineSubscaleScore_{ij,k} + \beta_2 Intervention_j + Strata_j'\beta_3 + \beta_4 age_{ij} + u_j + e_{ij}$$

where pupils (i) are nested within schools (j), and k is the specified subscale. β_0 is an overall intercept, $BaselineSubscaleScore_{ij,k}$ is the corresponding subscale score at baseline, and the intervention effect is estimated by β_2. The $Strata_j'$ variables represent the stratification blocks while age_{ij} is age in months. The term u_j is a school-level random effect and e_{ij} is the error term.

Due to multiple secondary outcomes being analysed, p-values for these outcomes were reported before and after Romano-Wolf step-down p-value adjustment, which considered the greater likelihood of Type I errors while allowing for dependence among outcomes in the data (Romano and Wolf, 2005; Romano and Wolf, 2016).

Subgroup analyses

Subgroup analysis was conducted both for primary and secondary outcomes.

RQ3: Is any impact (RQ1 and RQ2) moderated by eligibility for FSM?

RQ4: Is any impact (RQ1 and RQ2) moderated by any of the following characteristics? (a) gender, (b) exposure to science, (c) Key Stage 1 maths category, and (d) pupils' pre-intervention interest in science.

[9] Schools were stratified by geographic region and school level Key Stage 1 maths and reading attainment.

For RQ3, we assessed whether Focus4TAPS had a differential impact on pupils who were eligible for FSM and received the Focus4TAPS programme, compared to pupils who were eligible for FSM and received teaching as usual. To carry out this analysis, FSM eligibility was based on the variable "EVERFSM_6_P" in the NPD, which indicated if a pupil had been recorded as eligible for FSM at any time in the last six years.

Two analysis approaches were used for the FSM subgroup analysis. First, a similar model to that applied for the primary outcome analysis was applied to the restricted subgroup of pupils eligible for FSM. The β_3 coefficient in this model provides the estimated treatment effect specifically for the FSM subgroup. Second, an interaction model with interactions between the binary indicator of intervention allocation and FSM eligibility was run for all pupils in the analysis sample:

$$ScienceAttainment_{ij} = \beta_0 + \beta_1 KS1Maths_{ij} + \beta_2 KS1Reading_{ij} + \beta_3 Intervention_j + \beta_4 FSM_{ij} + \beta_5 Intervention_j * FSM_{ij} + Strata_j'\beta_6 + \beta_7 age_{ij} + u_j + e_{ij}$$

where pupils (i) are nested within schools (j). β_0 is an overall intercept, $KS1Maths_{ij}$ and $KS1Reading_{ij}$ are baseline Key Stage 1 maths and reading attainment categories, and the overall intervention effect is captured by β_3. The β_4 coefficient represent the differential attainment for pupils eligible for FSM and the interaction term (β_5) captures the differential impact of Focus4TAPS for the FSM subgroup. The results from the two approaches are compared as a sensitivity check.

The above analysis was repeated for all secondary outcome subscales to assess whether the impact of Focus4TAPS varied by FSM eligibility. As with the analysis for RQ2, p-values were reported before and after Romano-Wolf step-down p-value adjustments were made to account for multiple hypothesis testing, for all secondary outcomes within each subgroup.

For RQ4, we assessed whether Focus4TAPS had a differential impact on pupils looking at the following characteristics, in alignment with the subgroups explored in the efficacy trial and to capture other characteristics that may be associated with science attainment:

1. Gender (as self-reported by pupils via the survey questionnaire and/or as indicated in the NPD) (0 = Boys; 1 = Girls).
2. Key Stage 1 maths attainment category[10] (from the NPD) (0 = Below expected standard; 1 = Expected standard or above).
3. Pupil pre-intervention interest in science (as reported by pupils in the baseline pupil survey questionnaire) (0 = Low interest; 1 = High interest).
4. Exposure to science, derived from a set of questions asked in the baseline pupil survey questionnaire, namely:
    i. whether pupils' parents/carers attended university;
    ii. whether pupils had a family member who worked as a scientist or in a job using science or medicine;
    iii. pupils thinking that their parents/carers were interested in science;
    iv. pupils having teachers with STEM (science, technology, engineering, and mathematics) backgrounds.

Exposure to science was derived from the above list of survey questions by applying a dimension reduction approach, based on factor analysis where the four variables of interest were combined into a single variable. A binary indicator was then derived from the factor variable, taking the value of 1 if the factor score was above the median (0 = Low exposure; 1 = High exposure).

For each of the subgroups in RQ4, interaction models were run with intervention terms between the intervention and subgroup binary indicators for all outcomes. This implied the following model for each subgroup (example shown for the primary outcome):

[10] For the Key Stage 1 maths category subgroup analysis, baseline Key Stage 1 attainment was not included as a covariate in the first model due to the subgroup restriction already controlling for this variable.

$$ScienceAttainment_{ij} = \beta_0 + \beta_1 KS1Maths_{ij} + \beta_2 KS1Reading_{ij} + \beta_3 Intervention_j + \beta_4 Subgroup_{ij} + \beta_5 Intervention_j * Subgroup_{ij} + Strata'_j \beta_6 + \beta_7 age_{ij} + u_j + e_{ij}$$

where pupils (i) are nested within schools (j). β_0 is an overall intercept, $KS1Maths_{ij}$ and $KS1Reading_{ij}$ are baseline Key Stage 1 maths and reading attainment categories, and the overall intervention effect is captured by β_3. The β_4 coefficient represents the effect of the subgroup binary indicator and the interaction term (β_5) captures the differential impact of Focus4TAPS for the subgroup. The $Strata'_j$ variables represent the stratification blocks, while age_{ij} is age in months. The term u_j is a school-level random effect and e_{ij} is the error term.

Following the approach in the efficacy trial, where a statistically significant interaction effect was revealed in the interaction model analysis, separate models were estimated for the subgroups separately. Effect sizes by subgroup were therefore only reported for subgroup-outcome combinations where the interaction model demonstrates statistically significant effects. Given the high number of estimates generated by the subgroup analysis, adjustments for multiple hypothesis testing were made using the Romano-Wolf adjustment within each subgroup.

Analysis in the presence of non-compliance

A CACE analysis was undertaken to show the impact of the Focus4TAPS programme on the primary outcome (science attainment) compared to individuals receiving teaching as usual, taking into account the level of compliance with Focus4TAPS. This addressed the following question:

RQ5: What is the CACE on the science attainment of Year 5 pupils?

The compliance measure was defined as a count measure (taking the value 0 for a non-compliant school, and values 1 or 2 corresponding to the counts of compliance criteria that the school fulfilled). This therefore allowed for partial compliance to indicate how impact estimates varied by level of compliance.

Where there was imperfect compliance across schools, a CACE analysis was undertaken, drawing on an instrumental variable to estimate the CACE for schools where the intervention allocation was complied with, with standard errors clustered at the school level. Specifically, the CACE estimated the effect of Focus4TAPS on science attainment outcomes for pupils in schools that complied with the intervention allocation.

Compliance in this trial was defined at the school level and two criteria captured compliance to Focus4TAPS:

1. Attendance of teacher training sessions by the Year 5 teacher.
2. Delivery of assessment class lessons.

As shown in Table 8, for a school to be compliant the Year 5 teacher needed to attend at least two out of three sessions of training and deliver a minimum of five assessment class lessons using Focus4TAPS by the end of the academic year.

Table 8: Compliance measure indicators

Number	Criteria	Data source	Indicator
1	Attendance of teacher training sessions	Attendance register	Year 5 teachers attend at least two out of three sessions of training
2	Delivery of lessons	Teacher self-report in endline survey	At least five assessment class lessons using Focus4TAPS delivered by the end of the year

Notes: Missing data on any criterion was scored as zero (that is, the setting was considered non-compliant on that criteria).

Initial group allocation was used as an instrumental variable for the compliance measure in a two-stage least squares (2SLS) estimation approach (Imbens and Angrist, 1994), therefore adjusting the impact estimates by the compliance rate.

The first stage of analysis was compliance regressed on all covariates that were used in the main primary outcome model and included, as an instrumental variable, a binary variable that indicated schools' pre-intervention treatment allocation. The first stage equation was as follows:

$$Comply_j = \beta_0 + \beta_1 KS1Mathsij + \beta_2 KS1Readingij + \beta_3 Intervention_j + Strata'_j\beta'4 + \beta_5 age_{ij} + u_j + e_{ij}$$

The first stage results were reported alongside the correlation between the group allocation and the compliance indicator and the associated F-statistic to indicate instrument relevance.

The second stage of analysis regressed science attainment on the predicted compliance estimated in the first stage. The second stage equation was as follows:

$$ScienceAttainment_{ij} = \beta_0 + \beta_1 KS1Maths_{ij} + \beta_2 KS1Reading_{ij} + Strata'_j\beta_3 + \beta_4 age_{ij} + \beta_5 \widehat{Comply}_j + u_j + e_{ij}$$

The coefficient (β5) is the CACE estimate of the compliance effect. Where there were no confounding factors affecting compliance and attainment, the CACE estimate would be equal to the ITT estimate.
The compliance analysis was estimated only for the primary outcome of science attainment.

Missing data analysis

We followed the protocol for missing data suggested by the EEF statistical guidance (see EEF, 2022). For missingness at a rate of 5% or less from randomisation to final analysis, a complete-case analysis would have been employed. However, the rate of missingness was 21.13%, exceeding the 5% threshold, and therefore an assessment of missing data was carried out.

As a first step, we examined the prevalence and pattern of missingness in outcomes and pre-intervention covariates descriptively with crosstabulations, including counts and percentages in each category. We then explored patterns of missingness using a multilevel logistic regression drop-out model, regressing a binary missingness indicator of the primary outcome (science attainment score) on a set of relevant covariates, including treatment allocation, gender, pupil age in months, FSM eligibility, Key Stage 1 maths and reading attainment, and baseline attitudes and beliefs in science subscale scores, as well as the randomisation strata variables region and school's prior level of Key Stage 1 attainment.

A second drop-out model was estimated to assess patterns of covariate missingness. In this model, the dependent variable was an indicator of whether any of the covariates included in the primary outcome model (Key Stage 1 reading and maths attainment, age in months, and the randomisation strata variables region and school's prior level of Key Stage 1 attainment) was missing. The model included pupil-level covariates: treatment allocation; gender; FSM eligibility; and baseline attitudes and beliefs to science subscale scores.[11]

If the pattern of missingness appeared to be unrelated to treatment allocation or other observed covariates, we continued with a complete-case analysis. If a covariate was considered missing at random (MAR) conditional on other covariates, the primary analysis was re-estimated through multiple imputation using chained equations (MICE) and results were presented alongside the complete data analysis as part of the sensitivity analysis. Multiple imputation was implemented using the *mice* package in R.

Where missing data patterns indicated a likelihood of missing not at random (MNAR), multiple imputation was to be used in addition to sensitivity analyses (Carpenter *et al.,* 2007). The results of the missing data analyses are presented in the 'Missing data analysis' subsection in the 'Impact evaluation results' section below.

[11] In estimating both drop-out model, missing values for continuous covariates were replaced with the sample mean, and missing categories were included for categorical variables.

Additional analyses and robustness checks

We ran imbalance checks to assess whether the analysis sample differed between the Focus4TAPS intervention and teaching as usual groups on baseline pupil- and school-level characteristics. We identified evidence of imbalance for pupil-level FSM eligibility and school-level FSM eligibility, Key Stage 1 maths attainment, and Key Stage 1 reading attainment, based on standardised mean differences exceeding 0.10 or effect sizes exceeding 0.05 (for baseline attitudes to science subscale scores). Where we identified imbalance, a sensitivity analysis was estimated to capture the effect of imbalanced covariates on the primary outcome. This involved re-running the primary outcome analysis model controlling for the covariates that were imbalanced at baseline. The effect size estimates were compared across both models to understand potential changes in the estimated intervention effect.

Additional robustness checks also included estimating White standard errors using the vcovCR() function in the clubSandwich package in R, which adjusted for clustering and heteroskedasticity in the data.

Estimation of effect sizes

We estimated effect sizes for cluster randomised trials, as adapted from Hedges (2007):

$$ES = \frac{(\overline{Y_T} - \overline{Y_C})_{adjusted}}{\sqrt{\sigma_s^2 + \sigma_e^2}}$$

where $(\overline{\boldsymbol{Y_T}} - \overline{\boldsymbol{Y_C}})_{\boldsymbol{adjusted}}$ is the mean difference between the Focus4TAPS and teaching as usual group adjusted for baseline characteristics, while $\sqrt{\boldsymbol{\sigma_s^2} + \boldsymbol{\sigma_e^2}}$ is an estimate of the pooled unconditional population SD. $\boldsymbol{\sigma_s^2}$ is the variance of the school-level intercept and $\boldsymbol{\sigma_e^2}$ is the variance of residuals.

We used each group's adjusted mean and variance to calculate the effect size. This variance was the total variance (across both pupil and school setting, without any covariates, as emerging from a 'null' or 'empty' multilevel model with no predictors). A 95% confidence interval (CI) for the ES, which considered the clustering of pupils in schools, was also reported.

Effect sizes were calculated for each of the models estimated, as specified in the analysis sections above. For the FSM subgroup analysis, ES were calculated using two approaches. First, we calculated the ES using the restricted sample of pupils eligible for FSM, following the same approach as the overall ES calculation above. Second, as a sensitivity check, we calculated ES_{FSM} for FSM pupils using an interaction model, which captured the adjusted treatment effect for FSM pupils:

$$ES_{FSM=} \frac{\beta_1 \mathrm{Treat_j} + \beta_5(\mathrm{Treat_j} * \mathrm{FSMEver_{ij}})}{sd}$$

where β_1 represents the main treatment effect for non-FSM pupils, and $\beta_5$captures the differential effect for FSM pupils. The SD incorporates both within-school ($\boldsymbol{\sigma_e^2}$) and between-school ($\boldsymbol{\sigma_s^2}$) variance to adjust for clustering effects.

Estimation of ICC

ICC measure how strongly pupils from the same school resembled one another in terms of their achieved scores. High ICC estimates indicate that pupils from the same school were likely to score similarly on the science assessment and attitudes to science outcomes.

The ICCs were estimated directly from the primary and secondary outcomes analysis models, using the variance estimates for each level of clustering. The ICC for schools $\boldsymbol{\rho S}$ was estimated using the following formula based on Hedges (2011):

$$\boldsymbol{\rho_S} = \frac{\boldsymbol{\sigma_{BS}^2}}{\boldsymbol{\sigma_{BS}^2} + \boldsymbol{\sigma_{WS}^2}} = \frac{\boldsymbol{\sigma_{BS}^2}}{\boldsymbol{\sigma_{WT}^2}}$$

where σ^2_{BS} is the between-school variance, σ^2_{WS} was the within-school variance and σ^2_{WT} the total variance.

Exploratory additional analyses

We conducted several exploratory analyses in addition to those specified in the SAP (Leonard *et al.,* 2025) to assess the robustness of our findings and the potential influence of baseline imbalances. First, to assess the potential influence of the imbalance in FSM eligibility between groups (see 'Pupil and school characteristics' subsection in the 'Impact evaluation results' section below), we re-estimated the previously specified interaction model that assessed whether Focus4TAPS effects on science attainment differed by FSM eligibility, adding adjustments for baseline pupil- and school-level characteristics that were imbalanced. These included school-level FSM eligibility and school-level average Key Stage 1 attainment. We compared these estimates with those from the restricted subgroup analysis and from the interaction model without baseline adjustment as a sensitivity check. We applied the same baseline imbalance adjustment across all secondary outcome interaction models.

Second, due to the low response rate to the endline teacher survey, we redefined compliance based solely on teacher training attendance (criterion 1) and re-ran the CACE analysis using this alternative definition. We classified schools as compliant if the Year 5 teacher attended at least two of the three training sessions.[12] These results were compared with those from the primary CACE analysis.

Third, we conducted a robustness check by re-running the primary outcome model using an alternative baseline specification. We included Key Stage 1 reading and maths attainment as categorical variables (BLW, PFK, WTS, EXS, and GDS), consistent with the approach used in the efficacy trial. We compared these results with those from the pre-specified primary outcome model.

Finally, we re-estimated the primary outcome model combining the alternative baseline specification with adjustment for baseline imbalances, for both the ITT analysis and the FSM subgroup. These results were compared with those from the pre-specified primary outcome model and FSM subgroup analysis.

IPE

Research methods

NatCen conducted a multi-stage IPE to explore how Focus4TAPS was delivered by trainers and in schools, how delivery aligned with the programme design, and how teachers and pupils experienced the programme. It also assessed the implementation of the 'train-the-trainer' model, in which training was delivered by a team of trainers rather than directly by the programme developer.

The IPE used a mixed-methods design including: pre- and post-intervention surveys with teachers, science leads, and pupils; observations of teacher training sessions and lesson delivery; interviews with the programme developer and trainers; focus groups with teachers and science leads; and desk-based review of programme documentation and attendance data. Using these different methods allowed us to build a triangulated picture of how the programme worked in practice.

Table 9 provides an overview of the IPE methods and links them to key IPE dimensions.

[12] For this analysis, all pupils within a school were assigned a compliance score of 1 if the Year 5 teacher attended at least two training sessions, and a score of 0 otherwise.

Table 9: IPE methods overview

Research method and mode	Sample size	Sampling criteria	Data analysis methods	IPE RQs addressed (including sub-questions)	IPE dimension[a]	Timeline
Developer interview (online)	Encounter: One Sample size per encounter: One programme developer	Not applicable	Qualitative: thematic analysis	1, 3	Implementation Fidelity, Dosage, Differentiation, and Wider Outcomes	June 2024
Document analysis (online)	Not applicable	Not applicable	Qualitative: thematic analysis	1	Implementation Fidelity, Dosage, Differentiation, and Wider Outcomes	July 2024 – August 2024
Teacher surveys (online)	Encounters: two surveys, one at baseline and one at endline	Both intervention and control groups	Quantitative: descriptive statistics	1, 2, 3	Implementation Fidelity, Dosage, Differentiation, Reach, Teachers' Responsiveness, Pupils' Responsiveness, and Wider Outcomes	September 2024 – October 2024; June 2025 – July 2025
Pupil surveys (paper)	Encounters: two surveys, one at baseline and one at endline	Both intervention and control groups	Quantitative: descriptive statistics (covered in impact evaluation)	2	Differentiation, Reach, and Pupils' Responsiveness	September 2024 – October 2024; June 2025 – July 2025
Training observations (in person)	Encounters: three observations (three separate days of training) Sample size per encounter: one pair of trainers	One training session observed per school term, each in a different geographical region	Qualitative: thematic analysis	1, 2	Implementation Fidelity, Dosage, and Teachers' Responsiveness	October 2024; January 2025 – February 2025; May 2025 – June 2025
Training surveys (online)	Encounters: three surveys throughout intervention	Intervention group only	Quantitative: descriptive statistics	1, 2, 3	Implementation Fidelity, Dosage, and Teachers' Responsiveness	October 2024; January 2025 – February 2025; May 2025– June 2025
Trainer interviews (online)	Encounters: 13 Sample size per encounter: one pair of trainers	One pair interviewed per geographical region	Qualitative: thematic analysis	1, 2, 3	Implementation Fidelity, Dosage, Reach, Teachers' Responsiveness, and Wider Outcomes	October 2024 – November 2024; February 2025 – March 2025; May 2025

Research method and mode	Sample size	Sampling criteria	Data analysis methods	IPE RQs addressed (including sub-questions)	IPE dimension[a]	Timeline
Teacher focus groups (online)	Encounters: two focus groups Sample size per encounter: two to five participants including Year 5 teachers, Science leads and Special Educational Needs Coordinators (SENCOs), Science curriculum advisors[b]	Case study schools: maximise differentiation between school characteristic, student population, and participating teacher characteristics	Qualitative: thematic analysis	1, 2, 3	Implementation Fidelity, Dosage, Differentiation, Reach, Teacher's Responsiveness, Pupils' Responsiveness, and Wider Outcomes	October 2024 – July 2025
Pupil focus groups (in person)	Encounters: six focus groups Sample size per encounter: five Year 5 pupils	Intervention group only	Qualitative: thematic analysis	2	Reach and Pupils' Responsiveness	October 2024 – July 2025
Teacher lesson observations (in person)	Encounters: six observations Sample size per encounter: one Year 5 teacher	Intervention group only	Qualitative: thematic analysis	1, 2	Implementation Fidelity, Reach, Teachers' Responsiveness, and Pupils' Responsiveness	October 2024 – July 2025
Teacher lesson reflections (online)	Encounters: one lesson reflection	Intervention group only	Qualitative: thematic analysis	1, 2	Implementation Fidelity, Reach, Teachers' Responsiveness, and Pupils' Responsiveness	January 2025 – February 2025
Science lead survey (online)	Encounters: one survey at endline	Both intervention and control groups	Quantitative: descriptive statistics	1, 3	Differentiation and Wider Outcomes	June 2025 – July 2025
Attendance data	Not applicable	Intervention group only	Quantitative: descriptive statistics	1, 2	Reach, Dosage	May 2025

[a] *IPE dimensions were informed by the EEF IPE guidance.*
[b] *The focus groups were initially only aimed at Year 5 teachers, but the final participants included teachers in other roles at schools.*

The rest of this section describes each IPE activity in turn.

Pre-intervention IPE activities

Interview with the programme developer

We conducted an interview with the programme developer at the beginning of the trial in June 2024. The interview explored the Focus4TAPS programme from the developer's perspective, including the delivery and training model and how this had evolved since the efficacy trial. It examined the 'train-the-trainer' approach, recruitment and quality assurance processes, and the training materials used. The interview also explored perceived facilitators and barriers to training, scaling up, and implementation in schools. Finally, it gathered the developer's views on the expected impacts of Focus4TAPS on teachers, pupils, and schools within the effectiveness trial.

The interview was conducted online and lasted 60 minutes.

Document analysis

We undertook desk-based analysis of key documents relating to the delivery of the programme. The documents included trainer guidance materials and the training materials provided to teachers. This helped us familiarise ourselves with the programme's approach and helped with the development of our research materials.

We organised the information in the documents by themes linked to our RQs. Insights from this analysis were used to create topic guides for interviews and focus groups as well as observation guides for lesson observations.

IPE activities related to the training

Training session observations

We observed three in-person training sessions that were delivered to teachers in the intervention group.

We observed one training session per term, each in a different region. This ensured that the observations captured delivery by a range of different trainers to different teacher groups, while also including the full content of all three training days.

Observations were guided by a detailed protocol template, which set out key dimensions to be captured in each session. These included information on the structure and content of the sessions, and how closely delivery aligned with the intended Focus4TAPS model. The template also recorded participant engagement and understanding, and trainer delivery and use of training resources.

Trainer interviews

We conducted paired interviews with all 25 trainers who delivered training sessions across the 13 regions. Invitations to participate in the interviews were sent to all the trainers via email. The interviews took place online or by telephone and lasted up to 60 minutes each. To capture data relating to each of the three training days without placing undue burden on participant recall, interviews were spread out over three phases, each focusing primarily on views and experiences relating to the most recent training.

The interviews explored trainers' views and experiences of implementation and delivery of training sessions for teachers. This included their views on the structure and content of training workshops; any adaptations trainers made during delivery, including their rationale; how well they felt teachers engaged with the training, including perceived barriers to engagement. The interviews also explored trainers' insights into effects of the programme on teachers, based both on their direct observations during the workshops and what teachers had told them about implementation in classes between training sessions. Insights into perceived impacts on pupils (shared by teachers during the training programme) were also explored. Finally, interviews also sought to gather trainers' perspectives on facilitators and obstacles to scaling up the training model.

Training surveys

We worked with Bath Spa University to develop training feedback survey questionnaires that captured attitudinal data about each training session and the intervention itself. The questionnaire built on the version used by Bath Spa University in previous stages of the programme's development. We reviewed the existing training surveys and suggested amendments so that the survey items could be matched to our IPE RQs. The survey approach, including sample questions, were piloted as part of Bath Spa University's pre-trial training practice in the 2023/2024 academic year. The survey questionnaire was refined and finalised for the trial itself, drawing on insights from the Bath Spa University pilot about teachers' engagement with the survey and logistical considerations of completing the survey during the training sessions.

Trainers asked attendees to complete these questionnaires at the end of each of the teacher training sessions. The expected completion time was ten minutes.

Training survey responses were high across all three training days: 211 attendees completed items relating to Day 1; 189 completed items for Day 2; and 185 for Day 3.

IPE activities related to programme delivery

Teacher surveys (baseline and endline)

We administered two online surveys to both Focus4TAPS and teaching as usual group Year 5 teachers: one before the programme began; and one after it had finished. Survey links were emailed to each school's lead contact (usually the science lead), who was asked to share them with their Year 5 teachers.[13] The surveys collected information on teachers' usual science teaching practices, their confidence in delivering enquiry and formative assessment, and their views on pupils' engagement in science. By comparing responses at pre- and post-intervention, we were able to examine changes in teachers' confidence and practice over time, and to compare these changes between Focus4TAPS and teaching as usual group teachers.

In addition to these common questions, the post-intervention survey for Focus4TAPS teachers included programme-specific questions about implementation. These captured the number and length of Focused Assessment lessons delivered, the extent to which different elements of Working Scientifically were taught, and the time teachers spent training, planning, and delivering lessons. These helped us understand how the programme was delivered in schools and estimate the cost per pupil. Control teachers did not receive these implementation questions and instead continued to report on their usual science practice, which allowed for a clear comparison between the two groups at both time points.

The pre-intervention survey was conducted at baseline in September 2024, and the post-intervention survey was conducted at endline in June – July 2025. The pre-intervention survey took an average of nine minutes to complete, while the post-intervention survey took 11 minutes to complete.

For the pre-intervention survey, we received responses from 198 Year 5 teachers: 97 from the control group teachers; and 101 from the intervention group teachers.

For the post-intervention survey, we received responses from 133 Year 5 teachers: 71 from the control group teachers; 53 from the intervention group teachers; and nine from teachers who were unsure which group they belonged to. The teachers who were unsure of their group were only asked background questions in the survey.[14]

Science lead survey (endline)

We conducted an online survey with science leads at intervention and control group schools at endline. This post-intervention survey took place in June – July 2025. We received responses from 91 science leads in the post-intervention survey.

Because some teachers were both class teachers as well as science leads in their schools, this survey was merged with the post-intervention teacher survey and was conducted using the same survey link. This meant that any respondents who indicated that they were a science lead at their school in the post-intervention teacher survey were asked additional questions (within the same survey link).

These questions were used to gather data from senior science teachers in each school about their school's usual practice. Additionally, for intervention schools, we asked questions about their understanding of the Focus4TAPS programme, the benefits and challenges of implementing it in their schools, and any plans for supporting more staff to implement the programme after the intervention.

[13] Due to this design, it is not possible to provide exact response rates for the survey as we do not know the total number of teachers that received the survey link.

[14] The post-intervention teacher survey was administered as a single, common instrument to both intervention and control group teachers, rather than as separate group-specific surveys. As a result, teachers' trial allocation could not be inferred from the survey itself.

School-level survey responses

At a school level, endline survey responses were collected from 101 schools with 55 control group schools, 42 Focus4TAPS group schools and four where teachers did not provide their schools' name. Table 10 provides more details about the split by role within schools.

Table 10: Number and type of endline survey respondents

Number and type of respondent for endline (per school)	No. of schools (n)	% of schools
One respondent – Year teacher only	11	11
One respondent – Science lead only	39	39
One respondent – Dual role of both science lead and Year 5 teacher	20	20
Two respondents – One Year 5 teacher and one science lead	30	30
Two respondents – One with a dual role (Year 5 teacher/science lead) and one science lead	1	1
Total	101	101[a]

[a] The total is 101% due to rounding.

Pupil surveys (baseline and endline)

We added two IPE questions to the impact evaluation pupil surveys to provide contextual information on science teaching. The two questions focused on whether parents/carers of the children had an interest in science. Although these were originally added for the IPE, they were analysed alongside the attitudes and beliefs questions for the impact evaluation.

Pupils in both intervention and control schools completed these questions as a part of the single paper-based surveys at baseline and endline. Bringing IPE and impact evaluation questions together in a single survey questionnaire streamlined the process as much as possible for schools to deliver these to pupils. The baseline survey was completed in September 2024; the post-intervention/endline survey was conducted in June – July 2025.

Teacher lesson reflections

As a part of the pre-trial training, Bath Spa University had piloted conducting a reflective exercise with the attendees. The university suggested that a similar exercise could be used for the effectiveness trial.

Based on this feedback, we asked teachers from intervention schools to complete a written task where they were asked to reflect on a Focused Assessment lesson that they delivered in their schools. This was carried out during the midpoint of the intervention in January and February 2025. Teachers were asked to take notes of how they delivered the lessons against their lesson plans immediately or soon after delivery of the lessons and bring those to their second training session for a reflective exercise. NatCen added questions to Bath Spa University's existing reflective training exercises that were completed during the second training sessions to minimise burden on participants. Bath Spa University collected these notes via an anonymous online form as part of Day 2 training and shared them with NatCen.

Bath Spa University collected lesson reflection notes from 184 teachers during Day 2 of the training. We sampled 50% of the lesson reflections (92 teachers) across regions, ensuring that all regions were covered and that a range of different teachers' experiences were reflected within each region. Data analysis was conducted on a representative subsample of lesson reflections (one per group).

Qualitative case studies

We conducted qualitative research in a sample of 13 Focus4TAPS schools to build a rich and detailed understanding of implementation and delivery in a range of contexts. Schools were selected to ensure that the sample was as representative as possible of schools in England, with diversity across school and staff characteristics. Of the 13 schools, seven were Multi-Academy Trusts (MATs).

Data collection across the 13 case study schools was organised to capture a diversity of views:

- Six schools were asked to participate in Year 5 lesson observations followed by pupil focus groups. Of the six schools that were selected to participate in Year 5 lesson observations and pupil focus groups, three schools had FSM higher than 24%.[15]

- Seven schools were selected to participate in teacher focus groups.

The case studies provided a more detailed picture of the delivery of Focus4TAPS under real-world conditions. School data collection took place between October 2024 and July 2025.

- **Teacher lesson observations.** To gain first-hand insight of how Focused Assessment lessons were implemented, we carried out in-person lesson observations in six schools. These observations aimed to capture classroom delivery, pupil engagement, and how teachers used Focus4TAPS materials during lessons. We used these observations alongside teachers' written reflections to inform the development of topic guides for teacher and pupil focus groups. Lesson observations were spread across the second part of the academic year, which helped provide an overview of the different stages of teachers' training and pupils' development.

- **Pupil focus groups.** Five pupils from each of the six schools where observations took place took part in in-person focus groups. These groups covered pupils' experiences of learning science, and aspects of Working Scientifically and the Focus4TAPS intervention specifically. The focus groups aimed to understand pupils' views on Working Scientifically, their engagement with Focused Assessments, and how Focus4TAPS compared with their usual science lessons. Discussions followed a semi-structured topic guide and covered themes such as enjoyment of science, perceived learning benefits, and differences between Focus4TAPS and other science activities. Each focus group lasted approximately 60 minutes.

  In advance of the school visits, teachers were asked to select pupils based on diversity across key characteristics including gender, ethnicity, and eligibility for FSM. Pupils also needed to be comfortable to speak to a researcher as a part of a group. We asked teachers to distribute information sheets to the parents/carers of the children who were selected. The information sheets provided information about the focus groups and asked for their consent for their child to take part. We followed an opt-out approach to consent where parents/carers were able to let the teacher know if they did not want their children to take part in the focus groups. Pupils were also asked to provide their verbal consent to participate in the focus groups. Pupil focus groups were scheduled to take place on the same day as lesson observations to minimise teacher burden in terms of organising visits.

- **Teacher focus groups:** We conducted two online focus groups with teachers to discuss their experiences of delivering Focus4TAPS. These groups were used to follow-up on findings from the surveys, lesson observations, and lesson reflections. They aimed to explore implementation experiences, perceived impacts, and contextual factors influencing classroom delivery. One focus group included two attendees, while the other included five attendees, with all teachers coming from different schools/regions. It is important to note that while focus groups were meant specifically for Year 5 teachers, four out of the seven participants mentioned during the discussions that they held other roles at their schools like SENCOs, science coordinators, and also taught other year groups like Year 2 and Year 4. Each focus group lasted approximately 60 minutes.

We provided case study schools with a financial incentive for their participation. We gave schools a total of £50 for participating in the teacher lesson observations and pupil focus groups. For teachers who participated in the focus groups, we gave them £20 per teacher.

Attendance data

We collected attendance data from Bath Spa University about the training attendees of the three training sessions. Bath Spa University recorded this information in a monitoring data spreadsheet and shared the data in May 2025.

[15] This threshold aligned with the national average FSM eligibility in primary schools which was 24.7% (DfE, 2025). Purposive sampling of schools against this characteristic ensured that settings serving communities with higher-than-average levels of socio-economic disadvantage were represented in the study.

Analysis

All interviews and focus groups were audio-recorded (with participant permission), and transcribed verbatim. Observation data was recorded in written fieldnotes using pro forma templates.

We managed and analysed the qualitative data using NatCen's Framework approach (Ritchie *et al.,* 2014), wherein data are systematically summarised into a thematic framework to ensure they are comprehensively ordered and accessible for qualitative analysis.

We developed thematic frameworks for each research activity (the lead teacher interview, trainer interviews, pupil focus groups), using themes from the topic and observation guides and emerging from the data. These frameworks were assembled into matrices, with rows representing each case (an individual interview, observation, or focus group), and columns themes/subthemes. Transcripts and observation notes were systematically reviewed and summarised into the matrices by theme, including illustrative verbatim quotes where appropriate.

The final analytic stage involved working through the managed data to map the full range of views and experiences across participants, including identification of commonalities and differences across groups.

As part of the IPE, we considered factors that might influence the pathways to change outlined in the Focus4TAPS logic model. These included characteristics of programme participants (such as pupil FSM status, teachers' level of teaching experience), contextual factors (e.g. including region and type of setting), and implementation factors (such as fidelity, dosage). However, the available data did not allow for formal testing of their moderating effects. Where differences were noticed by implementation factors, these have been highlighted within the IPE results.

We used SPSS software (SPSS Inc., Chicago, Illinois, USA) to manage and analyse the quantitative survey data using descriptive statistics including frequency distributions and crosstabulations. As part of this analysis, we looked at differences between control and intervention groups and, where possible, over time in the pre- and post-intervention survey data. Response rates were not sufficiently large to conducts tests for statistical significance.

We triangulated and synthesised data collected throughout the IPE data to answer our RQs.

Costs

The cost evaluation used data on the time and cost involved in Focus4TAPS programme delivery, collected from schools through the teacher questionnaire, and from the programme developer who delivered the training of trainers in the effectiveness trial (Bath Spa University).

Questions for teachers covered the time in hours spent on attending Focus4TAPS training, engaging with training materials outside of training, using Focus4TAPS resources to develop teaching practice outside of their usual teaching responsibilities, delivering Focus4TAPS lessons, and sharing Focus4TAPS learning with other school staff. The teacher questionnaire also included questions on additional costs incurred to schools in the training and delivery of Focus4TAPS.

The cost evaluation was conducted in line with the EEF cost guidance (EEF, 2023).[16] Analysis considered the additional cost incurred by Focus4TAPS schools in terms of additional teacher time spent delivering the programme. Results are reported as averages of available teacher responses for each question. As Focus4TAPS implementation relies on existing school facilities and its implementation involves changes to the delivery of science lessons rather than making use of new resources, the teacher questionnaire asked teachers to report any additional costs incurred outside of usual teaching practice.

[16] Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1765404326

Information was also collected from Bath Spa University on delivery costs, covering personnel costs for preparing programme delivery, training of the trainers, and any additional implementation costs. These costs are incurred by the developer and are not included in calculating the per pupil per year cost of implementation from the schools' perspective (EEF, 2023). In line with the EEF cost guidance (EEF, 2023), costs were estimated using market practices—training costs were applied as if outside of the trial. Teachers also reported individual expenditure associated with attending training, such as travel expenses that they may not have been able to claim back. These costs are not included in the cost evaluation as they were not incurred by the school.

Costs are presented as both monetary and time units, by estimating an hourly cost for teaching cover. Teacher cover costs were calculated using salary information provided in the teacher questionnaire. To estimate the monetary teacher cost of the programme, we calculate the average hourly pay, assuming a maximum of 1,265 hours completed by teachers in a year.

Salary information was provided by 21 teachers in the teacher survey, with average yearly salaries summarised in Table 11.

Table 11: Average reported teacher salaries

Teacher role	No. of responses	Mean (SD)	Estimated hourly pay (per 1,265 hours)
All teachers	21	£42,648 (£11,113)	£33.71
Year 5 teachers	10	£37,322 (£8,526)	£29.51
Science leads	16	£45,680 (£10,568)	£36.11

Costs for the programme were estimated as if implemented over three years, so costs related to engagement with Focus4TAPS resources and further capacity building within schools are assumed to involve recurring teacher time being spent in subsequent years. The cost per pupil per year was calculated by dividing the cost per year by the number of pupils, with an average of 23 pupils per school per class assumed (equivalent to the average [arithmetic mean] class size in the analysis sample).

Timeline

Table 12: Timeline

Dates	Activity	Staff responsible / leading
January 2024 – May 2024	Recruitment of settings	Bath Spa University
January 2024 – May 2024	Schools provided pupil and teacher data and administered parent/carer opt-out	Schools
June 2024	Randomisation of schools Schools notified of the outcome of randomisation (by July 2024)	NatCen
June 2024	IPE programme developer interview	NatCen
July 2024 – August 2024	IPE document analysis	NatCen
August 2024	IPE case study school sampling	NatCen
September 2024	Additional time for schools to administer new parent/carer opt-out	Settings
September 2024 – October 2024	Baseline data collection with all schools	NatCen
October 2024 January 2025 – February 2025 April 2025 –May 2025	Intervention training	Bath Spa University
October 2024 – July 2025	Intervention delivery	Bath Spa University
October 2024 – July 2025	Data collection from case study schools	NatCen
October 2024 January 2025 – February 2025 May 2025 – June 2025	IPE training observations and surveys	NatCen
October 2024 – November 2024 February 2025 – March 2025 May 2025	IPE trainer interviews	NatCen
January 2025 – February 2025	Teacher lesson reflections	NatCen
June 2025 – July 2025	Endline data collection with all schools	NatCen
October 2024 – October 2025	Data management, analysis, and reporting	NatCen
September 2025 – December 2025	Draft report submitted	NatCen
January 2026 – May 2026	Comments and report finalisation	NatCen, Bath Spa University, the EEF
June 2026	Archiving	NatCen

Impact evaluation results

This section presents the impact evaluation findings for the Focus4TAPS evaluation. We first describe participant flow through the trial, attrition, and baseline balance, before reporting the results from the primary and secondary outcome analyses and sensitivity, subgroup, and non-compliance analyses.

Participant flow including losses and exclusions

Figure 4 presents the participant flow diagram from initial school recruitment through to the primary outcome analysis stage. The participant flow diagram summarises the number of schools and pupils included in the trial at each stage from recruitment through to analysis.

In total, 260 schools registered interest in participating in the Focus4TAPS trial, of which 249 schools signed an MoU. Schools that registered interest but did not sign the MoU did so for a range of reasons, including some who were ineligible, and some schools deciding to take part in the EEF-funded trial PSQM. After signing the MoU, a further nine schools either withdrew from the trial or failed to upload pupil data (or provide assurances that pupil data would be uploaded in the future) in time for randomisation and were therefore removed from the sample of schools included at randomisation.

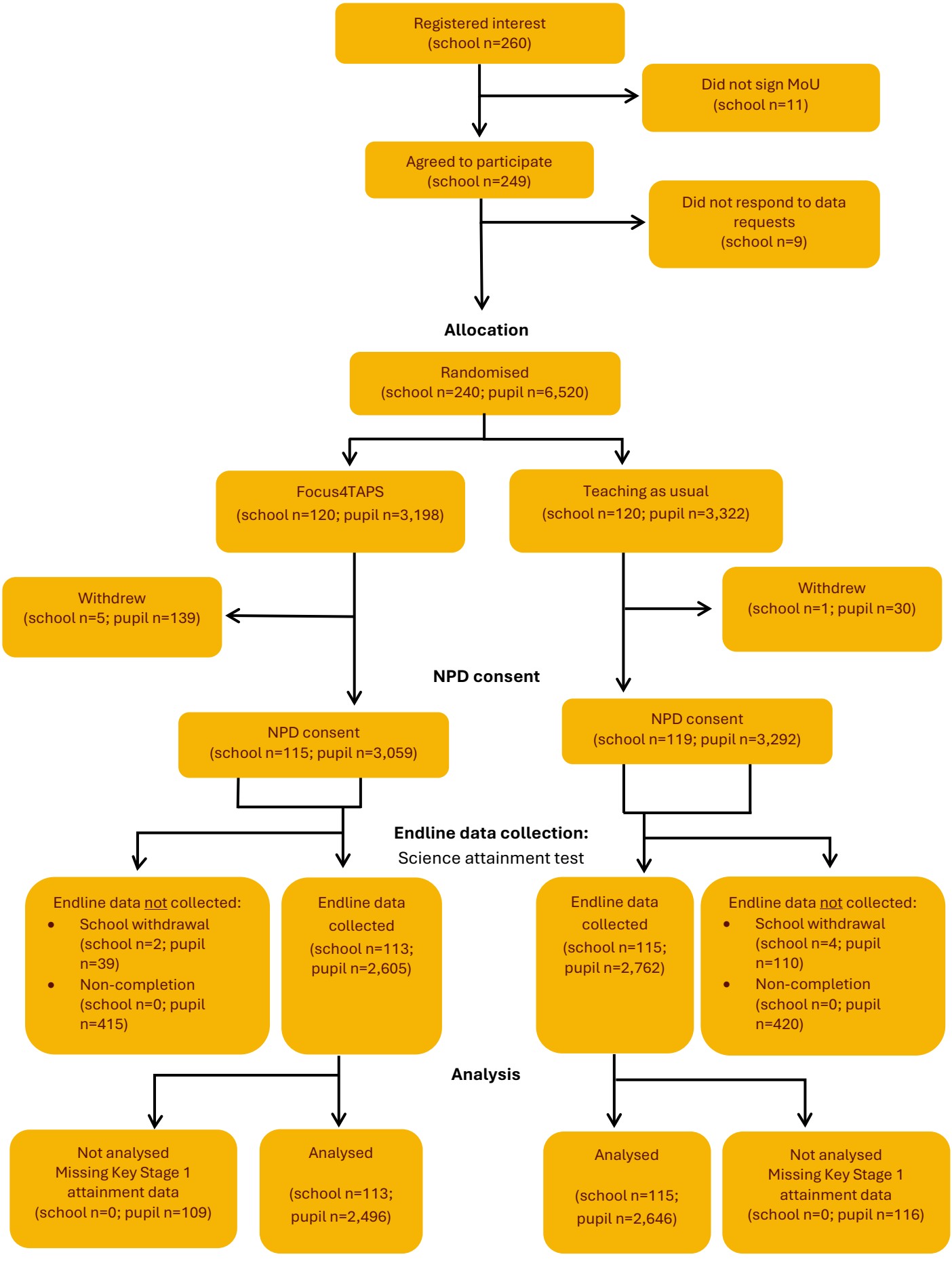
A total of 240 schools were randomised, with 120 schools allocated to receive the Focus4TAPS intervention (n=3,198 pupils) and 120 schools allocated to the teaching as usual group (n=3,322 pupils). Following randomisation, six schools withdrew from the trial (five Focus4TAPS schools, 139 pupils; one teaching as usual school, 30 pupils). To avoid statistical disclosure, the counts of pupil withdrawals from schools that withdrew from the trial and individual pupil-level withdrawals from remaining schools are reported together. Reasons for school withdrawal included changes in staff, schools moving sites which limited staff time and resources for the trial, and staff workload from trial participation being too burdensome. Data from these schools, and from pupils who withdrew from the trial, were not processed or linked to Key Stage 1 data in the NPD. Consent for linkage to Key Stage 1 NPD data was obtained for 6,351 pupils across 234 schools (115 Focus4TAPS schools, n=3,059 pupils; 119 teaching as usual schools, n=3,292 pupils).

Endline science attainment data were not collected for 984 pupils. This comprised 149 pupils from the six schools that withdrew before endline testing[17] (two Focus4TAPS schools, n=39 pupils; four teaching as usual schools, n=110 pupils) and 835 pupils from the remaining 228 schools (n=415 pupils in Focus4TAPS; n=420 pupils in teaching as usual) whose science attainment data were missing because of absence on the day of testing, movement between classes, or other unspecified reasons. These 984 pupils were therefore excluded from the primary analysis.

Following linkage to Key Stage 1 NPD data, a further 225 pupils (n=109 Focus4TAPS; n=116 teaching as usual) were excluded from the primary analysis because Key Stage 1 attainment data (reading and/or mathematics) were unavailable. The final primary analysis sample therefore comprised 5,142 pupils (n=2,496 Focus4TAPS; n=2,646 teaching as usual) across 228 schools (n=113 in Focus4TAPS; n=115 in teaching as usual). Further details are provided in the 'Attrition' section below.

[17] These schools did not withdraw consent for previously collected data to be processed.

Figure 4: Participant flow, recruitment to primary outcome analysis



Attrition

Pupil-level attrition was monitored throughout the trial from randomisation to the primary outcome analysis stage. Between randomisation and the NPD consent stage, pupil-level attrition was 4.35% in the Focus4TAPS intervention group and 0.90% in the teaching as usual group, representing 2.59% overall attrition. Further details on participant losses and exclusions are provided in the 'Participant flow including losses and exclusions' section above.

NPD consent was obtained for 6,351 pupils (n=3,059 Focus4TAPS; n=3,292 teaching as usual) across 234 schools.

From the NPD consent stage to the primary analysis stage, 1,209 pupils were excluded from the analysis sample. This comprised 984 pupils who did not have endline science attainment data available, including 149 pupils from schools that withdrew prior to endline testing and 835 pupils who did not complete the science attainment test. A further 225 pupils were excluded because Key Stage 1 attainment data (reading and/or maths) were unavailable in the NPD. The final primary analysis sample therefore comprised 5,142 pupils across 228 schools (n=2,496 Focus4TAPS; n=2,646 teaching as usual). For the primary analysis, attrition from the NPD consent stage to analysis was 18.40% in the Focus4TAPS group and 19.62% in the teaching as usual group, representing 19.03% overall.

Across the full trial period from randomisation to analysis, attrition was 21.95% for the Focus4TAPS group, 20.35% for the teaching as usual group, and 21.13% overall. This level of attrition was higher than the 16% pupil attrition rate assumed at the start of the trial. However, the achieved sample of 5,142 pupils was higher than expected, and the MDES at analysis stage was lower than expected (see Table 7 above for more information). Table 13 presents pupil-level attrition across the trial stages.

Table 13: Pupil-level attrition from the trial (primary outcome)

			Focus4TAPS	Teaching as usual	Overall
Randomised to NPD consent	No. of pupils	Randomised	3,198	3,322	6,520
		NPD consent	3,059	3,292	6,351
	Pupil attrition	Number	139	30	169
		Percentage	4.35%	0.90%	2.59%
NPD consent to analysed	No. of pupils	NPD consent	3,059	3,292	6,351
		Missing endline data	454	530	984
		Missing Key Stage 1 attainment data	109	116	225
		Analysed	2,496	2,646	5,142
	Pupil attrition	Number	563	646	1,209
		Percentage	18.40%	19.62%	19.03%
Randomised to analysed	No. of pupils	Randomised	3,198	3,322	6,520
		Analysed	2,496	2,646	5,142
	Pupil attrition	Number	702	676	1,378
		Percentage	21.95%	20.35%	21.13%

Pupil and school characteristics

To assess the balance of pupil and school characteristics at baseline, we present descriptive statistics for pupils and schools that provided consent to link to NPD data. Key Stage 1 reading and maths baseline attainment measures were not available at randomisation and could only be obtained through NPD linkage; therefore, balance is assessed for the NPD consent sample rather than the full randomised sample.

Table 14 presents school and pupil characteristics for the Focus4TAPS intervention and teaching as usual schools within the NPD consent sample. At the pupil level, gender and Key Stage 1 maths and reading attainment were similar across the Focus4TAPS and teaching as usual groups. FSM eligibility differed between the two groups, with a higher proportion of pupils in the Focus4TAPS group eligible for FSM (35.73%) compared with the teaching as usual group (30.47%), corresponding to an effect size of -0.11, higher than the threshold of 0.10 used to indicate imbalance. Baseline attitudes and beliefs to science subscale scores were comparable across the two groups, with all associated effect sizes falling below the pre-specified threshold of 0.05.

At the school-level, the Focus4TAPS and teaching as usual groups showed similar distributions across regions. However, schools in the teaching as usual group had higher mean Key Stage 1 maths attainment (defined as the mean proportion of pupils reaching the expected standard) than those in the Focus4TAPS group (effect size = 0.36), and higher mean Key Stage 1 reading attainment (effect size = 0.36). In addition, schools in the teaching as usual group had a lower mean proportion of pupils eligible for FSM than schools in the Focus4TAPS group (effect size = -0.25). As these effect sizes exceeded the pre-specified threshold of 0.10, we treated them as indicative of potential school-level imbalance.

As specified above in the 'Methods' section, we re-estimated the primary outcome model adjusting for pupil-level FSM eligibility and school-level FSM eligibility, Key Stage 1 maths attainment, and Key Stage 1 reading attainment to account for baseline imbalance in these pupil- and school-level characteristics. The 'Outcomes and analysis' section below report the results of this sensitivity analysis.

Table 14: Baseline characteristics of groups as randomised

Pupil-level categorical	Focus4TAPS		Teaching as usual		Overall
	n/N (missing)	Count %	n/N (missing)	Count %	Effect size Hedges' g
Gender: Boy	1,544/3,059 (0)	50.47	1,715/3,292 (SUPP[a])	52.10	0.03
Key Stage 1 maths attainment: Expected standard or above	1,990/3,059 (156)	65.05	2,203/3,292 (166)	66.92	0.04
Key Stage 1 reading attainment: Expected standard or above	1,982/3,059 (156)	64.79	2,129/3,292 (166)	64.67	0.00
FSM eligible: Yes	1,093/3,059 (SUPP)	35.73	1,003/3,292 (SUPP)	30.47	-0.11
Pupil-level continuous	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size Hedges' g
Interest and enjoyment in science subscale score	2,547/3,059 (512)	3.15 (0.67)	2,507/3,292 (785)	3.16 (0.65)	0.00
Confidence in science subscale score	2,551/3,059 (508)	2.92 (0.66)	2,508/3,292 (784)	2.90 (0.65)	-0.03
Perceptions of science teachers' teaching subscale score	2,543/3,059 (516)	3.26 (0.57)	2,500/3,292 (792)	3.28 (0.55)	0.04
Self-regulation of learning in science subscale score	2,536/3,059 (523)	2.91 (0.69)	2,498/3,292 (794)	2.92 (0.67)	0.01

Self-efficacy for Working Scientifically subscale score	2,518/3,059 (541)	2.95 (0.69)	2,489/3,292 (803)	2.97 (0.68)	0.02
Working Scientifically beliefs subscale score	2,490/3,059 (569)	3.33 (0.55)	2,466/3,292 (826)	3.35 (0.54)	0.02
Wider benefits of science subscale score	2,517/3,059 (542)	3.34 (0.66)	2,484/3,292 (808)	3.33 (0.64)	-0.01
School-level categorical	n/N (missing)	Count %	n/N (missing)	Count %	Effect size Hedges' g
Region: Bury	239/3,059 (0)	7.81	260/3,292 (0)	7.90	0.00
Region: Derby	193/3,059 (0)	6.31	214/3,292 (0)	6.50	0.01
Region: Essex	250/3,059 (0)	8.17	222/3,292 (0)	6.74	-0.05
Region: Liverpool	221/3,059 (0)	7.22	212/3,292 (0)	6.44	-0.03
Region: Newcastle	374/3,059 (0)	12.23	468/3,292 (0)	14.22	0.06
Region: Nottingham	170/3,059 (0)	5.56	130/3,292 (0)	3.95	-0.08
Region: SE London	186/3,059 (0)	6.08	256/3,292 (0)	7.78	0.07
Region: Sheffield	341/3,059 (0)	11.15	307/3,292 (0)	9.33	-0.06
Region: Southampton	179/3,059 (0)	5.85	201/3,292 (0)	6.11	0.01
Region: Suffolk	156/3,059 (0)	5.10	236/3,292 (0)	7.17	0.09
Region: Sunderland	310/3,059 (0)	10.13	327/3,292 (0)	9.93	-0.01
Region: West London	192/3,059 (0)	6.28	192/3,292 (0)	5.83	-0.02
Region: Wolverhampton	248/3,059 (0)	8.11	267/3,292 (0)	8.11	0.00
School-level continuous	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size Hedges' g
Key Stage 1 maths attainment (proportion of pupils reaching the expected standard or above)	2,619/3,059 (440)	65.23 (15.54)	2,800/3,292 (492)	70.47 (13.82)	0.36
Key Stage 1 reading attainment (proportion of pupils reaching the expected standard or above)	2,636/3,059 (423)	64.03 (17.49)	2,862/3,292 (430)	69.46 (12.91)	0.36
FSM eligibility (proportion of pupils eligible for FSM at any point in the last six years)	2,873/3,059 (186)	30.70 (17.35)	3,146/3,292 (146)	26.60 (15.60)	-0.25

Note: For binary variables, a single category is reported; the complementary category is not reported as results would be mirrored.

[a] *SUPP represents counts that have been suppressed (less than ten), in line with SRS statistical disclosure guidance.*

Outcomes and analysis

Primary analysis

The primary analysis explored the impact of Focus4TAPS on the science attainment of Year 5 pupils, using Key Stage 1 maths and reading attainment as the baseline measures.

Both Key Stage 1 maths and reading were coded as binary indicators of whether pupils were working below or working at/above the expected standard. Figure 5 shows that, in the primary analysis sample (5,142 pupils with available data), just under one-third of pupils (29.93%; n = 1,539) were working below the expected standard in Key Stage 1 reading, while just over two-thirds (70.07%; n=3,603) were working at or above the expected standard. Key Stage 1 reading attainment was moderately associated with science attainment (r = 0.49, p<0.001). Similarly, Figure 6 shows that just under one-third of pupils (28.67%; n=1,474) were working below the expected standard in Key Stage 1 maths, while just over two-thirds (71.33%; n=3,668) were working at or above the expected standard. Key Stage 1 maths attainment was moderately associated with science attainment (r = 0.50, p<0.001).

Figure 5:5 Distribution of Key Stage 1 reading attainment, by expected standard

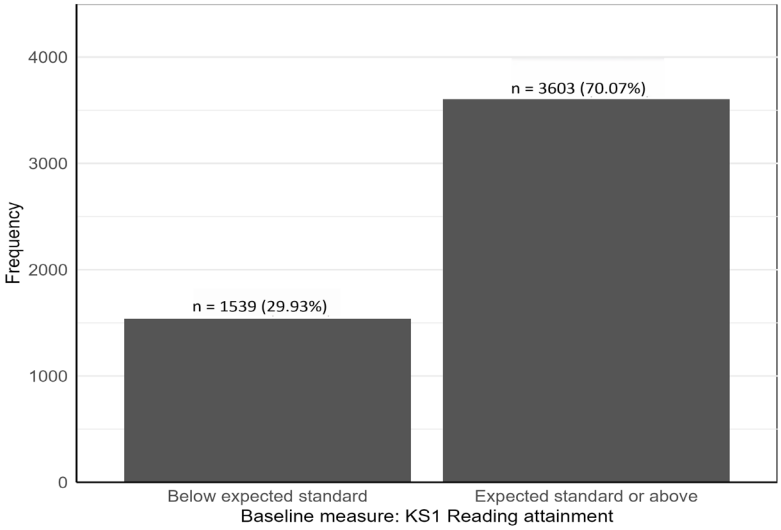
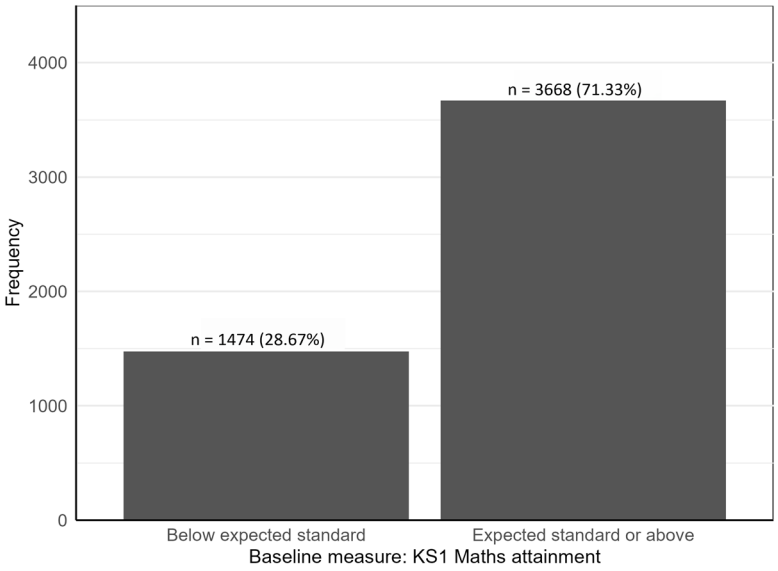
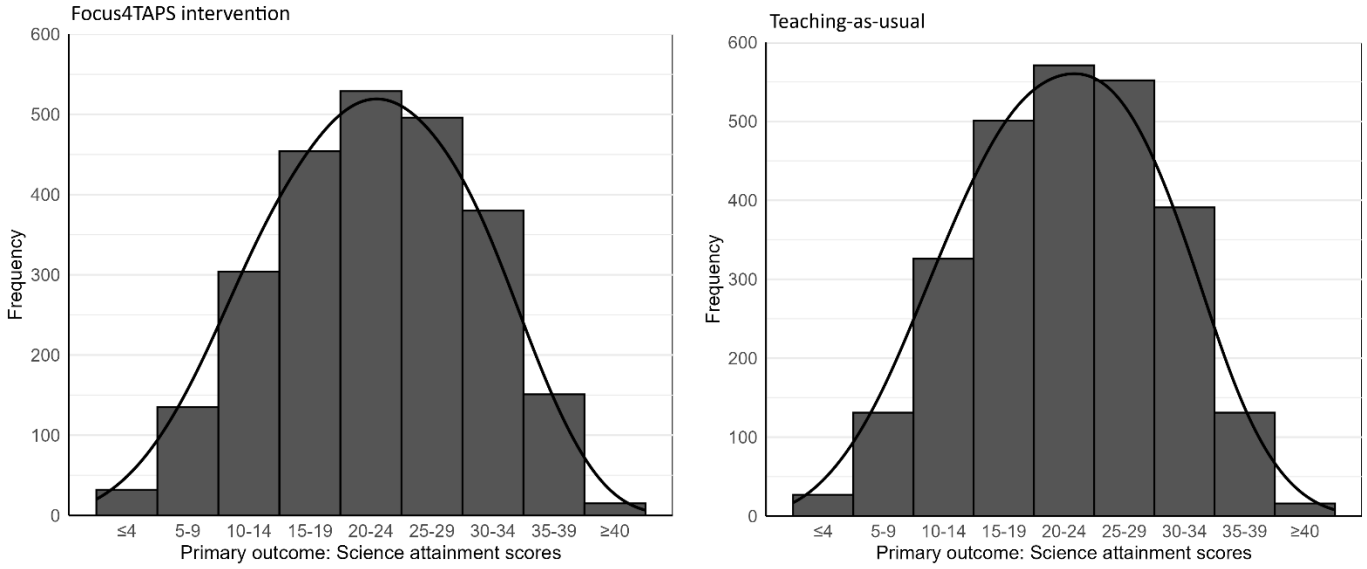


Figure 6: Distribution of Key Stage 1 maths attainment, by expected standard



Across the primary analysis sample, pupils had a mean science attainment score of 22.29 (SD = 8.13). Figure 7 presents the distribution of science attainment scores separately for the Focus4TAPS and teaching as usual groups.[18] Science attainment scores were approximately normally distributed in both groups. The Focus4TAPS group had a mean score of 22.31 (SD = 8.24), and the teaching as usual group had a similar mean score of 22.27 (SD = 8.03). In both groups, scores were concentrated towards the middle of the scale, with the highest proportion of pupils scoring within the 20–24 score band.

Figure 7: Distribution of science attainment scores, by group



Residuals were plotted against fitted values to assess potential heteroskedasticity (unequal variance of residuals across the variable values).[19] A Breusch-Pagan test indicated evidence of heteroskedasticity.[20]

The primary analysis followed an ITT approach using a multilevel model that controlled for baseline Key Stage 1 maths and reading attainment, randomisation strata, age in months, and school-level clustering. Results are presented in Table 15, with effect size estimates in Table 16. The adjusted mean difference was 0.24 (p = 0.69). The corresponding Hedges' g effect size was 0.03 (95% CI: -0.11 to 0.17), indicating a very small positive effect. The CIs, however, include both positive and negative values, providing no conclusive evidence that Focus4TAPS improved pupils' science attainment.

The achieved sample size for the primary analysis provided sufficient statistical power to detect a MDES of 0.149, which was smaller than the effect size specified at randomisation. The observed effect was below this threshold. Although the CI includes effects of the magnitude the study was powered to detect, it includes zero suggesting that any true effect of Focus4TAPS on pupils' science attainment is likely to be small (Hemming and Taljaard, 2021).

The post-intervention ICC was estimated directly from the null (unadjusted) multilevel model used for the primary outcome analysis, using schools in both the Focus4TAPS and teaching as usual groups. The estimated school-level ICC was 0.13 (95% CI: 0.10 to 0.15).

[18] In line with SRS statistical disclosure guidance, science attainment scores were presented in nine categories to ensure all cell counts met the minimum disclosure threshold of ten.
[19] In line with SRS statistical disclosure guidance, these plots are not presented.
[20] As a robustness check, as specified in the SAP (Leonard *et al.,* 2025), we estimated White heteroskedasticity-robust standard errors and re-estimated the primary outcome model. See 'Additional analyses and robustness checks' section for further detail.

Table 15: Primary outcome analysis results, by science attainment

Outcome	Unadjusted differences in means	Adjusted differences in means	Focus4TAPS		Teaching as usual		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Science attainment	0.04	0.24	2,496 (563)	67.86	2,646 (646)	64.53	66.15

Table 16: Primary outcome analysis, by science attainment effect size estimation

Outcome	Focus4TAPS		Teaching as usual		Total n (Focus4TAPS, teaching as usual)	Hedges' g (95% CI)	P-value
	n (missing)	Unadjusted means (95% CI)	n (missing)	Unadjusted means (95% CI)			
Primary outcome: Science attainment	2,496 (563)	22.306 (21.98, 22.63)	2,646 (646)	22.27 (21.96, 22.57)	5,142 (2,496, 2,646)	0.03 (-0.11, 0.17)	0.69

Secondary analysis

The secondary outcomes of the Focus4TAPS programme were used to assess its impact on pupils' attitudes and beliefs about science, measured in seven subscales collected in the attitudes and beliefs to science questionnaire. Each measure was collected in both the baseline and endline questionnaires, allowing the analyses to employ a similar approach as the primary analysis using multilevel linear models that adjust for the baseline score for that subscale, randomisation strata, pupil age in months, and school-level clustering. All secondary outcome subscales were calculated as the average of the relevant questionnaire items, each rated on a 4-point Likert scale: 1 ('Disagree a lot'); 2 ('Disagree a little'); 3 ('Agree a little'); and 4 ('Agree a lot'), giving a possible range of 1 to 4 for all subscales.

Distributions of the secondary outcomes

In the baseline questionnaire, average scores for the secondary outcomes ranged from 2.92 for confidence in science to 3.34 for wider benefits of science and Working Scientifically. All subscale scores were negatively skewed, with most responses clustered toward the higher end of the scale and a longer tail toward lower scores. The confidence in science, self-efficacy and self-regulation subscale scores were less strongly skewed, while the remaining subscale scores were more strongly concentrated at the upper end of the distribution.

In the endline questionnaire, average scores for the secondary outcomes ranged between 2.78 for confidence in science and 3.42 for wider benefits of science. The scores remained negatively skewed, with most responses clustered towards higher scores and a long tail towards lower scores. The confidence in science, self-efficacy, and self-regulation scores were less strongly skewed. The interest and enjoyment in science scores were also more normally distributed at endline compared to baseline, although the score remained more skewed towards the upper end of the distribution.

The skewness observed in both the baseline and endline scores may have resulted in skewed or heteroskedastic residuals in the linear regression. While linear regression does not require the outcomes themselves to be normally distributed, skewed outcomes can increase the likelihood of violating assumptions about the residuals, particularly normality and constant variance. As a sensitivity check to assess whether skewness affected the model's standard errors, the secondary outcome models were re-estimated using robust standard errors (see Appendix F, Table 2 in Further Appendices). The use of robust standard errors did not meaningfully change findings, and these results are therefore reported in the appendix and not discussed further in the main body of the report.

Tables containing descriptive statistics for each secondary outcome, along with histograms showing the distribution of each outcome for the full sample at baseline and separately for the Focus4TAPS and teaching as usual group at endline, are presented in Appendix G in Further Appendices.

Secondary analysis results

The unadjusted mean differences in scores for each secondary outcome are presented in Appendix F, Table 1 in Further Appendices. These ranged from 0.01 for the perceptions of science teachers' teaching subscale to 0.08 for the confidence in science subscale. Overall, these results indicated a small positive difference between the Focus4TAPS and teaching as usual groups across most attitudinal outcomes.

The secondary analyses followed an ITT approach using a multilevel model that controlled for the corresponding baseline subscale score, randomisation strata, age in months, and school-level clustering. Effect size estimates are presented in Table 17.

After adjustment, the analysis found evidence of positive differences in adjusted mean scores between Focus4TAPS and teaching as usual schools for six of the attitudes to science subscales. These were: interest and enjoyment in science (ES: 0.16, 95% CI: 0.03 to 0.29); confidence in science (ES: 0.15, 95% CI: 0.04 to 0.26); perceptions of science teachers' teaching (ES: 0.17, 95% CI: 0.02 to 0.32); self-regulation of learning in science (ES: 0.18, 95% CI: 0.06 to 0.31); self-efficacy for Working Scientifically (ES: 0.11, 95% CI: 0.01 to 0.24); and Working Scientifically beliefs subscales (ES: 0.12, 95% CI: 0.01 to 0.23). The exception was pupils' views of the wider benefits of science, where there was no conclusive evidence of a difference in attitudes towards science (ES: 0.08; 95% CI: -0.03 to 0.19). In all cases Hedges' g effect sizes were below the threshold for a small effect (0.2), with estimates of effects ranging from 0.11 to 0.18. These findings indicated that, although there was evidence of a positive impact on pupils' attitudes towards science across several dimensions, the magnitude of impacts was small.[21]

As several secondary outcomes were examined, the p-values shown in Table 17 are presented after applying a Romano-Wolf step-down adjustment.[22]

[21] The effect sizes for the secondary outcomes cannot be expressed as months of extra progress because these outcomes are not direct measures of attainment.
[22] This procedure accounts for the increased chance of Type I errors while accommodating dependence among the outcomes. When evaluating statistical significance, only the Romano-Wolf p-value is referred to.

Table 17: Secondary analysis, by effect size estimations

Outcome	Unadjusted means				Effect size			Model coefficient		
	Focus4TAPS		Teaching as usual							
	N (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value	Coefficient (95% CI) Romano-Wolf adjusted	Romano-Wolf p-value	Unadjusted p-value
Interest and enjoyment in science	2,283 (776)	3.04 (3.01, 3.07)	2,215 (1,077)	2.99 (2.97, 3.02)	4,498 (2,283; 2,215)	0.16 (0.03, 0.29)	0.02	0.10 (0.02; 0.21)	0.02*	0.02
Confidence in science	2,289 (770)	2.82 (2.79, 2.85)	2,215 (1,077)	2.74 (2.72, 2.77)	4,504 (2,289; 2,215)	0.15 (0.04, 0.26)	0.01	0.10 (0.03; 0.17)	0.01**	0.00
Perceptions of science teachers' teaching	2,278 (781)	3.18 (3.15, 3.2)	2,201 (1,091)	3.16 (3.14, 3.19)	4,479 (2,278; 2,201)	0.17 (0.02, 0.32)	0.04	0.10 (0.02; 0.18)	0.02*	0.01
Self-regulation of learning in science	2,258 (801)	2.80 (2.77, 2.83)	2,197 (1,095)	2.78 (2.75, 2.81)	4,455 (2,258; 2,197)	0.18 (0.06, 0.31)	0.01	0.13 (0.04; 0.23)	0.01**	0.00
Self-efficacy for Working Scientifically	2,226 (833)	3.03 (3.01, 3.06)	2,164 (1,128)	2.98 (2.95, 3.01)	4,390 (2,226; 2,164)	0.11 (0.01, 0.24)	0.04	0.08 (0.01; 0.14)	0.02*	0.02
Working Scientifically beliefs	2,189 (870)	3.37 (3.35, 3.40)	2,138 (1,154)	3.34 (3.32, 3.37)	4,327 (2,189; 2,138)	0.12 (0.01, 0.23)	0.05	0.06 (0.01; 0.12)	0.02*	0.03
Wider benefits of science	2,235 (824)	3.44 (3.41, 3.46)	2,172 (1,120)	3.41 (3.38, 3.44)	4,407 (2,235; 2,172)	0.08 (-0.03, 0.19)	0.18	0.05 (-0.01; 0.11)	0.13	0.20

*Asterisks indicate statistical significance based on Romano-Wolf adjusted p-values: * p<0.05; ** p<0.01.*

Subgroup analyses

Pupils eligible for FSM

Subgroup analyses examined whether Focus4TAPS had a different impact for pupils who had been eligible for FSM in the last six years. Following the EEF statistical guidance (EEF, 2022), this was conducted using a restricted-sample approach, whereby the primary analysis model was re-estimated on a subgroup of pupils eligible for FSM with available data on all variables for the model. A multilevel model was estimated controlling for baseline Key Stage 1 maths and reading attainment, randomisation strata, pupils' age in months, and school-level clustering. Results are presented in Table 18, with effect sizes in Table 19. The adjusted mean difference was 0.55 (p = 0.48). The corresponding Hedges' g effect size was 0.07, indicating a small positive effect. However, as the CI includes both zero and negative values (ranges from -0.12 to 0.25), we do not find conclusive evidence that Focus4TAPS improves science attainment among FSM pupils.

Table 18: FSM subgroup analysis results, by science attainment

Outcome	Unadjusted differences in means (95% CI)	Adjusted differences in means (95% CI)	Focus4TAPS		Teaching as usual		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Primary outcome: Science attainment	-0.34 (-1.11, 0.43)	0.55 (-0.92, 2.01)	864 (229)	63.849	769 (234)	61.347	62.67

Table 19: FSM subgroup analysis, by science attainment effect size estimation

Outcome	Focus4TAPS		Teaching as usual		Total n (Focus4TAPS; teaching as usual)	Hedges' g (95% CI)	P-value
	n (missing)	Unadjusted means (95% CI)	n (missing)	Unadjusted means (95% CI)			
Primary outcome: Science attainment	864 (229)	19.172 (18.64, 19.71)	769 (234)	19.512 (18.96, 20.07)	1,633 (864; 769)	0.07 (-0.12, 0.25)	0.48

In addition, an interaction model was estimated to assess whether Focus4TAPS had a different effect for FSM pupils compared with non-FSM pupils. This model used the full analytical sample and included treatment allocation, an FSM eligibility indicator, and an interaction term between treatment allocation and FSM status, while controlling for baseline Key Stage 1 maths and reading attainment, randomisation strata, pupils' age in months, and school-level clustering. The interaction term coefficient was -0.21, with a p-value of 0.61 suggesting that Focus4TAPS did not have a differential impact on FSM pupils' science attainment compared with non-FSM pupils. The interaction model coefficients are reported in Appendix D, Table 1 in Further Appendices.

In line with the EEF statistical guidance (EEF, 2022), we derived the estimated effect size for FSM pupils from the interaction model and compared it with the effect size from the restricted FSM subgroup analysis. The resulting effect size was smaller than that from the restricted subgroup analysis (ES: 0.02, 95% CI: -0.13 to 0.18). However, the CIs were consistent across both approaches, providing no conclusive evidence that Focus4TAPS improves science attainment among FSM pupils.

As baseline differences in FSM eligibility were observed between the two groups (as presented in the 'Pupil and school characteristics' section above), we conducted a sensitivity analysis by re-estimating the interaction model to assess whether these differences contributed to the variation in effect size estimates between the restricted subgroup and the interaction analysis. The model was extended to control for all characteristics that were imbalanced at baseline, including school-level FSM eligibility and Key Stage 1 maths and reading attainment.

The interaction term coefficient was -0.35 (p = 0.45), suggesting that Focus4TAPS did not have a differential impact on FSM pupils' science attainment. The corresponding effect size was slightly higher than in the interaction model not controlling

for imbalance (ES: 0.04, 95% CI: -0.12 to 0.21) but remained very small, with CIs consistent across both models. The interaction model coefficients are reported in Appendix D, Table 10 in Further Appendices.

Taken together, findings across all approaches provide no conclusive evidence that Focus4TAPS improves science attainment among FSM pupils.

Secondary outcomes, among pupils eligible for FSM

When the existing model for each secondary outcome was extended to include an interaction effect between FSM and Focus4TAPS, the analysis found no evidence of differential effects for any of the secondary outcomes. These models, therefore, provide no conclusive evidence that the impact of Focus4TAPS on secondary outcomes was moderated by FSM eligibility. The results of these interaction models are presented in full in Appendix F, Tables 4 and 5 in Further Appendices.

The analysis was then repeated without the interaction term, restricting the analysis sample to FSM pupils with available data on all variables for the model. Within this subgroup, the analysis found evidence of a positive effect for two outcomes: self-regulation of learning in science (ES: 0.26, 95% CI: 0.07 to 0.45); and perceptions of science teachers' teaching (ES: 0.23, 95% CI: 0.03 to 0.43) (Table 20). For all other secondary outcomes, after adjustment for multiple comparisons, there was no conclusive evidence of an impact of the Focus4TAPS intervention. The effect sizes are presented in Table 20.

It should be noted that this subgroup analysis was based on a smaller sample size, meaning that effects on other secondary outcomes may not have been detected with sufficient statistical power, given the available data. For example, the number of cases included in the FSM subgroup model for interest and enjoyment in science was 1,390, compared with 4,498 for the analysis of all pupils (Table 20).

Given the subsample analysis found evidence of a positive effect in the FSM subsample, and considering the imbalance in the sample at baseline identified in pupil-level FSM eligibility and school-level FSM eligibility, Key Stage 1 maths attainment, and Key Stage 1 reading attainment, the interaction model was then repeated—including additional controls for these imbalanced school-level covariates (pupil-level FSM is already included in the model). However, the inclusion of these additional covariates did not change our initial finding of there being no differential effects for any of the secondary outcomes by FSM (Appendix F, Tables 14 and 15 in Further Appendices). This finding suggests therefore that there is no conclusive evidence of differential impacts for FSM-eligible pupils on secondary outcomes, after accounting for imbalance in characteristics.

Table 20: Secondary analysis, by effect size estimation for pupils eligible for FSM

Outcome	Unadjusted means				Effect size			Model coefficient p-values		
	Focus4TAPS		Teaching as usual							
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value	Coefficient (95% CI) Romano-Wolf adjusted	Romano-Wolf p-value	Unadjusted p-value
Interest and enjoyment in science	742 (351)	3.00 (2.95, 3.05)	648 (355)	2.97 (2.92, 3.03)	1,390 (742; 648)	0.17 (-0.01, 0.35)	0.08	0.12 (-0.03; 0.26)	0.11	0.07
Confidence in science	749 (344)	2.72 (2.68, 2.77)	647 (356)	2.68 (2.63, 2.73)	1,396 (749; 647)	0.17 (-0.00, 0.34)	0.06	0.11 (-0.01; 0.23)	0.06	0.04
Perceptions of science teachers' teaching	744 (349)	3.19 (3.14, 3.23)	644 (359)	3.14 (3.09, 3.19)	1,388 (744; 644)	0.23 (0.03, 0.43)	0.03	0.14 (0; 0.28)	0.05*	0.03
Self-regulation of learning in science	732 (361)	2.83 (2.78, 2.89)	639 (364)	2.77 (2.72, 2.83)	1,371 (732; 639)	0.26 (0.07, 0.45)	0.01	0.19 (0.03; 0.35)	0.02*	0.01
Self-efficacy for Working Scientifically	723 (370)	2.95 (2.9, 3.01)	628 (375)	2.91 (2.86, 2.97)	1,351 (723; 628)	0.15 (-0.04, 0.35)	0.14	0.11 (-0.03; 0.25)	0.14	0.11
Working Scientifically beliefs	706 (387)	3.27 (3.23, 3.32)	617 (386)	3.26 (3.21, 3.30)	1,323 (706; 617)	0.08 (-0.11, 0.27)	0.42	0.05 (-0.07; 0.17)	0.43	0.36
Wider benefits of science	725 (368)	3.36 (3.31, 3.41)	631 (372)	3.32 (3.27, 3.37)	1,356 (725; 631)	0.03 (-0.15, 0.21)	0.76	0.02 (-0.06; 0.09)	0.63	0.72

*Asterisks indicate statistical significance based on Romano-Wolf adjusted p-values: * p<0.05; ** p<0.01.*

Additional subgroup interaction effects

Further subgroup analyses examined whether participation in Focus4TAPS had differential effects across different pupil groups, as described in the 'Methods' section.

Table 21 presents the results of the subgroup interaction analyses for the primary outcome science attainment. Overall, the analysis did not find that the effect of Focus4TAPS differed across gender, pupils' pre-intervention exposure to science, or prior Key Stage 1 maths attainment. However, for pupils' pre-intervention interest in science, a small differential effect was observed. The associated Hedges's g effect size was -0.13 (95% CI: -0.23 to -0.04, p = 0.01), indicating that Focus4TAPS had a smaller impact for pupils who reported higher levels of interest in science at baseline.

Table 21: Subgroup interaction analysis for science attainment

Subgroup	Total n (Focus4TAPS; teaching as usual)	Intervention		Subgroup		Subgroup x intervention	
		Hedges' g (95% CI)	P-value	Hedges' g (95% CI)	P-value	Hedges' g (95% CI)	P-value
Gender (0 = Boys; 1 = Girls)	5,142 (2,496; 2,646)	0.05 (-0.10, 0.20)	0.554	-0.04 (-0.11, 0.02)	0.156	-0.03 (-0.12, 0.06)	0.49
Exposure to science (0 = Low exposure; 1 = High exposure)	4,187 (2,124; 2063)	0.00 (-0.15, 0.16)	0.957	0.02 (-0.04, 0.10)	0.421	0.05 (-0.05, 0.14)	0.36
KS1 Maths category (0 = Below expected standard; 1 = Expected standard or above)	5,142 (2,496; 2646)	0.00 (-0.16, 0.16)	0.996	0.64 (0.56, 0.72)	<0.001	0.04 (-0.06, 0.14)	0.40
Pupils' pre-interest in science (0 = Low interest; 1 = High interest)	4,333 (2,199; 2,134)	0.08 (-0.08, 0.24)	0.311	0.27 (0.20, 0.34)	<0.001	-0.13 (-0.23, -0.04)	0.01

Notes: Following the approach of the efficacy trial, effect sizes (Hedges' g) were calculated by dividing unstandardised regression coefficients and their CIs by the SD of the outcome measure (science attainment test scores; SD = 8.13).

In line with the EEF statistical guidance (EEF, 2022), restricted subgroup analyses were conducted to support this finding. For pupils with lower baseline interest in science, the estimated effect was small and positive (ES: 0.06, 95% CI: -0.11 to 0.24), while for pupils with higher baseline interest, the estimated effect was small and negative (ES: -0.08, 95% CI: -0.25 to 0.09). We therefore do not find conclusive evidence that Focus4TAPS improved science attainment within either subgroup.

Taken together, while the interaction model suggests that pupils' pre-intervention interest in science may have moderated the impact of Focus4TAPS, the restricted subgroup analyses do not show conclusive evidence that pre-intervention interest moderated science attainment outcomes. Full results are reported in Appendix D, Tables 2 to 5 in Further Appendices.

We also conducted subgroup analyses for secondary outcomes but did not find conclusive evidence that the effect of Focus4TAPS differed across the subgroups examined after adjustment for multiple comparisons. The full results are reported in Appendix F, Tables 6 to 13 in Further Appendices.

Analysis in the presence of non-compliance

Compliance with Focus4TAPS was defined at the school level. A school was considered compliant if the Year 5 teacher attended at least two of the three teacher training sessions and delivered a minimum of five Focus4TAPS assessment lessons by the end of the academic year. Table 22 summarises compliance with these criteria.

Overall, compliance with teacher training was high, with Year 5 teachers from 107 Focus4TAPS schools (93.04% of Focus4TAPS schools) attending at least two of the three training sessions. Information on lesson delivery was collected through the endline teacher survey, but response rates were low, with teachers from 39 Focus4TAPS schools (33.91%) completing the survey and reporting on assessment lesson delivery. Of these, 15 schools (13.04%) self-reported delivering at least five Focus4TAPS assessment lessons by the end of the year.

As specified in the 'Methods' section above, we treated non-response as non-compliance with the assessment lesson delivery criterion. As a result, 76 Focus4TAPS schools (66.09%) were considered non-compliant in relation to assessment lesson delivery due to missing data on this component. This approach was used to avoid counting schools as compliant in the absence of evidence that the minimum delivery requirement was met and may therefore lead to conservative estimates of compliance. However, some non-responding schools may have delivered the required lessons but not completed the endline survey, meaning true compliance may be higher and estimates of compliance effects should therefore be interpreted with some caution.

Table 22: School-level compliance with Focus4TAPS

Criteria	No. of schools that met criteria (%)
Attendance of teacher training	107 (93.04%)
Delivery of at least five Focus4TAPS lessons	15 (13.04%)

Across the compliance criteria, eight Focus4TAPS schools met neither criterion nor received a compliance score of 0. A further 92 Focus4TAPS schools met one of the two criteria and received a compliance score of 1, while 15 schools met both criteria and received a compliance score of 2. This indicated higher levels of complete non-compliance compared to the efficacy trial, where all Focus4TAPS schools met at least one compliance criterion.

Given limitations in measuring compliance, there is limited scope to assess whether higher levels of compliance with Focus4TAPS were associated with higher science attainment scores. Results were robust to a sensitivity analysis using an alternative compliance definition based on teacher training attendance only (criterion 1). Please see Appendix D, Tables 17 and 18 in Further Appendices for results from the CACE analyses.

Missing data analysis

From randomisation to analysis, there was a high level of pupil-level attrition, driven by non-completion of the science attainment, at 21.13% overall (21.95% in the Focus4TAPS group and 20.35% in the teaching as usual group).[23]

The missing data analysis considered the pupils in the primary outcome analysis sample. As Key Stage 1 maths and reading attainment measures were obtained through linkage to NPD records, pupils (and whole schools) who withdrew from the trial and did not provide consent for their data to be processed or linked to the NPD, could not be included in the missing data analysis, as no baseline Key Stage 1 data were available for them.

Among pupils who provided NPD consent, attrition from the NPD consent sample to the analysis sample was 19.03% (Table 13 in 'Attrition' section above), which exceeded the 5% threshold, below which patterns of missingness can typically be ignored. We estimated a series of drop-out models to explore potential patterns of missingness, as described in the 'Methods' section.

Missingness in the primary outcome was associated with several baseline characteristics. Pupils eligible for FSM were more likely to have missing science attainment scores compared to those not eligible for FSM. Missingness was also associated with Key Stage 1 attainment: pupils with missing Key Stage 1 maths data were more likely to have missing science attainment scores compared to pupils with Key Stage 1 maths attainment below the expected standard, while pupils with Key Stage 1 reading attainment at or above the expected standard were less likely to have missing scores compared to pupils with Key Stage 1 reading attainment below the expected standard. Higher baseline scores on the Working Scientifically subscale were also associated with a greater likelihood of missingness. Finally, missingness varied by school region, with pupils in several regions more likely to have missing science attainment scores.[24] No other covariates

[23] Full details are included in the 'Attrition' section of this report in Table 13.
[24] Pupils in Derby, Essex, Nottingham, South-east London, Sheffield, and Southampton were more likely to have missing science attainment scores compared to pupils in Bury.

significantly predicted missingness in the primary outcome. Coefficient estimates from this drop-out model are reported in Appendix E, Table 1 in Further Appendices.

Furthermore, missingness in the primary outcome model covariates was associated with FSM eligibility and baseline self-regulation of learning in science subscale scores. No other covariates significantly predicted missingness. Coefficient estimates from this drop-out model are reported in Appendix E, Table 2 in Further Appendices.

Given these patterns of missingness, we re-estimated the primary outcome model under alternative model specifications as a robustness check. We estimated an unadjusted model with no covariates and an adjusted model that included covariates associated with pupil drop-out, alongside randomisation strata and pupil age in months. Results from both alternative specifications were consistent with the primary analysis, indicating that the findings were consistent across the alternative model specifications.[25]

Overall, the results of the drop-out analyses were consistent with a MAR assumption conditional on the observed covariates. In line with the SAP and the EEF statistical guidance (Leonard *et al.,* 2025; EEF, 2022), we therefore re-estimated the primary analysis using multiple imputation. We conducted MICE, imputing missing values for the primary outcome and covariates included in the primary outcome model. The imputation model incorporated a wide set of pupil- and school-level variables, including treatment allocation, demographic characteristics, and prior attainment measures, to support the MAR assumption.[26] We generated 100 imputed datasets and re-estimated the primary outcome model across these datasets. The results are reported in Tables 23 and 24.

Table 23: Analysis of imputed datasets, by science attainment

Outcome	Unadjusted differences in means (95% CI)	Adjusted differences in means (95% CI)	Focus4TAPS		Teaching as usual		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Primary outcome: Science attainment	-0.11 (-0.90, 0.68)	0.15 (-0.88, 1.18)	3,059 (0)	68.529	3,292 (0)	65.44	66.93

Table 24: Analysis of the imputed datasets, GCSE science attainment effect size estimations

Outcome	Focus4TAPS n (missing)	Focus4TAPS unadjusted means (95% CI)	Teaching as usual n (missing)	Teaching as usual unadjusted means (95% CI)	Total n (Focus4TAPS; teaching as usual)	Hedges' g (95% CI)	P-value
Primary outcome: Science attainment	3,059 (0)	21.89 (21.58, 22.20)	3,292 (0)	22.04 (21.75, 22.33)	6,351 (3,059; 3,292)	0.02 (-0.11, 0.15)	0.78

The re-estimated results were consistent with those from the complete-case analysis. The adjusted difference in means was 0.15 (p = 0.78), compared with 0.24 in the complete-case analysis (Table 16 above). The corresponding Hedges' g effect size was 0.02 (95% CI: -0.11 to 0.15), compared with 0.03 in the complete-case analysis (Table 16 above), indicating a very small positive effect. As in the complete-case analysis, the CI includes both zero and negative values, and the results therefore remain consistent with not finding conclusive evidence that Focus4TAPS improves science attainment.

[25] In the unadjusted model, the estimated difference in mean science attainment scores was 0.04 (95% CI: -0.84 to 0.92; p = 0.93). In the adjusted robustness model, the estimated difference was 0.32 (95% CI: -0.79 to 1.44; p = 0.57), compared with 0.24 in the primary analysis model. In all cases, CIs included zero and negative values.

[26] The imputation model included treatment allocation and school Unique Reference Number (URN), together with a wider set of auxiliary variables used to inform the imputations: the randomisation strata variables region and school's prior level of Key Stage 1 attainment, age in months, Key Stage 1 maths and reading attainment, gender, pupil-level FSM eligibility, and baseline science subscale scores.

Additional analyses and robustness checks

Robust standard errors

A robustness check for the primary analysis was conducted using White (heteroskedasticity-robust) standard errors, which adjust for both clustering and heteroskedasticity in the data. We re-estimated the primary outcome model using these robust standard errors. With the effect estimates remaining the same as in the primary analysis, the robust standard errors resulted in more statistical uncertainty (p = 0.69, 95% CI: -0.89, 1.37). This is consistent with the primary analysis results and does not provide conclusive evidence that Focus4TAPS had an impact on science attainment. Please see Appendix D, Tables 6 and 7 in Further Appendices for full findings.

Imbalance at baseline

Where imbalance was indicated, we estimated an additional model as a sensitivity analysis that included all relevant pre-Focus4TAPS characteristics as covariates. As presented in the in the 'Pupil and school characteristics' section above, pupil-level FSM eligibility and school-level FSM eligibility, Key Stage 1 maths attainment, and Key Stage 1 reading attainment showed evidence of imbalance, with standardised mean differences exceeding 0.10 or effect sizes exceeding 0.05. We therefore included these measures as additional covariates in the primary outcome model to assess whether the main findings were robust to adjustment for baseline imbalance.

In the sensitivity model, the adjusted difference in means was 0.59 (p = 0.33). The corresponding Hedges' g effect size was 0.07 (95% CI: -0.07 to 0.21), indicating a small positive effect that was slightly larger than that observed in the primary analysis (0.03; Table 16 above). However, as in the primary analysis, the 95% CI includes both zero and negative values, and the results therefore remain such that we do not find conclusive evidence that Focus4TAPS improves science attainment. Full results from the sensitivity analysis are presented in Appendix D, Tables 8 and 9 in Further Appendices.

Alternative baseline specification

A robustness check was conducted by re-estimating the primary outcome model using an alternative baseline specification. In this model, Key Stage 1 reading and maths attainment were included as categorical variables (BLW, PFK, WTS, EXS, and GDS), in alignment with the approach used in the efficacy trial. The adjusted difference in means was 0.48 (p = 0.40). The corresponding Hedges' g effect size was 0.06 (95% CI: -0.08 to 0.19), indicating a small positive effect that was slightly larger than that observed in the primary analysis (0.03; Table 16 above). However, the CI overlapped with that from the primary analysis and included zero; we therefore do not find conclusive evidence that Focus4TAPS had an effect on science attainment. Please see Appendix D, Tables 11 and 12 in Further Appendices for these results.

Combined specification

As a final robustness check, we re-estimated the primary outcome model combining both the alternative baseline specification and adjustment for the pre-Focus4TAPS characteristics that were imbalanced at baseline.[27] The adjusted difference in means was 0.71 (p = 0.24) and the corresponding Hedges' g effect size was 0.09 (95% CI: -0.05 to 0.23). This effect estimate was larger than that observed in the primary analysis (0.03; Table 16 above). The CI, however, included zero, and the overall conclusion remained unchanged, with no conclusive evidence that Focus4TAPS improved science attainment. Please see Appendix D, Tables 13 and 14 in Further Appendices for these results.

The same model was also estimated for the FSM subgroup. The adjusted difference in means was 1.20 (p = 0.15) and the corresponding Hedges' g effect size was 0.15 (95% CI: -0.04 to 0.34), which was larger than that observed for the pre-specified FSM subgroup analysis (0.07; Table 19 above). The CI overlapped with that from the pre-specified FSM subgroup analysis and included zero; we therefore did not find conclusive evidence that Focus4TAPS improved science attainment for the FSM subgroup. Please see Appendix D, Tables 15 and 16 in Further Appendices for full findings.

[27] Key Stage 1 reading and maths categories were transformed into numeric working scales (1 = 'BLW', 2 = 'PKF', 3 = 'WTS', 4 = 'EXS', 5 = 'GDS') to estimate the individual-level correlations between the baseline attainment measures and endline science attainment. The observed correlations were moderate (reading: r = 0.576, p<0.001; maths: r = 0.585, p<0.001).

Implementation and Process Evaluation results

Summary of IPE findings

- Focus4TAPS teacher training was delivered broadly as intended. Evidence from training observations, trainer interviews, and attendance data showed that trainers followed the planned sequence and training materials closely, and the three full-day workshops were well-attended by teachers.

- When delivering Focus4TAPS in schools, our lesson observations and teachers' own lesson reflections showed that teachers generally selected Focused Assessment plans, focused lessons on a single Working Scientifically element, and used formative assessment in line with the intended model, adapting activities when needed for their pupils.

- Among the teachers who completed the endline teacher survey, just over half (54%, n=28) reported delivering the minimum five Focused Assessment lessons, while others reported delivering fewer lessons, most commonly citing difficulty fitting lessons in with their school's or MAT's curriculum and assessment style, or a lack of capacity to deviate from the existing curriculum. However, given the low completion rate of the survey, this finding should be interpreted with caution. The reasons provided for not delivering the intended number of Focused Assessment lessons were echoed in teacher focus groups.

- Teachers responded positively to the training and resources. They valued the practical activities, clear modelling, and structured materials. In the survey, teachers reported increased confidence in formative assessment and Working Scientifically.

- Lesson observations, pupil focus groups, and teacher reflections showed that pupils planned parts of investigations and recorded their own results. Teachers also reported consistently high-pupil interest in Working Scientifically lessons and increased engagement with science more broadly. Teachers described pupils as more confident and more aware of why they were doing particular activities. They also mentioned that the practical activities in the lessons were more inclusive than usual class work, which helped pupils who were less confident in science or had Special Educational Needs and Disabilities (SEND) to participate and feel involved.

- Survey data and focus groups indicated that one view among teachers was that Focus4TAPS improved pupil-led enquiry and formative assessment. Another view was that aspects of the approach resembled their existing practice, with some schools already using TAPS resources, which reduced the distinctiveness of the programme in those schools.

Differentiation from usual practice

This section outlines the typical approaches to science teaching and formative assessment in teaching as usual schools prior to and during the delivery of the Focus4TAPS programme. It also explores usual practice pre-intervention in the Focus4TAPS schools. Understanding the difference between usual practice and the Focus4TAPS programme helps us interpret the contribution that the programme made to the observed outcomes from the impact evaluation and the perceived wider outcomes. This section examines the following question:

- IPE RQ1: (a) What is 'teaching as usual' and how does Focus4TAPS differ?

Overall, the data showed that usual practice in both teaching as usual and Focus4TAPS schools was characterised by broadly stable patterns of weekly science lessons, similar lesson lengths, established schemes of work, and variable use of practical work and formative assessment. Prior to the intervention, intervention schools resembled control schools in terms of structure, resources, and support. While elements of formative assessment and practical work were present in usual practice, they were not consistent across schools.

Usual practice in teaching as usual schools

This section describes usual practice in the control schools prior to the intervention period, and whether those schools changed their practice during the trial.

We asked about teaching as usual schools' usual practice in the Year 5 teachers' baseline survey and the programme developer's pre-intervention interview. We asked about these schools' practice during the intervention period as part of the endline survey of teachers and science leads, to enable assessment of any change over time that was not associated with the Focus4TAPS programme.

Lesson frequency and length

The majority of teaching as usual teachers (88%, n=85) reported in the baseline survey that they taught science lessons once a week (base n=97). More frequent lessons were reported by 8% (n=8); 1% (n=1) said they taught less than once a week. The remaining 3% (n=3) said the frequency of their teaching differed between weeks.

Broadly, this pattern was replicated in the endline data (base n=84),[28] where the majority of teaching as usual group teachers and science leads again reported that science lessons were delivered once a week during the intervention period. At endline, 79% (n=66) reported that science lessons were typically weekly. More frequent delivery was reported by 11% (n=9); the smallest proportion (4%, n=3) said they taught science less than once a week, and the proportion whose lesson frequency varied between weeks sat between these two groups at 7% (n=6).

In the baseline survey, the majority of teachers in teaching as usual schools reported that their science lessons typically lasted 60 or 90 minutes—37% (n=36) said their usual science lessons generally lasted 60 minutes; 37% (n=36) reported lessons of around 90 minutes. Much smaller proportions said their lessons were longer than 90 minutes (18%, n=17) or around 45 minutes (8%, n=8).

Endline survey data indicated that control group schools' practice was broadly consistent during the Focus4TAPS intervention period: again, the majority reported that lessons lasted 60 minutes (37%, n=31) or 90 minutes (36%, n=30). Around a quarter delivered lessons of more than 90 minutes (25%, n=21). Only a very small minority of respondents reported that usual science lessons lasted 45 minutes (2%, n=2).

Lesson structure and planning

In the baseline survey, all teaching as usual Year 5 teachers reported that they used resources and schemes for their science planning and teaching. The range of resources they used included: a scheme of work provided by their schools (58%, n=56); online resources from organisations like PSTT or STEM Learning (53%, n=51); a scheme of work from their MATs (7%, n=7); and other resources including a mix of multiple schemes (26%, n=25).[29] It is worth noting that TAPS resources are developed by and accessed via PSTT. As such, teachers who reported that they were accessing online resources from organisations like PSTT or STEM Learning might already have been using TAPS-based resources.

Four specific resources that respondents mentioned under other resources were:[30]

- **Pan-London Assessment Network (PLAN) primary science resources.** Support schools to plan and assess the science National Curriculum. PLAN focuses on skills like eliciting pupils' prior knowledge, teaching new concepts and skills, checking understanding, and revisiting learning.[31]

- **Explorify.** A website run by STEM Learning that provides free online activities that aim to develop pupils' critical thinking skills, discussion skills, curiosity, and creativity. [32]

[28] This base refers to the number of responses rather than individual respondents, because some respondents held dual roles as both Year 5 teachers and science leads. This meant they submitted an answer to both the Year 5 teacher and science lead versions of the usual practice questions in the endline survey.
[29] This was a multi-coded question where respondents could select more than one response.
[30] Descriptions of the advertised aim/content of these schemes have been added by the research team
[31] See: www.planassessment.com/introduction
[32] See: www.stem.org.uk/explorify/top-tips

- **White Rose.** Provides schemes for learning that teachers could use to plan and teach science content aligned with the National Curriculum, including Working Scientifically skills.[33]

- **Oak National Academy.** Free teaching and lesson resources for science, structured by topics that aim to build pupils' knowledge progressively.[34]

The baseline survey also asked what support teaching as usual teachers received to support their science teaching. A small minority (8%, n=8) of Year 5 teachers reported having no existing forms of support for their science teaching. Of those who received support, the most common sources included:

- support from their school's science lead (65%, n=62);
- specific resources for teaching and assessing science (29%, n=28);
- support from their senior leadership team (SLT) (26%, n=25);
- support from other Year 5 teachers (20%, n=19);
- in-house training for teaching and assessing science (20%, n=19);
- external training for teaching and assessing science (18%, n=17).

In the endline survey, teaching as usual group teachers and science leads reported that they most often planned and taught science using their school's own schemes of work (56%, n=47), online resources from organisations such as PSTT (which includes TAPS resources) or STEM Learning (51%, n=43), other external resources (39%, n=33), or their MAT's schemes of work (10%, n=8).[35]

Practicals

Practical work in science refers to hands-on science activities where children actively investigate, observe, and use equipment to explore scientific concepts and processes. For ease of understanding, we asked the teaching as usual respondents to answer this question in the context of 'using practical equipment within their usual science lessons'. In the baseline survey, 72% (n=70) of teaching as usual teachers reported that they used practical equipment in their usual science lessons a few times in every science topic or unit, and 16% (n=16) used practical equipment in every science lesson. Around 5% (n=5) said they used practical equipment in every science topic or unit, and 1% (n=1) once per term. A small proportion of Year 5 teachers (5%, n=5) reported varying the frequency of using practical equipment and explained that they used them more or less often depending on the specific science topic.

Responses to this question were broadly similar at the endline survey stage when we asked teaching as usual group teachers and science leads whether and how often they had used practical science equipment during the intervention period. The majority (73%, n=61) of respondents reported that they used practical equipment a few times in every science unit or topic, while others used practical equipment in every science lesson (18%, n=15), once in every science unit or topic (6%, n=5), or at a varying frequency (4%, n=3).[36]

Assessment approaches

The programme developer told us in their interview that schools commonly equated science assessments with formal testing, and that assessment tended to be implemented inconsistently. In their experience, where assessment was undertaken, many schools used commercial tests that did not fully align with the National Curriculum or pupils' needs.

[33] See: **https://whiteroseeducation.com/resources/science/primary**
[34] **See: www.thenational.academy/teachers/programmes/science-primary-ks2/units**
[35] It is worth noting that TAPS resources are also developed and accessed via PSTT.
[36] Science leads who indicated that they did not teach Year 5 pupils were not asked this question.

Assessment results were also not systematically collected or analysed to identify which concepts and skills pupils had understood.

In the baseline survey, we asked teachers to reflect on formative assessments (defined in the survey as 'a way of assessing pupils' needs and adapting teaching to pupils' level of understanding in science lessons'). Just under a third (27%, n=26) of Year 5 teachers in teaching as usual schools reported that they did not use a set approach for formative assessments in science lessons. Of those teachers who used formative assessments, 58% (n=56) reported that they used an approach developed by their school or MAT and 16% (n=15) used TAPS resources from PSTT.[37] A minority also reported using Assessment for Learning in Science resources from STEM Learning (6%, n=6).

A similar pattern emerged in the endline teacher survey, where this question was asked of teachers and science leads who taught Year 5 pupils. About a third (35%, n=29) of teaching as usual Year 5 teachers and science leads reported that they did not use formative assessment in their science lessons during the intervention period. Of the 65% (n=61) of teaching as usual Year 5 teachers and science leads who used them, the most common formative assessment approaches were those developed in their school or MAT (56%, n=47). Equal proportions said they used resources from STEM Learning (8%, n=7) and TAPS resources from PSTT (8%, n=7).

Year 5 teachers who used formative assessment in their usual practice were also asked how frequently they did so. At the baseline stage, almost two-thirds (61%, n=43) used formative assessment in every science lesson and around a quarter (24%, n=17) used it a few times in every science topic or unit. Formative assessment was used once in every science topic or unit by 11% (n=8), and 3% (n=2) only used it once per term.

Endline survey respondents reported using formative assessment every science lesson (42%, n=23), a few times in every science unit or topic (36%, n=20), or once in every science unit or topic (18%, n=10) during the intervention period. A very small minority reported using formative assessment once per term (2%, n=1), or at a varying frequency (2%, n=1).

Patterns of lesson frequency, lesson length, use of practical work, and formative assessment did not indicate any systematic changes over time in teaching as usual schools. Overall, evidence from baseline and endline surveys suggested that usual practice in teaching as usual schools remained broadly stable across the intervention period.

Usual practice in Focus4TAPS schools prior to the intervention

This section outlines baseline science teaching practices in Focus4TAPS schools prior to the intervention, including lesson length, frequency, structure, use of practical work, and assessment approaches. The data suggests that most teachers taught science weekly in lessons lasting 60–90 minutes, drawing on a wide range of existing schemes of work and resources that were broadly similar to those used in teaching as usual schools. While practical activities and formative assessment were present in usual practice, their frequency, purpose, and implementation varied considerably. This section also sets out similarities and differences between usual practice and the Focus4TAPS approach.

Lesson length

In the baseline survey (base n=101), the majority of teachers in Focus4TAPS schools reported that their science lessons typically lasted 60 or 90 minutes—44% (n=44) said their usual science lessons generally lasted 60 minutes; 33% (n=33) reported lessons around 90 minutes. Much smaller proportions said their lessons were longer than 90 minutes (17%, n=17) or around 45 minutes (7%, n=7).

Lesson frequency

The majority of teachers (90%, n=91) reported that they taught science lessons once a week. More frequent lessons were reported by 4% (n=4) of teachers, while 4% (n=4) said they taught less than once a week. The remaining 2% (n=2) said the frequency of their teaching differed between weeks.

[37] This is the model on which Focus4TAPS was based.

Lesson structure and planning

The range of resources used in intervention schools prior to Focus4TAPS was similar to those used in teaching as usual group schools. They included: a scheme of work provided by their schools (56%, n=57); online resources from organisations like PSTT (which includes TAPS resources) or STEM Learning (55%, n=56); a scheme of work from their MATs (10%, n=10); and other resources including a mix of multiple schemes (26%, n=26).[38] Other resources mentioned included PLAN primary science resources, Explorify, Oak National Academy, Plymouth Science Schemes, and Twinkl.

Intervention teachers identified a similar range of support they received for their science teaching to control group schools. Again, a small minority (9%, n=9) of Year 5 teachers reported having no existing forms of support for their science teaching. Of those who received support, the most common sources included:

- support from their school's science lead (67%, n=68);
- support from other Year 5 teachers (33%, n=33);
- specific resources for teaching and assessing science (32%, n=32);
- in-house training for teaching and assessing science (22%, n=22);
- external training for teaching and assessing science (18%, n=18); and
- support from their SLT (17%, n=17).

Practicals

In line with teaching as usual schools, the majority of Focus4TAPS group teachers (65%, n=66) said that they taught practical science lessons a few times in every science topic or unit, and 14% (n=14) taught practical science in every science lesson. Around 11% (n=11) said they taught one practical in every science topic or unit, and 1% (n=1) once per term. A small proportion of Year 5 teachers (9%, n=9) reported varying the frequency of practical science lessons and explained that they taught them more or less often depending on the specific science topic.

Assessment approaches

A third (33%, n=33) of Year 5 teachers in the Focus4TAPS schools reported that they did not use a set approach for formative assessments in science lessons prior to the Focus4TAPS intervention. Of those teachers who used formative assessments, 55% (n=56) reported that they used an approach developed by their school or MAT and 13% (n=13) used TAPS resources from PSTT.[39] A minority also reported using Assessment for Learning in Science resources from STEM Learning (9%, n=9).

At the baseline stage, almost half (49%, n=33) of those intervention schoolteachers who did use formative assessment said they did so in every science lesson. More than a quarter (28%, n=19) used it a few times in every science topic or unit. Formative assessment was used once in every science topic or unit by 22% (n=15), and 1% (n=1) only used it once per term. This was comparable with responses in control group schools at baseline.

In teacher focus groups and trainer interviews, Focus4TAPS teachers reported mixed levels of experience and familiarity with formative assessment. Some found the Focus4TAPS approach to formative assessment to be the same as their usual assessment approach since they were already using some of the programme's elements in their teaching; others reported that formative assessment was new to them.

[38] This was a multi-coded question where respondents could select more than one response.
[39] This is the model on which Focus4TAPS was based.

How Focus4TAPS differs from usual practice

Focus4TAPS differed from usual practice as it focused more strongly on formative assessment and use of practical work. These distinctions are central to understanding how Focus4TAPS may have contributed to the outcomes observed in the impact evaluation and IPE.

Similarities and differences in lesson planning

According to the programme developer, usual science lessons covered a specific topic and were delivered in a linear way with topics addressed in turn without revisiting the concepts and skills explored in earlier lessons. However, pupils in focus groups recounted that most of their usual science lessons used examples and concepts from previous lessons.

By contrast, the developer mentioned that the Focus4TAPS model encouraged teachers to adapt pre-made lesson plans based on how pupils responded to and understood previous lessons. Focus4TAPS promoted flexibility by encouraging teachers to 'pause, observe, revisit, and reframe lessons' to target specific skills in pupils.

Year 5 teachers and science leads also identified several specific schemes of work that they had already used before the Focus4TAPS intervention. Some schemes incorporated elements similar to Focus4TAPS, particularly in supporting teachers' understanding of pupils' scientific thinking. For example, PLAN primary science resources included materials designed to develop pupils' Working Scientifically skills and provided examples of pupil work. Similar to Focus4TAPS, PLAN encouraged teachers to use formative assessment during lessons to adapt teaching in response to pupils' needs. Teachers in the focus groups also described Explorify as effective for encouraging whole-class reflection and helping pupils become more comfortable sharing their thoughts.

Despite some overlaps, the schemes teaching as usual teachers used, lacked some features that were central to Focus4TAPS. For instance, PLAN, White Rose, and Oak National Academy did not emphasise focused recording, self-assessment, or peer assessment; provide a structured framework to embed formative assessment practices consistently across classrooms and at whole-school level (the TAPS pyramid model in Focus4TAPS; or include an explicit emphasis on formative assessment). In addition, Oak National Academy did not offer resources that specifically targeted the development of pupils' Working Scientifically knowledge and skills.

Use of practical work

In the pre-Focus4TAPS interview, the programme developer shared the view that standard teaching practice tended not to make effective use of practical work in science, as its potential to engage pupils and support understanding was often misunderstood or undervalued. Rather than offering opportunities to explore concepts or develop specific scientific skills, practical activities tended to be teacher-led, where experiments were carried out by the teachers or by pupils following a fixed set of instructions, and pupils' involvement focused on written work (e.g. writing up methods, results, and conclusions for every experiment).

In focus groups, we asked pupils to reflect on their usual science lessons before Focus4TAPS was implemented in their schools. Pupils mentioned that usual science lessons were delivered in a range of ways: examples included watching a video to learn about a specific topic; taking notes of a teacher-led presentation; and learning from their textbook. Pupils also discussed working in groups and carrying out experiments in their usual lessons prior to Focus4TAPS.

In focus groups, some pupils reported that most of their usual science lessons (prior to Focus4TAPS) did not involve practical work. When practical activities did occur, pupils described a structured and teacher-led format. Typically, either in the same lesson or a prior lesson, the teacher first explained the key content, often using PowerPoint slides. The teacher then walked the class through the practical activity, before pupils carried out the practical themselves. This contrasted with Focus4TAPS lessons, where practical work was more central to the lesson and pupils reported more hands-on practical activities.

Focusing on one element and recording

The programme developer described that Focus4TAPS encouraged teachers to focus on a single element of enquiry and use practical work for formative assessment. Some Focus4TAPS teachers mentioned that their usual practical lessons were similar to this and already focused on one element of Working Scientifically (i.e. Plan, Do, or Review).

Fidelity

This section examines the following question:

- IPE RQ1: How is the Focus4TAPS programme delivered, and what are the key facilitators and barriers for delivery?

Under the broad dimension of fidelity, we cover the extent to which Focus4TAPS was delivered as intended and the extent to which changes were made to the delivery model. We also comment on the quality of the intervention's training activities and workshops. We link our data to the sub-questions that fall under IPE RQ1.

Overall, the evidence indicated that Focus4TAPS was delivered with fidelity to the intended delivery model across trainer training, teacher training workshops, and implementation in lessons. Training workshops were delivered consistently across regions, with trainers adhering closely to the planned structure, content, and sequencing. Trainers made permitted, responsive adaptations to support engagement and understanding among teachers. Teachers generally implemented Focus4TAPS as intended, using Focused Assessment lesson plans, focusing on individual elements of Working Scientifically, and drawing on formative assessment to adapt teaching and support pupil learning. Where adaptations occurred, these were aligned with programme guidance, and reflective of pupil-level needs rather than departures from the intended model.

It is important to note that delivering a minimum of five Focused Assessment lessons is also a component of the intended delivery model. This section focuses primarily on whether the core pedagogical components of Focus4TAPS were implemented as intended when lessons did take place. Patterns in lesson dosage are examined in detail in the 'Dosage' section below.

Implementation fidelity

Teacher training workshops

Provision of three training days to teachers is a key input set out in the programme logic model. Training was provided by Focus4TAPS trainers (themselves trained by the programme developer). This section aims to answer the following question related to teacher training, drawing on data from our observations of training sessions and qualitative interviews with trainers.

- IPE RQ1: (c) Was Focus4TAPS training implemented as intended by trainers?

Our observations of training delivery and interview discussions with trainers found that Focus4TAPS training was largely implemented as intended by trainers.[40] Across regions, trainers adhered closely to the planned structure, content, and sequencing of the training workshops, using the resources and guidance provided by the programme developer. While trainers exercised some flexibility in the selection of the activities included in each workshop, this was consistent with the intended training model.

[40] More information on the total number of training sessions observed and the stages of delivery at which data were collected can be found in the 'Methods' section.

Structure and format of the training workshops

The delivery of teacher training workshops closely followed the structure and format set out in the Focus4TAPS model. We observed that each training workshop comprised four main elements:

- **Trainer-led presentations.** Using the presentation slides and speaking notes provided, trainers delivered key information to the training group. Examples included presentations on key elements of the Working Scientifically curriculum, use of focused recording in science lessons, and about how to undertake Focused Assessments.

- **Practical activities.** Opportunities to try out some of the practical activities teachers could deliver in classrooms as part of Focus4TAPS lessons.

- **Group discussions.** Whole-group and small group discussions where teachers reflected on their own teaching practice and shared their reflections and experiences with others. In the second and third workshops, this included discussion of teachers' reflections on the Focused Assessment lessons they had delivered in classrooms.

- **Agreeing 'gap tasks'.** At the end of the first and second workshop, trainers asked teachers to put their learning into practice in classrooms before the next workshop by delivering lessons focused on Working Scientifically skills using Focused Assessment lesson plans.

Overall, we observed that there was a consistent structure across all teacher training workshops. Each followed a logical, sequential structure aligned with the prepared PowerPoint slides presented by the trainers. Practical activities and handouts were prepared in advance and used throughout the day. The workshops all included a mix of trainer-led presentations, group discussions, practical activities, and opportunities for teachers to reflect on their own practice. Teachers' views on the usefulness of the training are discussed under IPE RQ2(c) in the 'Responsiveness' section below.

Content of the training workshops

Observation data indicated that trainers across all regions aligned content delivery closely with Bath Spa University training guidance. Across observed workshops, trainers followed the intended sequencing of content.

Across observed workshops, trainers consistently covered the intended content for each session and clearly linked individual concepts and activities to the Focus4TAPS approach. Trainers revisited key content across workshops, supporting teachers to develop their understanding over time, in line with the programme model.

Practical activities

Observation and interview data indicated that trainers implemented the practical activities in line with the intended Focus4TAPS approach. Trainers selected activities from a list provided by the programme developer, allowing them to tailor delivery to their specific groups. In interviews, trainers described this flexibility as important for ensuring that activities resonated with teachers and their individual contexts.

Teachers mentioned that the practical activities were useful for them to start thinking about how they could teach Working Scientifically skills, and to model and try out different methods for encouraging more independent work for pupils.

Across regions, trainers consistently incorporated multiple hands-on activities within each session, tailoring their selection to each group. Appendix M, Table 1 in Further Appendices, describes the practical activities delivered across the three workshops we observed.

Overall, trainers delivered practical activities in a way that aligned with the intended Focus4TAPS training model. Trainers used flexibility in activity selection and format to tailor delivery to teachers' needs.

Resources and materials

Observation and interview data indicated that trainers used resources and materials as intended. The programme developer designed the training to be supported by resources that trainers would use during workshops and that teachers could later

apply in their own classrooms. Across observed workshops, trainers provided teachers with these materials, including training workshop slides, handouts explaining the practical activities, a TAPS booklet folder with the TAPS pyramid model on it, the TAPS pyramid model self-evaluation tool, and a handout specifying different Focused Assessment lesson plans for different Working Scientifically topics.

Trainers also used online resources as intended during the observed workshops. In the first workshop, they showed teachers how to access the PSTT website where they could find the TAPS lesson plans, examples of pupil work, guidance booklets, and the TAPS pyramid model. In the third workshop, trainers introduced additional resources and how to access them:

- **TAPS resources.** This included the Focused Assessment lesson plans, searchable on the PSTT website, with lessons and examples of pupil work linked directly to Working Scientifically.
- **Stoke exemplification for primary science.** A new resource that supported teachers to make end of Key Stages 1 and 2 judgements.
- **PLAN.** Resources provided a range of activities and examples of pupil work for all lessons, covering different science topics.

This section examines the following question:

- IPE RQ1 (d) What were the drivers and obstacles to implementation of the intervention/training, and was this affected by region/school context?

Observation data and trainer interviews highlighted that several features supported the implementation of the training. These included the train-the-trainer model, structured but flexible workshop design, staged delivery across the school year, and the use of practical activities that allowed teachers to engage directly with the Focus4TAPS approach. Trainers and teachers valued the balance of presentations, discussions, and hands-on activities, as well as the manageable pacing of content and opportunities for reflection between workshops.

Trainers also identified some obstacles to the delivery. Where challenges arose, these related primarily to logistical factors such as venue constraints or differences in teachers' prior experience of teaching primary science, which trainers addressed by making adaptations. There was limited evidence that implementation varied by region, with adaptations more commonly driven by the experience of teacher groups rather than school or regional context.

Teachers' views on the implementation of the intervention and the training are discussed in the 'Responsiveness' section below.

Drivers to training implementation

Observation data and trainer interviews highlighted the following drivers to implementation of the training:

1. **Train-the-trainer model.** In interviews, trainers consistently described the training delivered by Bath Spa University and the programme developer as high quality and well-paced. They valued the detailed explanations given by the programme developer, the centralised and well-organised resources and materials, and the opportunity to discuss, share best practice, and learn from their peers who would be delivering in other regions.

   > *It really gave you complete confidence, actually, in terms of how to deliver it. …it was just covered in such a lot of detail that you did leave that training feeling like you had everything you needed to deliver it.* (Trainer 5, Trainer interview)

   They appreciated the opportunity to deliver the training in the pre-trial phase as it helped them practise and adequately prepare their delivery of the training. The programme developer echoed this view in their pre-intervention interview, saying that it was beneficial for trainers to gain this experience so that they can deliver the workshops in an effective way.

2. **Balanced workshop structure.** In their interviews, trainers commented that teachers found it helpful to have a balance of different elements in each workshop. Exploring a range of different practical activities, for example, allowed teachers to consider which they were more interested in or would be more applicable to their school. Trainers said group discussions offered useful opportunities for teachers to share reflections and experiences with their peers, which teachers would not often otherwise have time or opportunity to do.

   > *I think the teachers that came to our training seemed really grateful for that time to talk [...] People took pride, in fact, in bringing back what they'd been doing in school and sharing it and that emphasis on, 'This doesn't have to be perfect, just bring what's happening in your classroom'.* (Trainer 1, Trainer interview)

3. **Staged delivery over time.** Trainers also felt it was beneficial to deliver training in stages, with teachers implementing lessons in their classrooms between the workshops. Key benefits of this structure that were identified in trainer interview data were:

   - Provision of a manageable amount of information, building over time. This included, for example, focusing on teacher-level, pupil-level, and school-wide elements of the model in turn.
   - Greater opportunities to learn and develop from reflective practice and feedback on experiences in the classroom.
   - Better sharing of reflections and experiences, supported by rapport and comfort developed between group members over the course of the training programme. Trainers reported that as teachers built supportive working relationships with their training groups, willingness to discuss their own work increased, facilitating richer discussions of experiences with Focused Assessment lessons, and reflections on their own teaching practice.

4. **Flexible delivery within the intended framework.** Trainers reported that they adapted the timing and emphasis of activities within workshops to focus on what they thought would be most helpful for their group of teachers. The two main reasons for this were:

   - *To support engagement and attention.* Some trainers found it useful to begin workshops with a practical activity, which they felt 'jumpstarted' teachers' engagement. Others preferred to start with trainer-led presentations of theory and key concepts in the morning when teachers were more focused.
   - *To support comprehension and utility.* Trainers tailored the time spent on particular content based on teachers' prior experience and knowledge of primary science pedagogy. For example, trainers extended the time spent exploring how Focused Assessment lessons could focus on a single 'Working Scientifically' element when teachers initially struggled to understand how this would work in practice. In contrast, trainers spent less time covering types of scientific enquiry when this content was already familiar to their group of teachers.

5. **Accessible, well-pitched content and resources.** Trainers and teachers consistently reported that the Focus4TAPS training content was pitched at an appropriate level, which reduced the need for adaptations to the content itself. A large majority of teachers (98%, n=51) 'agree' that the content and activities were pitched at the right level (77% [n=40] 'strongly agree', 21% [n=11] 'agree', and only 2% [n=1] 'neither agree nor disagree').

   Trainers and teachers in interviews and focus groups agreed with this view, describing the content as accessible, relevant, and well suited to their roles. Teachers particularly valued the explicit links made to the National Curriculum, as these helped them understand how Focus4TAPS could be integrated into current lesson plans.

   According to the programme developer, all resources were designed to be accessible, free to use, and easily applied in school settings, which reduced practical barriers to implementation. Trainers echoed this view and highlighted that the materials provided to them for the delivery of the workshops, and those given to teachers to support their implementation of the programme, were detailed, well-organised, and required minimal set-up. Trainers did not report adapting the resources in any way.

Obstacles to training implementation

The trainer interview data and observation data identified some obstacles to implementing the Focus4TAPS training. Where challenges did arise, these were primarily practical in nature and did not prevent delivery of the training as intended.

One challenge related to variation in teachers' prior knowledge and experience of primary science pedagogy. Trainers mentioned that in some training groups, teachers initially struggled to understand how Focused Assessment lessons could focus on a single element of Working Scientifically. This required trainers to spend additional time on certain concepts to support understanding, which occasionally affected the pacing of workshops.

Observations indicated that venue size influenced how some practical activities were implemented, such as limiting opportunities to rotate teachers during carousel-style activities. In addition, time constraints occasionally led trainers to adjust the planned structure of sessions, for example, by omitting an afternoon carousel activity in order to prioritise other elements of the training.

Another challenge related to interpretation of some lesson plan resources. Trainers reported that certain lesson plans were more difficult for teachers to understand without first experiencing them as practical activities. As a result, trainers sometimes emphasised hands-on exploration of these resources before teachers could confidently apply them in their own classrooms.

Overall, these obstacles reflected contextual and logistical considerations, and trainers were generally able to address them through flexible delivery.

Regional and school context

There was no strong evidence that implementation of the Focus4TAPS training varied by region. Across regions, we observed that the structure and delivery of the training workshops were consistent, with trainers following the same core programme design. Where adaptations were mentioned in trainer interviews, these were driven primarily by the composition of training groups rather than by regional or school-level context. In particular, trainers adjusted delivery in response to teachers' prior experience and familiarity with primary science. This flexibility was used to support engagement and understanding and did not reflect any differences linked to geography or region.

Implementing Focus4TAPS in schools

This section examines the extent to which teachers implemented Focus4TAPS in line with the intended delivery model. We assessed adherence to delivery of Focus4TAPS in schools through the teacher endline survey, Focus4TAPS lesson observations, teacher lesson reflections from the second workshop training survey, and the teacher focus groups. Our analysis compared these data to the delivery model outlined in the protocol and against the intended activities described in the logic model. The section aims to answer the following questions:

- IPE RQ1 (b) Was Focus4TAPS implemented by teachers as intended?

This section focuses on how teachers delivered the core components of Focus4TAPS in the lessons they taught. It is important to note, however, that among teachers who responded to the endline survey, just over half (54%, n=28) reported delivering the intended minimum of five Focused Assessment lessons. This is discussed further in the 'Dosage' section below.

Overall, the evidence suggests that, when they were delivered, Focus4TAPS lessons were largely implemented by teachers in line with the intended delivery model. Most teachers reported focusing on a single element of Working Scientifically in each Focused Assessment lesson and using formative assessment to plan, reflect on, and adapt their science teaching. Observation, survey, and interview data indicated that teachers generally followed Focused Assessment lesson plans and used formative assessment to adjust the level of support and challenge within lessons, consistent with the programme's logic model. Teachers also implemented a range of strategies to support pupil engagement in scientific enquiry, including self-assessment and, to a lesser extent, peer assessment. This was reflected in the training surveys, where 96% (n=102) of Year 5 respondents described their pupils' engagement during a specific Focused Assessment lesson, commonly reporting

high levels of engagement and participation. Where adaptations were made, these were aligned with guidance that encouraged flexibility within the Focus4TAPS approach.

IPE RQ1 (b) (i) Did teachers focus on one element of Working Scientifically in each lesson?

The logic model suggested that teachers must focus on one element of Working Scientifically in each lesson. The training guidance recommended that teachers should aim to cover different elements of the Working Scientifically wheel but there was no strict guidance around covering all the elements.

The majority of endline survey respondents reported having focused on a single Working Scientifically element in each of the Focused Assessment lessons they delivered (Table 25). All Year 5 teacher respondents (100%, n=23) 'agree' with this statement, with the majority (65%, n=15) saying they 'strongly agree'. Among science leads, 84% (n=30) 'agree' with this statement, while 6% (n=2) 'neither agree nor disagree', 3% (n=1) 'disagree', and 8% (n=3) reported that they 'Did not deliver any Focused Assessments to my class'. This difference could be due to the lower levels of direct involvement in classroom delivery by science leads.

Table 25: Whether each Focused Assessment lesson focused on one element of Working Scientifically

	Year 5 teachers	Science leads
Strongly agree	65%	53%
Agree	35%	31%
Neither agree nor disagree	0%	6%
Disagree	0%	3%
Strongly disagree	0%	0%
Did not deliver any Focused Assessments to my class	0%	8%
Base	23	36

Endline survey data also showed variation in which elements of Working Scientifically teachers chose to focus on. The most common elements of Working Scientifically that Year 5 teachers delivered were:

- observe and measure (91%, n=21);
- ask questions and plan enquiry (87%, n=20);
- record (78%, n=18);
- setting up enquiry (70%, n=16);
- evaluate (57%, n=13); and
- interpret and report (57%, n=13).

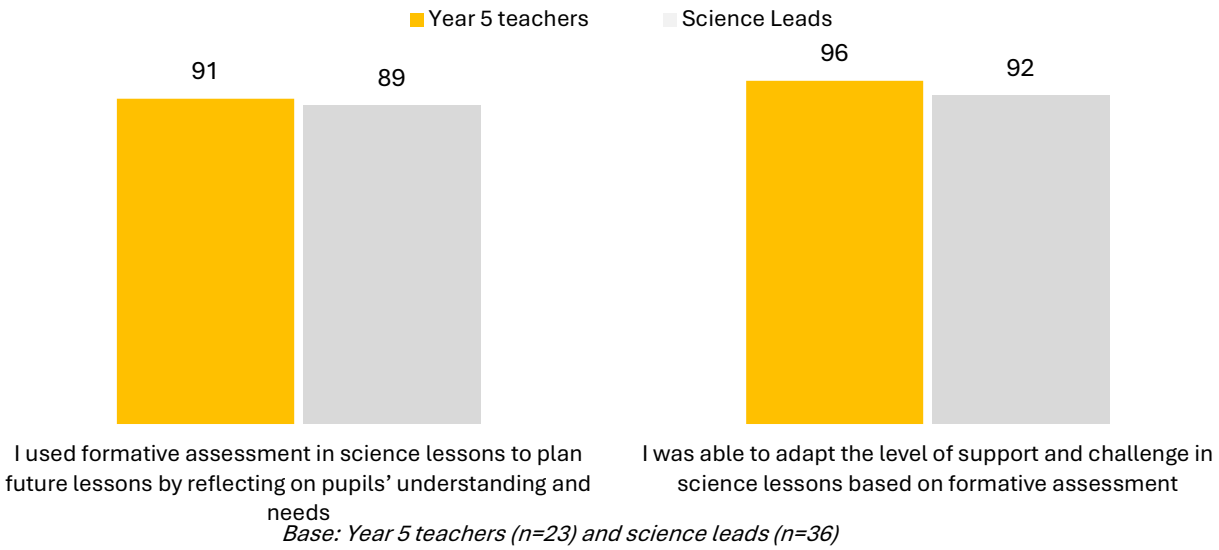
A similar pattern emerged among science leads, who also most frequently reported focusing on elements within the 'Plan' and 'Do' stages rather than 'Review'.

This pattern was reflected in the lesson reflections and training survey data, where more teachers reported trying out lessons focused on 'Plan' or 'Do' elements than on 'Review'. All Focused Assessment lessons that we observed focused on 'Do'. Teachers selected which element to teach during the lesson observations, and both trainers and teachers suggested that teachers often perceived 'Do' lessons as more engaging and practical, which may have influenced these choices.

IPE RQ1 (b) (ii) Did teachers plan and reflect on formative assessment within science lessons?

The Focus4TAPS logic model set out that teachers must follow and use Focused Assessment lesson plans in their science lessons. The endline survey asked Focus4TAPS teachers and science leads about their use of formative assessment in science lessons. Large majorities 'agree' or 'strongly agree' with the statement: 'I used formative assessment in science lessons to plan future lessons by reflecting on pupils' understanding and needs', with very similar levels among teachers (91%, n=21) and science leads (89%, n=32). This result is presented in Figure 8.

Figure 8:6 Use of formative assessment in science lessons (agreed), %



Base: Year 5 teachers (n=23) and science leads (n=36)

Evidence from teacher focus groups and lesson reflections indicated a good level of adherence to this expectation. In focus groups, Focus4TAPS teachers reported that they had, as intended, accessed the Focused Assessment lesson plans provided on the PSTT website. Teachers reported selecting lessons that aligned with their school curriculum, current topics, or particular Working Scientifically elements or enquiry types.

Lesson reflection data and the training surveys from Year 5 teachers and science leads showed that teachers used a wide range of Focused Assessment lesson plans. Common examples included lessons on forces (such as spinners, parachutes, and zip wires), dissolving, space, water filtration, and seasonal change. Teachers explained that they selected these lessons either to fit with planned curriculum coverage or to focus on developing a specific Working Scientifically skill.

A more mixed picture was evident in our observation data. Five of the six lessons that we observed at case study schools closely followed Focused Assessment lesson plans. In the remaining school, we observed a lesson that did not do so. The school explained that they had merged Focus4TAPS strategies into existing lesson plans provided as part of the scheme for science planning and teaching used at their school. This aligned with guidance provided during the training, which encouraged teachers to integrate Focus4TAPS principles into existing schemes where appropriate, and therefore represented a planned adaptation of intended implementation in schools.

IPE RQ1 (b) (iii) To what extent were teachers able to adapt the level of support and challenge in lessons based on their formative assessment?

The logic model indicated that teachers must also adapt the level of support and challenge based on formative assessment. In the lesson reflections data, Focus4TAPS Year 5 teachers and science leads described using formative assessment to reflect and plan their future science lessons. In Day 2 training surveys, 93% (n=99) of Year 5 teachers described adapting a specific lesson in response to their pupils' needs. Examples of adaptations they made in response to assessment results included delivering follow-up lessons to address misconceptions; and additional demonstrations and modelling of scientific concepts to help firm up pupils' knowledge and skills where assessments indicated further support might be needed.

Respondents were also asked to reflect on the statement: 'I was able to adapt the level of support and challenge in science lessons based on formative assessment'. Almost all Year 5 teachers (96%, n=22) and science leads (92%, n=33) 'agree' or 'strongly agree' with this statement (Figure 8).

Observation data provided examples of teachers using formative assessment to adapt the level of support within lessons, in line with the Focus4TAPS model. For example, in one Focused Assessment lesson we observed, pupils did not initially carry out the investigation systematically. In response, the teacher prompted pupils to reflect on what they had done well and less well and supported them to evaluate how they could improve their approach. This adaptation reflected the intended use of formative assessment to adjust support and deepen understanding.

In interviews, teachers described a shift in how they responded to pupils' understanding during science lessons, in line with the Focus4TAPS emphasis on formative assessment. Teachers reported revisiting the same content, reflecting on how lessons had gone, and using assessment information to decide whether further explanation was needed.

> *So, that has given us more of a focus, and actually, it's fine that they don't fully get it on the first one. So, we reiterate, we reiterate, and we assess, and that has worked really well.* (Teacher 1, Teacher focus group)

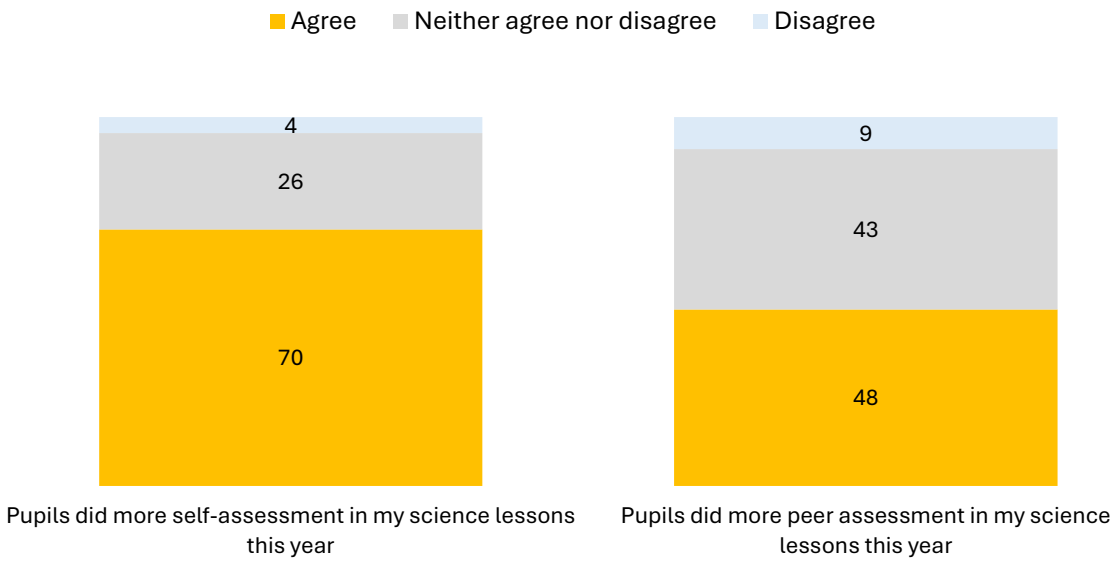
IPE RQ1 (b) (iv) To what extent were students involved in more self-assessment and peer assessment compared to teaching as usual?

Finally, the logic model suggested that teachers should involve pupils in more self-assessment and peer assessment.

We observed pupils doing peer assessment or self-assessment in our lesson observations. For example, we observed teachers asking pupils to reflect on the practical lesson they had just completed and identify what worked well and less well and what they could do next time to get more accurate results. Focus4TAPS teachers in focus groups also described using peer assessment and self-assessment in their science lessons.

In the endline survey, we asked Focus4TAPS Year 5 teachers and science leads to indicate their agreement with the statement: 'Pupils did more self-assessment in my science lessons this year'. Among Year 5 teachers, who were most likely to be delivering Focus4TAPS lessons directly, 70% (n=16) 'agree' with this statement, indicating increased use of pupil self-assessment in their science teaching (Figure 9). Just over half of science leads (53%, n=19) 'agree'. However, science leads were less likely to be teaching Focused Assessment lessons themselves, meaning that this question might have been less directly applicable to their role.

Figure 9:7 Use of self-assessment and peer assessment in science lessons, %



Base: Year 5 teachers (n=23)

Increases in the use of peer assessment were more mixed. A little less than half (48%, n=11) of Focus4TAPS Year 5 teacher respondents 'agree' with the statement: 'Pupils did more peer assessment in my science lessons this year', while 43% (n=10) 'neither agree nor disagree', and 9% (n=2) 'disagree' (Figure 9).

In the focus group discussions, one view among teachers was that they had focused their peer assessment on written feedback, rather than also considering broader forms of assessment such as verbal feedback between peers. Further, some participants said that they had not used peer assessment as it was not appropriate for some of their pupil groups, such as those who struggled to work with peers rather than independently due to additional needs.

Use of other Focus4TAPS strategies in lessons

Focus group and lesson observation data also showed that teachers implemented other Focus4TAPS strategies that aimed to encourage a more active role for pupils in science lessons. Examples included pupils recording data and results themselves, and involvement in planning their enquiries (e.g. choosing variables to investigate or contributing ideas to planning board displays). These practices were broadly aligned with the intended Focus4TAPS approach, which emphasised pupil engagement in the enquiry process.

Adaptations to the Focus4TAPS approach

The observations and lesson reflections data showed that teachers made small adaptations to the Focus4TAPS approach to tailor their teaching to pupils' needs. It is important to note that such changes were permitted and encouraged within the delivery model to help make the intervention more effective for pupils.

In the lesson reflections data, Focus4TAPS Year 5 teachers and science leads described adapting Focused Assessment lessons in response to pupils' needs. These adaptations included:

- **Providing additional teacher time and support for SEND or English as an Additional Language (EAL) pupils.** To help them access lesson instructions and scientific vocabulary. For example, conducting experiments one-on-one outside of science lessons with pupils who needed additional support or who struggled to work independently.

- **Giving pupils structured recording templates.** With examples included in the templates so that pupils can refer to them while recording their own answers.

- **Using planning boards in other lessons.** To support pupils in organising their ideas before writing or recording outcomes.

- **Including teacher modelling.** Where teachers demonstrated how practical activities were meant to be conducted and gave additional examples.

Another adaptation was the way in which teachers made use of the Working Scientifically wheel in their classrooms. The Focus4TAPS teacher training guidance set out that the wheel was for teachers to use to support coverage or to refer to within lessons, rather than specifying that it needed to be used directly with pupils. However, our teacher focus groups and lesson observations found that teachers did share the Working Scientifically wheel directly with pupils in a range of ways. While some teachers referred explicitly to the wheel during Focused Assessment lessons, for example, others had pictures of it on the classroom wall, and/or also pasted in pupils' workbooks.

Train-the-trainer model

Another key input of the logic model is trainer training which was delivered by the programme developer to the Focus4TAPS trainers. The section aims to answer the following questions:

- IPE RQ1 (e) How well was the Focus4TAPS developer able to monitor fidelity and quality of the training programme?

Evidence from trainer interviews indicated that the Focus4TAPS programme developer was able to monitor the fidelity and quality of the training programme through a combination of standardised materials, structured support, and ongoing

communication with trainers. Although there was no formal fidelity monitoring framework associated with the train-the-trainer model, trainers reported a shared understanding of the core content and pedagogical approach of Focus4TAPS. This was supported by standardised materials and modelling via videos.

Ongoing quality assurance was further supported through regular opportunities for interaction between the programme developer and trainers. These included:

- **Pre-workshop drop-in webinars.** These were hosted by the programme developer before each of the teacher training workshops. Trainers reported that these offered a helpful opportunity to ask any questions, confirm logistical arrangements for workshops, and refresh their knowledge about the training content.

- **Post-workshop feedback meetings.** These gave trainers opportunities to reflect on workshop delivery, share tips and tricks with each other, and discuss any improvements that could be made.

- **Regular emails from the programme developer.** With reminders and guidance about communicating with their teacher training participants. These emails included anonymous feedback that teachers had provided via on-the-day training surveys.

- **Videos.** Trainers could access online training videos that recapped the key messages for each teacher training workshop, including demonstrations of the practical activities.

In interviews, trainers described that they had focused on content preparation, such as reviewing materials and jointly planning sessions with training partners, and practical preparation related to organising resources and logistics. The programme developer provided additional support to trainers who were new to delivering Focus4TAPS through observations of experienced trainers or joint planning, helping to build confidence and consistency across the trainer cohort.

Together, these mechanisms enabled the programme developer to maintain oversight of training delivery while allowing trainers flexibility in how sessions were facilitated.

- IPE RQ1 (f) How scalable is the Focus4TAPS developer's approach to quality assurance of the training programme?

In terms of scalability, the evidence suggested that the Focus4TAPS developer's approach to quality assurance was well suited to delivery at scale. We observed that the use of standardised materials, recorded training videos, and structured but flexible support mechanisms allowed consistency across multiple regions without requiring direct oversight of every training session. In their interviews, trainers reported that the training model supported consistency while also enabling them to adapt delivery to their own style and the needs of their training groups.

Additionally, in interviews, trainers identified several features of the approach to training and supporting trainers that enhanced their engagement. These broadly related to:

- **Collaborative ways of working.** Trainers valued opportunities to work closely with both the programme developer and other trainers. They said that the pre-meetings and debriefs that were offered as a part of their initial training provided space to ask questions, reflect on what worked well or less well during teacher training delivery, and revisit key content for upcoming workshops. Trainers also enjoyed connecting with other trainers in these meetings, and sharing tips and discussing best practice as a part of their debriefs. They described the programme developer as supportive and respectful of their input. For some trainers, involvement was further enhanced by the opportunity to work with a programmer developer who was well regarded within the primary science community.

- **Accessibility and convenience.** Practical arrangements were also seen as important in facilitating engagement with the trainer training workshops and wider support. Participation was made more accessible through flexible scheduling of online meetings, with polls to check and accommodate trainers' availability as much as possible, as well as sharing recordings for those unable to attend sessions. Trainers also appreciated that travel expenses were covered for central training sessions.

These views suggested that the quality assurance approach could be sustained as the programme expanded if these key facilitators of the train-the-trainer model were kept in place.

Quality

To understand how well the teacher training workshops and Focused Assessments were delivered, data on quality was collected through training workshop observations, interviews with trainers and the programme developer, and lesson observations.

In summary, the Focus4TAPS programme was delivered to a high standard across the teacher training workshops, and lessons. Trainers were well prepared and demonstrated strong facilitation skills, creating supportive and well-managed training environments. Most teachers showed a good understanding of the Focus4TAPS approach and integrated Focused Assessments effectively into their lessons. However, lesson observations indicated some variation in the quality of teachers' explanations of scientific concepts, suggesting that while many teachers were confident in this area, others required additional support.

Teacher training workshops

In the three teacher training workshops we observed, trainers demonstrated very strong facilitation skills.[41] The training delivery was interactive, responsive, and encouraging. Overall, we observed that trainers created a comfortable atmosphere where teachers were able to share their views and experiences with the group.

We observed that trainers in all regions used plain English and accessible language when describing scientific and pedagogical concepts (e.g. 'activating pupils' prior knowledge' was described as 'exploring what children already know').

We also observed that trainers managed their time effectively during the delivery of the training workshops. They allowed time for questions while ensuring they covered all the training slides. Trainer interviews and training observations both indicated that trainers sometimes provided one-on-one support to teachers during breaks to ensure that the scheduled sessions stayed on track that the timing of the different elements was not derailed due to some teachers needing additional support.

A key facilitator for high-quality delivery of the workshops was having trainers deliver the workshops with a partner (or partners). The programme developer's approach to pairing trainers was to bring together an experienced primary science CPD provider and a current or former teacher in each pair. In interviews, trainers said this supported effective delivery by:

- **Maximising expertise.** Each trainer brought their unique experience, skills, and expertise. As a result of the pairing approach, one trainer usually had a greater grasp on the theory and concepts behind Focus4TAPS, while the other could give realistic advice about applying Focus4TAPS in a school environment. Delivery as a pair meant that both aspects could be comprehensively addressed during training delivery.

- **Division of tasks to streamline delivery.** Trainer pairs split the preparation work between them. Additionally, when one trainer presented content to the teachers, the other checked the PowerPoint slides and presenter notes to ensure nothing was missed and also prepared the practical kit for the next activity together.

Delivery of Focused Assessment lessons

The lesson observations found that teachers used the Focus4TAPS approach in their lessons.[42] We observed that they checked pupils' understanding of key scientific concepts and vocabulary. Teachers used targeted questioning, practical activities, and pupil responses to build their lessons. The delivery of lessons was well-structured, with teachers using specific examples from the lesson to demonstrate the impact of Focused Assessment on pupil learning. Overall, the quality of reflections indicated a good understanding among teachers of using Focused Assessments. The analysis of the lesson

[41] We observed three in-person training sessions, delivered by different trainers and in different regions at separate stages of the programme. Further information about the observation approach is provided in the 'Methods' section.
[42] We observed six lesson observations in six different schools. More information on this approach is provided in the 'Methods' section.

reflections supported this view, and teachers detailed different ways in which they integrated Focused Assessments in their lessons to help support pupils' needs.

We observed that scientific concepts were explained inconsistently across lessons. Teacher lesson observations showed that, in some cases, teachers gave little or no explanation of the scientific principles behind the experiment before pupils began. Others provided only brief guidance, such as limited time spent clarifying the planning board slide for practicals or key variables in the experiment. In contrast, some teachers demonstrated strong practice by giving clear instructions and explaining both the rationale and objectives of the investigation. This variation showed that while some teachers were confident in explaining scientific concepts, others struggled to explain the concepts effectively when giving practical work to the pupils.

Reach

This section examines the following questions:

- IPE RQ2 (a) Did the intervention reach its intended population?
- IPE RQ2 (b) Are there patterns in terms of pupil characteristics for those who were/were not reached by the intervention?

Our data indicated that Focus4TAPS largely reached its intended population at the school and teacher levels, and that lessons were delivered to whole classes as intended.

Around 115 schools were allocated to the Focus4TAPS group, and attendance data showed that all but one engaged with the programme by attending at least one of the teacher training workshops. For most schools, Year 5 teachers and science leads were nominated and attended the training sessions as intended.

Table 26 shows the number of staff members and number of schools who attended each day of the training session.

Table 26: Attendance per training workshop

Training sessions	No. of staff members who attended	No. of schools who attended	% of schools who attended (total=115)
Day 1	214	114	99%
Day 2	203	112	97%
Day 3	199	108	94%

Almost all intervention group teachers and schools attended the three in-person training workshops. However, attendance across the three sessions varied between schools. Where schools missed training sessions, reasons for absence that were recorded in the administrative data included staff sickness, staff leave, a lack of lesson cover, or teachers having to attend other training. In some cases, schools did not have a dedicated Year 5 teacher in post, and other staff attended training in their place. Where staff were unable to attend the workshops (which built on one another sequentially) or did not have opportunities to deliver lessons to Year 5 between workshops, this might have affected classroom delivery and outcomes.

Focus4TAPS was intended to be delivered as a whole-class activity. In our observations of lesson delivery, teachers delivered Focused Assessment lessons to their whole class, as intended.

Data on participation at an individual pupil level was not collected as this would have placed significant evaluation burden on classroom teachers. As such, it is not possible to analyse reach by pupil characteristics such as gender, SEND, or socio-economic background. Data from lesson observations and teacher focus groups, however, suggested that Focus4TAPS was delivered as a whole-class programme, as intended. There was no evidence from observations or qualitative data that particular groups of pupils were selectively included or excluded from participation in Focused Assessment lessons.

Focus group data also indicated that some teachers were using aspects of Focus4TAPS in year groups beyond Year 5. Teachers gave examples of lessons with pupils in Year 2 and Year 4. In part, this is because some of the science leads and teachers who took part were directly involved in lesson delivery to other year groups and did not themselves teach Year 5 pupils.[43]

Dosage

This section examines the following questions:

- IPE RQ1 (g) How many Focused Assessment lessons were delivered in intervention schools?
- IPE RQ1 (h) What was the amount of time spent delivering the Focused Assessment lessons?

This section explores how much of the intervention was delivered and whether this was in line with the intended dosage. Data on dosage was collected in two ways: as part of observations of teacher training delivery; and in the survey of teachers and science leads in the intervention group.

Teachers in intervention schools were expected to attend three full-day training sessions (one in each school term) and to deliver a minimum of five Focused Assessment lessons. These requirements formed the core delivery structure of the programme and were outlined in the training guidance provided by Bath Spa University.

Teacher training dosage

All three in-person training workshops were delivered as intended across 13 regional cohorts. As intended, each workshop comprised a full-day's training: those we observed were all delivered across a full six-hour day and generally finished on time. Teacher attendance data is discussed in the 'Reach' section above.

Classroom dosage

Taken together, the data on intervention dosage suggests that pupils' exposure to Focus4TAPS varied across intervention schools, with teachers delivering varied numbers of lessons of different lengths.

Number of lessons

The Day 2 training survey data indicated that a high number of teachers were beginning to implement Focused Assessment lessons, with 99% (n=105) of Year 5 respondents reporting that they had tried at least one lesson by January – February 2025.

In the endline survey, we asked how many lessons teachers had delivered over the course of the programme. It is important to note that these data were limited: the endline survey was answered by only 52 of the Focus4TAPS group teachers, spread across 42 Focus4TAPS group schools, meaning data was not available for the other 61 Focus4TAPS schools. It is not possible, therefore, to comment on classroom dosage across the intervention group as a whole. As such, the findings regarding lesson delivery should be interpreted with caution and understood as relating to the subset of schools for which dosage data were collected.

Nonetheless, the endline survey data suggests that, in practice, a substantial proportion of teachers were unable to deliver the intervention at the expected level. Just over half of teachers (54%, n=28) who completed the endline survey reported that they had delivered five or more lessons, which meant that they had met or exceeded the minimum intended dosage. Just over a quarter (27%, n=14) reported that they delivered four lessons; the remaining fifth (19%, n=10) reported that they had delivered three lessons or fewer.

[43] One of the focus groups we conducted with school staff was comprised of science leads and staff who taught year groups other than Year 5. This could explain why Focus4TAPS approaches were used with other pupils and why experiences with year groups other than Year 5 were discussed.

Factors influencing lower-than-intended dosage were also captured in the endline survey data. The most common reasons Year 5 teachers selected when reporting why they had delivered fewer than five minimum Focused Assessment lessons were that:

- 'It was difficult to fit the lessons in with my school's or MAT's curriculum and assessment style' (63%, n=5);
- 'I didn't have the capacity to deviate from the existing curriculum' (25%, n=2);
- 'Lack of support from senior leadership' (13%, n=1); or
- 'Other' reasons (25%, n=2).

A small proportion of Year 5 teachers reported that they did not know the reasons behind not delivering the intended number of lessons (13%, n=1).

In some of our observations of training delivery, we noted that trainers did not always explicitly state the minimum requirement for lesson delivery. This is likely to have played a part in lower numbers of lessons being delivered by some teachers.

Length of lessons

The logic model did not specify an intended length of delivery for the Focused Assessment lessons. In the endline survey, a range of lesson lengths for Focused Assessment lessons were reported. Most teachers delivered lessons lasting either 60 minutes (50%, n=26) or 90 minutes (31%, n=16). Smaller numbers reported longer or shorter lessons: Around 10% (n=5) delivered lessons more than 90 minutes in length; 4% (n=2) lasted around 45 minutes. Finally, 6% (n=3) of respondents, all of whom were science leads, reported that they did not deliver any lessons.[44]

Responsiveness

This section examines the following question:

- IPE RQ2: How engaged are teachers and pupils in Focus4TAPS, and how is engagement affected by context, teacher characteristics, and pupil characteristics?

This section explores how teachers and pupils engaged with Focus4TAPS, using data from the endline survey, lesson and training observations, and teacher and pupil focus groups.

To summarise, teachers showed high engagement with Focus4TAPS. Teachers found the training useful and took part actively in workshops. Engagement was supported by helpful trainers, clear resources and opportunities to work with other teachers, while challenges mainly related to workload, staffing pressures, and school priorities.

Pupils were generally engaged in Focus4TAPS lessons, especially during practical and group activities. Engagement was less consistent during written tasks. There was no evidence of differences by pupil characteristics. Overall, Focus4TAPS supported positive engagement in science lessons.

Teacher engagement

Teacher engagement with Focus4TAPS training

We first consider teachers' engagement with the Focus4TAPS training programme, exploring how actively teachers participated in the training and how useful they perceived it to be. This section explores findings on teacher engagement

[44] The percentages here are greater than 100% due to rounding.

during the Focus4TAPS training based on the training observations, lesson observations, trainer interviews, survey data from teachers, and teacher focus groups. This section aims to answer the following question:

- IPE RQ2 (c) How engaged were teachers with the training programme and how useful did they find it?

Teachers had a positive view towards the usefulness of the training workshops. When asked about this in the endline survey, the majority of teachers and science leads (96%, n=50) stated that they 'strongly agree' or 'agree' that the Focus4TAPS training was useful while only 4% (n=2) stated that they 'neither agree nor disagree'. No respondents 'disagree' with the statement.

Science leads and teachers had positive views towards the training content. All Focus4TAPS science leads 'strongly agree' (69%, n=25) or 'agree' (31%, n=11) with the following statement: 'the Focus4TAPS training made it clear how Focus4TAPS strategies could be implemented in my school'.

Teachers were also very engaged in delivering Focus4TAPS and most Year 5 teachers and science leads (96%, n=22) reported in the endline survey that their experience with delivering was 'excellent' or 'good', while only 4% (n=1) rated their experience as 'fair'.

In training observations, teachers showed consistent engagement across all three training workshops. They participated enthusiastically in practical activities, group discussions, and reflective tasks. In the second workshop, we observed that discussions continued into break times, with teachers showing interest in the training. During the final workshop, teachers appeared confident and comfortable in sharing their views with peers. However, in some sessions, some teachers seemed to lead the group discussions more than others.

In interviews, trainers similarly reported that teachers were highly involved in all activities and demonstrated a strong willingness to contribute, ask questions, and provide reflections during the training.

Trainers mentioned that the programme included structured activities to scaffold thinking and build engagement gradually. For example, they made teachers start with low-stakes tasks on Day 1 of the training and moved to higher-stakes tasks like sharing pupil work. Training survey responses highlighted that teachers enjoyed practical activities and valued the mix of discussion, theory, and hands-on tasks spread across all sessions.

> *So, we break the activities down, so it becomes more manageable, so, 'There are all of these resources here, but which one do you want to start with first?' So, it's that bigger picture into a more focused, manageable way forward for them in their school, and they really appreciate that.* (Trainer 6, Trainer interview)

One view among teachers was that the workshops and resources included 'lots of information to absorb', which felt overwhelming. There were instances where teachers mentioned that they remained confused about key concepts, such as the difference between focused recording and focused planning even after attending the training.

Another view among trainers was that teachers sometimes became very absorbed in certain group activities during the training and needed prompting to move on to the next task.

Trainers also noted that teachers grew more confident and comfortable in discussions as the training sessions progressed. By the second and third day of the training, teachers contributed more freely and appeared at ease when sharing ideas with peers and willingness to share pupil work was noticeably higher in later training sessions when compared to the first training session. Additionally, trainers described that the 'gap tasks' that teachers had to complete between each workshop (e.g. trying out a Focused Assessment lesson) were pitched at the right level and easy for teachers to try out.

We also observed that teachers engaged with the training content. Teachers took notes during presentations and worked through topics such as formative assessment and summative assessment. All teachers mentioned that they had trialled Focused Assessment lessons and they shared examples of pupil work to share and reflect on with peers during the training.

During training observations, teachers reflected on how to adapt lesson plans to meet pupils' needs. These reflections often focused on practical adjustments they could make in their own classrooms and ways to tailor activities so that all pupils could access the science content.

We observed that teachers asked questions about choosing Focused Assessment lessons, adapting plans, using photos for assessment, and reporting pupil outcomes. Such questions indicated that teachers were thinking ahead to how they would apply the approaches in practice.

Additionally, teachers with prior TAPS experience shared their insights and helped others understand how to use the materials. Examples included explaining how they had previously integrated assessment strategies and offering tips to make the strategies work in different contexts. This showed that teacher engagement during the training supported practical uptake of Focus4TAPS approaches among peers.

Teacher engagement with Focus4TAPS lesson resources

Alongside engagement with the training itself, teachers' experiences were shaped by their interactions with the Focus4TAPS lesson resources, which are explored in the next section. This section aims to answer the following question:

- IPE RQ2 (d) How did teachers feel about the intervention lesson resources?

Teachers and trainers praised the quality and accessibility of the Focus4TAPS lesson resources. In focus groups, teachers reported that the resources were straightforward to use. They praised materials such as planning boards, lesson plans, and the TAPS pyramid model, and said these resources made it easier to deliver lessons consistently. In interviews, trainers also noted that the practical activities were flexible and could be carried out with everyday materials, which helped teachers include them in regular classroom practice.

Teachers in focus groups also mentioned that the resources helped them use different approaches in lessons and supported pupils with varied needs. They highlighted that clear activities and optional scaffolds allowed pupils to work at their own level, while giving teachers flexibility to adapt tasks. One example of the effectiveness of these resources was the life cycle activity. The activity included brief instructions, and suggested ways in which pupils could demonstrate their learning in creative ways. One view among teachers was that such straightforward lesson plans encouraged them to give pupils more freedom, while the scaffold remained useful for those who needed it.

In interviews, trainers reported that some teachers had faced challenges in engaging with the Focus4TAPS resources. Trainers mentioned that some teachers had provided feedback that the Focus4TAPS website was difficult to use. This made it harder for them to engage with the lesson materials and activities.

Costs and benefits of involvement in Focus4TAPS

Teachers' views of the training and resources informed their broader perceptions of the costs and benefits of participating in Focus4TAPS. This section answers the following question:

- IPE RQ2 (e) What were the perceived costs and benefits of being involved in the intervention?

Benefits of being involved in Focus4TAPS

In focus groups, teachers identified a range of benefits associated with participating in Focus4TAPS. Teachers reported increased confidence in teaching science, particularly in relation to using formative assessment and developing pupils' Working Scientifically skills.

Teachers also valued the professional learning opportunities provided through the training, including developing a clearer understanding of formative assessment and how it could be applied within science lessons.

Teachers in focus groups reported that engaging with Focus4TAPS improved their understanding of the Working Scientifically curriculum, and some described seeing positive changes in pupil engagement during science lessons, which increased their motivation to continue using the approach.

Teachers also valued opportunities to talk with other teachers and schools. In interviews, trainers mentioned that when multiple staff from the same area attended the training, they were able to build a shared understanding and exchange ideas. Teachers described the opportunity to engage with other schools as enjoyable and supportive for their practice. They stated that they often helped other teachers make sense of tasks during the training. Teachers also found it particularly useful to compare their pupils' workbooks with those from nearby schools so that they could learn more about different ways of assessment from their peers. Trainers also mentioned that collaborative learning worked best when both science leads and Year 5 teachers attended training together, creating space for joint planning and discussion.

Costs of being involved in Focus4TAPS

In focus groups, teachers also described some perceived drawbacks or challenges associated with participation in Focus4TAPS. Some felt it was challenging to commit the time required to attend training and implement Focused Assessment lessons within an already crowded curriculum. Teachers noted that curriculum pressures and competing priorities sometimes limited their capacity to engage fully with Focus4TAPS.

Linked to this, another view among teachers in focus groups was that Focus4TAPS contributed to workload challenges which could result in feelings of pressure and stress. Some teachers reported that juggling their existing responsibilities while trying new approaches in lessons could be stressful. In interviews, trainers mentioned that when reflecting on how they felt about training workshops, one view among teachers was that sharing examples of their own practice and pupils' work with peers could cause concern about being judged negatively.

Drivers and challenges to teachers' engagement

In this section, we explore the key drivers and challenges shaping teachers' engagement with the Focus4TAPS training and then examine how these factors influenced engagement during implementation of the intervention in schools. This links to the following question:

- IPE RQ2 (f) What were the drivers/challenges to teachers' engagement with the training and the intervention?

Drivers to teachers' engagement with the training

Trainer interviews, teacher observations, teacher focus groups, and endline survey responses highlighted five factors that supported teacher engagement with the Focus4TAPS training. These included:

1. **Training mode.** Teachers reported that face-to-face interaction helped them feel comfortable and form peer relationships with other teachers. They stated that this was something that they had found harder to achieve through other online CPD.

2. **Accessible venues.** In interviews, trainers reported that teachers also had a positive experience with the venues at which training took place, highlighting that the venues were in good and accessible locations.

3. **Structure of the training.** Trainers mentioned that each training session built on the last, and the three-day structure gave teachers time to reflect, experiment, and develop a deeper understanding of the programme. Additionally, teachers mentioned that it was useful for them to have training structure where they could start using the resources immediately after the training.

4. **Delivery by trainers.** Teachers described trainers as knowledgeable, enthusiastic, and motivational. They also mentioned that trainers helped them understand the challenges that they would face during delivery. This approach made teachers feel valued and created space for open discussion. Teachers appreciated trainers' helpful and engaging style.

   In addition, teachers stated that having two trainers with varied school experience was beneficial, and this 'tag-team approach' allowed one trainer to present while the other trainer answered questions. Teachers mentioned that trainers worked towards creating a welcoming atmosphere by introducing attendees, reassuring latecomers, and ensuring no one sat alone during the activities.

In interviews, trainers described the sessions as a safe space without judgement, where teachers felt comfortable to discuss what went well and challenges to using Focus4TAPS. Training observations showed that trainers facilitated one-on-one conversations about school-specific needs, and personal anecdotes were used to make content feel relatable to the teachers.

Trainers mentioned that they used personal anecdotes to flesh out theoretical concepts such as formative assessment and to provide examples of how teachers could apply these concepts in practice. Trainers with teaching experience also said they drew on this to illustrate concepts and ground them in examples that would be more familiar to teachers.

5. **Adapting training to teachers' experiences.** Trainers highlighted that the content and delivery of the training workshops enabled teachers with varying experience and skills to 'get something out of the training'. Those with less experience learned something new, while those with more experience revisited and further embedded what they already knew. Trainers managed this variation in experience by giving more basic explanations of concepts and ideas when needed, or if teachers were already familiar with something (like the five types of enquiries), they would spend less time going over it in detail.

Challenges for teachers' engagement with the training

While several factors supported engagement with the training, trainers and teachers also identified challenges that limited participation and uptake. These challenges fell into two broad categories: barriers linked to school contexts; and challenges related to the design and delivery of the training itself.

School-based challenges included:

1. **Attendance and staff cover issues.** Trainers reported that attendance was a recurring challenge during the training. They identified three main reasons for this: last-minute cover difficulties in schools; personal circumstances affecting individual teachers; and clashes with Ofsted inspections. These factors led to late cancellations or partial attendance, occasionally disrupting continuity for teachers and limiting peer learning during the training.

2. **Lack of SLT support or permission.** Trainers mentioned that resistance from senior leaders at schools was the biggest barrier to the training. Without SLT approval, teachers struggled to prioritise attending the training or make changes in their classrooms.

3. **Teacher workload and stress.** Heavy workloads and stress were common barriers, which limited teachers' focus during sessions. However, informal chats during the training helped teachers decompress and share concerns.

> *Yes, the noise that goes on outside of what we were doing. Sometimes you had to almost facilitate that to be offloaded first. That's why the coffee room was really helpful. We'd be like, 'how are things? How is it since the last time we've seen you?' …Sometimes you can just read people's faces and you just know that you should make a beeline for them and go and chat to them, and I think that helped.* (Trainer 7, Trainer interview)

Training-related challenges included:

1. **Fear of judgement from peers.** One view among teachers was that they were hesitant to share pupil work, particularly during the initial training, as they were concerned their work would be criticised or compared during the training.

2. **Preference for individual learning.** Some teachers preferred one-on-one learning and found group discussions challenging. This preference sometimes reduced participation in collaborative tasks during the training.

3. **Group dynamics.** Trainers observed that some groups were louder and more dominant in discussions, while others were quieter. In order to help with this, trainers considered adjusting seating arrangements to create more balanced and productive group discussions.

Drivers of engagement with programme delivery

This section explores drivers to teachers' engagement in delivering Focus4TAPS lessons. These were very similar to facilitators to engagement with the training.

First, school leadership and support played a key role in teacher engagement with delivering Focus4TAPS. According to teachers, SLT buy-in and enthusiastic science leads created a supportive environment for teachers to try new approaches and prioritise science within the curriculum. Trainers also mentioned that schools that were already invested in science CPD (e.g. through programmes like the Ogden Trust or PSQM) showed high levels of engagement. This evidence was supported by the programme developer interview, who highlighted SLT permission to use Focus4TAPS methods as a factor which helped implement the programme in schools.

Second, peer engagement and ongoing communication supported teachers in delivering Focus4TAPS in their classrooms. Teachers mentioned that they shared feedback with each other, which helped generate new ideas and practical solutions.

Finally, trainer support and reassurance helped teachers feel confident in using Focus4TAPS within their classrooms. Teachers mentioned that this reassurance reduced anxiety and encouraged them to persist with the new methods, even if early results were challenging. Trainers had mentioned in the training that a reduction in pupil performance during the transition to the Focus4TAPS approach was expected, potentially due to children working more independently at first, and formed a part of moving towards formative assessments.

Challenges for engagement with programme delivery

Trainers and teachers also discussed several factors that acted as challenges to delivering Focus4TAPS lessons in schools. These can be categorised into school-based challenges and individual-level challenges.

School-based challenges included:

1. **Leadership and school priorities.** SLT buy-in varied across schools. According to teachers, while some schools' SLT supported training, others were hesitant to change science assessments in their schools. One perspective among trainers was that school priorities might not align with Focus4TAPS, which would hinder delivery.

   The programme developer also reported that some schools were required to follow strict approaches set by their MAT, which limited flexibility in introducing new approaches. In other schools, school policies mandated that pupils had to write down every aspect of their practical activities.

   Trainers also noted that science was often deprioritised in schools with tight budgets or where it was not seen as a core subject. They observed this was particularly common in some faith-based schools. The programme developer also reiterated this evidence, stating that one barrier to implementation was that there was not enough time for science in some schools.
2. **Resourcing challenges.** Teachers reported structural barriers including limited equipment, staff turnover, and lack of time. Trainers also mentioned high staff turnover was also seen as particularly disruptive, as it led to a loss of expertise and momentum within schools.

   > *When you lose a staff member, you've lost that impetus, that school has lost all that knowledge. Another school has gained it, but then that new subject leader's got to build up their reputation and things to be able to make an impact in a new school.* (Trainer 8, Trainer interview)

   > *It's not uncommon for 50 per cent of a primary school staff to change at the end of a year.* (Teacher 2, Teacher focus group)

The programme developer also highlighted limited school funding, which restricted resource purchases and contributed to an ongoing recruitment crisis, making it harder for schools to prioritise science and embed Focus4TAPS.

3. **Curriculum time and subject priorities.** Trainers noted that science was often deprioritised in schools facing tight budgets or where science was not considered a core focus. This was reported as particularly common in some faith-based schools. The programme developer echoed this, highlighting limited curriculum time for science as a key barrier to implementation.

Individual-level challenges included:

1. **Lack of time and skills development.** Trainers perceived that some elements of Focus4TAPS required skills that were difficult to develop within three training sessions. Teachers reported that lessons could be particularly challenging to facilitate in schools with behavioural difficulties or with younger pupils (in cases where they implemented the approach in other classes too). Trainers felt that teachers often needed more time and ongoing support to develop these skills before they could effectively implement the approach.
2. **Staff resistance and confidence.** Trainers reported that some staff were unwilling to change, preferring to use familiar methods over implementing new Focus4TAPS methods. Trainers highlighted that teachers without a strong scientific background were hesitant to start using formative assessments.
3. **Expectations around pace of implementation.** Trainers noted that some teachers expected to implement Focus4TAPS rapidly, despite it being designed as a longer-term process. This mismatch in expectations sometimes led to frustration or partial implementation among teachers.

Pupil responsiveness

This section draws together data from lesson observations, teacher focus groups, and pupil focus groups to explore findings on pupil engagement in Focused Assessment lessons, including drivers and challenges.

Pupil engagement with usual science lessons

Pupil attitudes towards usual science lessons provide useful context to understand their engagement with science lessons. In pupil focus group discussions, attitudes towards their usual science lessons were mixed. Some participants said that their usual science lessons were boring because they typically involved a lot of reading, writing, and listening. A contrasting view was that science lessons were enjoyable because participants got to learn new things and fun facts. Some pupils said they particularly enjoyed lessons involving practicals, which they considered easier and more fun than other lessons that were weighted towards written tasks.

> *…it helps us learn, but also it's fun while we do it because we get to like experiment on different things.*
> (Pupil 5, Pupil focus group)

On the other hand, some pupils mentioned that practical lessons included elements that they found challenging, such as writing up scientific findings. Reasons why this was difficult were not explicit in the pupil data.

Pupil engagement with Focus4TAPS lessons

This section examines the following question:

- IPE RQ2 (g) How engaged were pupils in the intervention Focused Assessment lessons and did this differ from their engagement in other science lessons?

Survey data, pupil focus groups, teacher focus groups, and lesson observations indicated that pupils had a good level of engagement with the Focus4TAPS lessons. In the endline survey, 70% (n=16) of Year 5 teachers agreed that pupils were 'highly engaged' in their Focused Assessment lessons, and the remaining 30% (n=7) reported that pupils were 'moderately

engaged'.[45] When asked whether pupil engagement changed over the course of the lessons, 61% (n=14) of teachers said that pupil engagement had increased, while 26% (n=6) stated that engagement stayed the same throughout. The remaining 13% (n=3) reported that it was difficult to comment on this.

Pupil engagement during whole-class activities

During whole-class activities, pupils commonly demonstrated engagement through participation and willingness to contribute. In lesson observations, pupils frequently raised their hands to answer questions, asked for clarification when unsure, and responded verbally to teacher prompts. Pupils appeared comfortable sharing ideas publicly, including contributing in front of the whole class or explaining their thinking to the class. Observations showed enthusiasm during practical experiments, especially when pupils could see changes occurring and record new observations.

In some cases, positive peer responses such as cheering appeared to increase pupil willingness to participate. When teachers asked at the end of the lessons whether pupils enjoyed them, almost all pupils would put their hands up. In focus groups, pupils mentioned that they could not think of anything boring about their lessons.

Pupil engagement during group work

Pupil engagement was good during group-based and collaborative activities. Observations showed that pupils were concentrating when they worked together on practical tasks, sharing measurements, and discussing results. For example, pupils watched closely as data were plotted on the interactive whiteboard and were keen to compare their findings with peers.

In interviews, teachers reflected that 'stepping back during lessons' increased pupil independence and ownership, which in turn supported engagement. We observed that group work allowed pupils to share ideas and ask questions.

Pupil engagement during individual tasks or recording

Pupil engagement during individual written tasks varied. We observed pupils actively recording predictions, methods, and results during practical lessons, particularly when writing was closely linked to hands-on activity.

In focus groups, pupils shared a range of views about writing. Some felt that recording results was less enjoyable than practical work, either because they considered it 'boring' or because they found recording data challenging and said they made mistakes. A contrasting view was that writing their own notes was preferable to copying from the teacher. Pupils who held this view explained that writing independently helped them remember information and work at their own pace.

Pupil engagement with peer assessment and self-assessment

In lesson observations, we noted quiet, focused classrooms during peer assessment and self-assessment tasks. Pupils in focus groups reported valuing feedback from peers as it helped them identify mistakes and improve their work.

> *When we do our work, we can't see what's wrong with it, but if other people look at it, they can find out what's wrong and help us.* (Pupil 1, Pupil focus group)

Teachers reported that pupils became skilled at identifying errors and suggesting improvements after using peer assessment. However, some teachers noted that while pupils were confident in giving feedback, they were less comfortable receiving it, particularly when feedback was recorded in their books.

[45] The question wording was 'Overall, how would you describe pupils' engagement with the Focus4TAPS programme?' (Highly engaged, Moderately engaged, Not very engaged, Not at all engaged).

Pupil engagement with Working Scientifically

This section explores pupils' views of science lessons under Focus4TAPS, with a particular focus on their experiences of Working Scientifically to answer the following question:

- IPE RQ2 (h) How did pupils feel about science lessons, and Working Scientifically in particular?

Pupil responses to Working Scientifically activities were mixed. Some pupils disliked focusing on a single Working Scientifically element. They found it 'stressful' and 'confusing' as it was not always clear to them what element was being focused on. On the other hand, other pupils stated that focusing on one element made the lessons easier for them and less rushed for them because they did not have to grasp many concepts at the same time.

Opinions on the Working Scientifically wheel were also mixed: one group of pupils liked using it to track their own skills, but other pupils did not recognise it or could not explain its purpose. This could be because some teachers had mentioned that they displayed the Working Scientifically wheel in classrooms while others stated that they did not do so.

In focus groups, both pupils and teachers noted that the 'Do' element of the Working Scientifically wheel was more engaging than the planning ('Plan') or evaluation ('Review') stages.

Additionally, pupil engagement during teacher explanations of scientific concepts was mixed. We observed that while some pupils listened attentively and were able to recall instructions, safety guidance, and investigative steps in detail, others appeared less focused during extended explanations. We noted instances of pupils handling equipment while teachers were speaking. Pupils themselves reported in focus groups that lengthy instructions of concepts were 'boring' when compared to doing the practical work.

Drivers and challenges to pupils' engagement

This section aims to answer the following question:

- IPE RQ2 (i) What were the drivers/challenges to pupils' engagement?

Drivers to pupil engagement

Analysis of lesson observations, teacher and pupil focus groups, and endline survey data suggested that pupil engagement in Focus4TAPS lessons was facilitated by five factors: pupil ownership and choice; hands-on and practical learning; group work and peer support; support from teachers; and focusing on one scientific element within lessons.

1. **Pupil ownership and choice.** In focus groups, pupils reported enjoying working independently without direct teacher instruction, as this allowed them to conduct experiments in their own way. Teachers also highlighted that pupils were more engaged when they were able to design aspects of investigations, choose variables, and complete their own recording. Teachers mentioned that they provided extension tasks for pupils who finished early, which helped sustain their focus.

   Survey findings supported this. Teachers were presented with a list of facilitators to pupil engagement to choose from in the survey. Some of the key facilitators that they identified were: having a choice of fun activities for pupils (70%, n=16); pupils doing the recording of information themselves (52%, n=12); and doing something different from usual lessons (35%, n=8).

2. **Hands-on activities and practical learning.** Pupils in focus groups consistently described practical activities as fun and interesting, particularly those involving movement or outdoor learning. Practical activities were also seen as more inclusive than usual class work, with pupils reporting that everyone could participate and feel involved, including those who were less confident in science or had limited prior experience of practical work.

   > *I also liked how we all get to participate, and people who don't like science as much…they still don't get left out.* (Pupil 2, Pupil focus group)

3. **Group work and peer support.** In focus groups, pupils stated that they liked group work because it reduced confusion since they could check their understanding with their peers. We observed that while some pupils dominated discussions, teachers and teaching assistants actively supported quieter members, ensuring everyone contributed. Overall, group work created opportunities for collaboration and pupils described in focus groups that this made lessons more enjoyable for them.

   > *I liked when we talked about it so none of us got confused on what we were doing.* (Pupil 3, Pupil focus group)

   When picking from the list of facilitators in the survey, 30% (n=7) of teachers reported that opportunities for peer assessment and self-assessment had acted as a facilitator for pupil engagement.
4. **Support from teachers.** Lesson observations showed that teachers supported engagement by providing clear instructions, modelling tasks, reinforcing key terminology, and managing transitions within lessons. When pupils became distracted, teachers intervened quickly and provided one-on-one support, or behaviour management. In focus groups, teachers mentioned that they provided tailored support for pupils with SEND, ensuring they could participate fully in practical and group activities.

5. **Focusing on one scientific element within lessons.** In the endline survey, when asked to pick from a list of facilitators, 91% (n=21) of teachers stated that focusing on one element of Working Scientifically had helped pupil engagement. Similarly, 83% (n=19) of teachers indicated that recording one element of the practical lesson/experiment helped pupil engagement.

   Teachers in the focus groups mentioned that focusing on one scientific element helped pupils understand what was expected of them and reduced the pressure pupils felt during practical work. They mentioned that narrowing the lesson objective helped maintain pupils' enjoyment of science while still giving them a clear purpose.

   > *Knowing that 'I have to record one thing and I'm being assessed on that one thing—I'm not being assessed on x, y and z' has kept them going. It's like stamina for learning. If I had to sit there and write out the whole thing every time I did an investigation, it takes the joy out of it. So part of it is giving them that joy, but with a key focus.* (Teacher 3, Teacher focus groups)

<u>Barriers to pupil engagement</u>

This section explores factors that made it harder for pupils to stay engaged during the Focused Assessment science lessons using data from pupil focus groups and survey data.

In focus groups, pupils were asked to reflect on any factors that made it harder for them to engage with their lessons. One view among pupils was that they struggled with using unfamiliar apparatus such as tape measures and pipettes. Another view was that mathematical elements within the lessons made it difficult for them to engage, such as calculating averages or planning bar graphs.

Also, written activities were less engaging for some pupils. Pupils reported that recording results or writing down information was less interesting than completing practical activities.

In the endline survey, 61% (n=14) of teachers reported that they had not experienced challenges related to pupil engagement. Other teachers selected barriers from the list provided as follows: challenges to providing opportunities for self-assessment or peer assessment (17%, n=4); challenges related to doing something different from pupils' usual lessons (17%, n=4); pupils doing the recording of information themselves (13%, n=3); followed by challenges related to recording one element of the practical lesson or experiment (9%, n=2).

Differences based on pupil characteristics

This section examines the following question:

- IPE RQ2 (j) Did patterns of engagement and drivers/challenges differ by pupil characteristic, i.e. by pupil disadvantage?

Across lesson observations and pupil focus groups, there was no evidence of differences in pupil engagement with Focus4TAPS lessons by pupil characteristics. Pupils across classes were observed to participate in whole-class activities, and teachers did not report differences in engagement linked to disadvantage or other characteristics. In focus groups, pupils reported varying levels of enjoyment across different lesson components, and these views were not linked to specific characteristics.

However, one view among teachers in focus groups reported that the flexibility of Focus4TAPS lessons made them more accessible for some pupils with SEND who would typically engage less with written work.

> *I've seen big changes in accessibility for children with SEND. In the past, some of them would switch off as soon as they saw a long worksheet. Now, because the recording can be flexible—drawings, oral responses, post-it notes – they're much more involved.* (SENCO, Teacher focus group)

Another teacher presented an example where a child who was usually below age-related expectations produced detailed and accurate scientific diagrams when given the freedom to choose how to record their work. The teacher felt that this flexibility enabled the child to engage in science lessons.

> *There's one girl in particular that really comes to mind [...] academically, in terms of maths and English, she's always been below age-related [expectations], but God, her life cycles were the most beautiful thing [...] she wanted to draw those scientific diagrams. [...] but alongside that, she still did all the writing, and it all matched all the pictures. I think it could have very easily gone down the route of, 'Here, cut this out and stick it,' but she was given the opportunity to do what she thought she was capable of.* (Teacher 7, Teacher focus group)

Lesson observations supported this view, showing that pupils with SEND were able to engage in practical and group activities when teachers provided clear instructions.

Wider outcomes

This section explores the wider perceived outcomes of Focus4TAPS for pupils, teachers, and schools, as identified by teachers and pupils. The extent to which these findings align with the intended outcomes set out in the programme logic model, and the mechanisms and barriers influencing them, are also explored. This section brings together findings from the teacher and science lead surveys, focus groups with teachers and pupils, and trainer interviews (in which participants shared insights from teachers' feedback to them during the training programme about implementation of lessons in their classrooms).

This section aims to answer the following question:

- IPE RQ3: What are the longer-term implications of Focus4TAPS for teaching and learning science in primary schools?

Overall, our data suggested that Focus4TAPS was associated with positive perceived outcomes at both teacher and pupil levels, particularly within classrooms. At the teacher level, findings indicated increased confidence in teaching Working Scientifically, increased knowledge of formative assessment in science, and high confidence in teaching science more broadly. Teachers also reported intentions to continue using the approach beyond the intervention.

At the pupil level, teachers reported consistently high interest in Working Scientifically lessons, alongside perceived positive attitudes towards science overall after Focus4TAPS and increased engagement with learning and school more broadly. Teachers also reported increased pupil confidence in science abilities and improvements in science attainment for pupils.

Sustainability of Focus4TAPS implementation

This section aims to answer the following question:

- IPE RQ3 (a) Could teachers in the intervention group continue to implement Focus4TAPS without any further support or training?

In the endline survey, almost two-thirds (67%, n=24) of Focus4TAPS teachers and science leads 'agree' with the statement: 'Teachers at my school will continue to implement the Focus4TAPS approach in the future without any further support or training'. The remaining 17% (n=6) 'disagree' with the statement while 17% (n=6) 'neither agree nor disagree'.

Among focus group participants, there was less variation in views: all teachers reported that they intended to continue to use Focus4TAPS beyond the intervention and that they had embedded the approach in their planning routines.

Some teachers described adaptations to their Focus4TAPS approach. One example was moving away from discussing the Working Scientifically wheel with pupils as part of their lessons, which some teachers felt could be time-consuming. They planned instead to display the Working Scientifically wheel and focus interactive sessions/discussions on other lesson plan elements.

Survey responses indicated that many teachers reported having achieved key outcomes by the end of the Focus4TAPS programme, particularly in relation to confidence and formative assessment practice. At endline, Year 5 teachers reported increased confidence in teaching Working Scientifically (74%, n=17), increased knowledge of formative assessment within science (70%, n=16), and high confidence in teaching science in general (70%, n=16). These outcomes are discussed in more detail later in this section.

However, when compared with anticipated outcomes reported at baseline, levels of agreement were consistently lower at endline, suggesting that the extent to which outcomes were realised in practice was more limited than initially expected. In particular:

- Fewer teachers reported increased networking and collaboration with other teachers or schools at endline (43%, n=10) compared with baseline expectations (59%, n=115).
- Perceived access to future training related to formative assessment within science declined substantially, from 83% (n=162) at baseline to 39% (n=9) at endline.
- Expectations regarding school-wide application of the Focus4TAPS Focused Assessment approach also reduced, with agreement falling from 68% (n=134) at baseline to 39% (n=9) at endline.

These findings suggested that teachers experienced positive outcomes at the individual classroom level. However, reduced access to future training, lower levels of networking, and limited whole-school adoption may constrain the programme's longer-term sustainability.

Scaling up of Focus4TAPS

This section aims to answer the following question:

- IPE RQ3 (b) What are the drivers and obstacles to scaling up the training model to increase the number of trainers and teachers delivering Focus4TAPS?

This section explores the factors that may support or constrain the scaling up of the Focus4TAPS training model, drawing on survey data, focus groups with teachers, and trainer interviews. Scaling is understood here as the expansion of Focus4TAPS beyond individual classrooms to a greater number of teachers, schools, and trainers.

Drivers to scaling up

Several drivers that could support the scaling up of Focus4TAPS were identified across the data. One key mechanism was the sharing of Focus4TAPS strategies and resources within schools. In the endline survey, almost all Focus4TAPS group

teachers and science leads (94%) reported that they planned to support other teachers in their school to implement the approach.

Additionally, qualitative data suggested that science leads played an important role in supporting the spread of Focus4TAPS practices. Teachers described sharing resources and approaches through joint planning, informal discussions, and wider staff conversations. In some schools, this sharing extended beyond science, with teachers reporting that they had applied Focus4TAPS principles to other curriculum subjects, such as foundation subjects and music. These practices were perceived as supporting clearer understanding of pupil progression and more use of formative assessment across other subjects.

Teachers also described other drivers of the programme in focus groups and through feedback to the trainers:

- **Increased access to peer networks.** One view among teachers was that, by bringing together teachers and science leads in regional groups, the programme facilitated useful access to a supportive network of their peers. Sharing ideas and best practice within training sessions supported teachers' knowledge and understanding, contributing to outcomes described in the logic model.

- **Increased focus on science within schools**: In focus groups, one view was that sharing Focus4TAPS materials had led to a greater attention on science among the wider school—for example, science was discussed more frequently in staff meetings and governors' meetings, positioning it alongside core subjects such as English and maths:

  > *Governors have asked more questions about science since we shared some of the TAPS work with them. It's put the subject back on the map.* (Teacher 6, Teacher focus group)

These findings highlighted several mechanisms that could support the scaling up of Focus4TAPS, including internal sharing within schools, the role of science leads, access to peer networks, and increased institutional attention to science.

Obstacles to wider implementation

The data also pointed to a number of obstacles that may constrain the wider scaling of Focus4TAPS. Application of the Focus4TAPS Focused Assessment approach across schools was more limited than anticipated. Less than half (39%, n=9) had applied Focused Assessments across their schools at the time of the endline survey, although 68% (n=134) of respondents had anticipated that they would do so when completing the baseline survey.

Teachers in focus groups reported that schools were at different stages of embedding Focus4TAPS. For some schools, this was limited to Year 5, while others had progressed to trialling the approach with other year groups. Some teachers described integrating the Year 5 content into their regular lesson planning and said that they intended to continue to use the approach on an ongoing basis while others discussed implementing Focused Assessments across their school. These findings indicated uneven uptake and suggested that scaling beyond initial implementation may require sustained support over time.[46]

Role of science leads and teachers

Teachers and science leads also described variation in capacity to support wider roll-out within schools. Differences in remit and prior familiarity with the TAPS model appeared to shape how participants engaged with scaling up within schools. While science leads often reported greater responsibility for supporting whole-school implementation, Year 5 teachers tended to focus on applying the approach within their own classrooms. Trainer interviews similarly highlighted variation in how confidently teachers felt able to support others without additional guidance and indicated that science leads were often more confident in supporting others.

[46] This trial focused specifically on Year 5 and typical delivery of Focus4TAPS is intended to extend across all year groups.

Wider outcomes for schools and teachers

This section explores the perceived effects of Focus4TAPS on schools and teachers, drawing on survey data, focus groups with teachers, and trainer interviews. Findings are structured around three areas: teachers' understanding of formative assessment in science; teachers' confidence and attitudes towards Working Scientifically and science more broadly; and the extent to which teachers have begun to apply the Focused Assessment approach across the science curriculum. It links these areas to the short-term and longer-term outcomes set out in the programme logic model. Bringing these findings together, we aim to answer the following question:

- IPE RQ3 (d) What are the perceived effects of Focus4TAPS on schools and teachers?

IPE RQ3 (d) (i) To what extent have teachers gained an enhanced understanding of formative assessment in science?

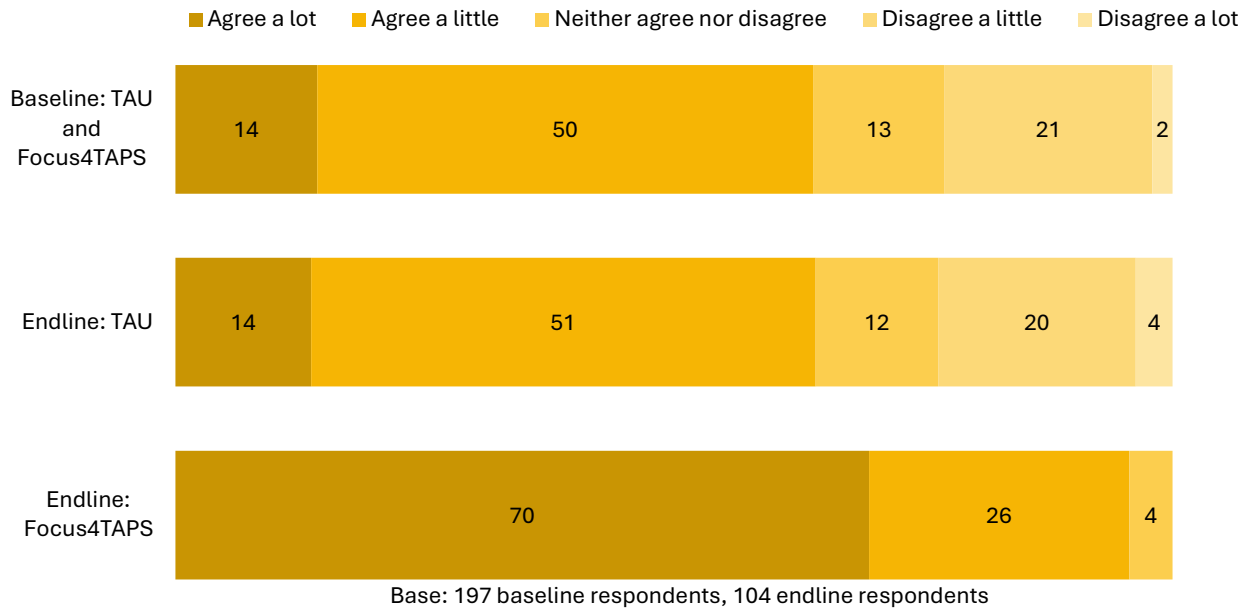
Evidence from surveys, focus groups, and trainer interviews suggested that Focus4TAPS supported teachers in gaining an enhanced understanding of formative assessment in science. This aligned with an intended short-term outcome of the programme.

Short-term outcome: Teachers gain an enhanced understanding of applying assessment strategies and progression in science

Teachers who took part in the baseline and endline survey were asked two questions about their understanding of formative assessment. The first asked the extent to which they 'agree' with the statement: 'I have a good understanding of formative assessment in Working Scientifically lessons'. The second question asked about formative assessment in science more broadly, through agreement with the statement: 'I do not have a good understanding of formative assessment in science lessons'.

Across both questions, response patterns for teachers in the teaching as usual group remained largely stable between baseline and endline, indicating little change in reported understanding over time. In contrast, teachers in the Focus4TAPS group showed marked differences between baseline and endline responses for both questions. These response patterns are shown in Figure 10 and Figure 11.

Figure 8: Whether teachers 'agree' with the statement: 'I have a good understanding of formative assessment in Working Scientifically lessons' (baseline survey, endline survey is split by teaching as usual and Focus4TAPS group), %

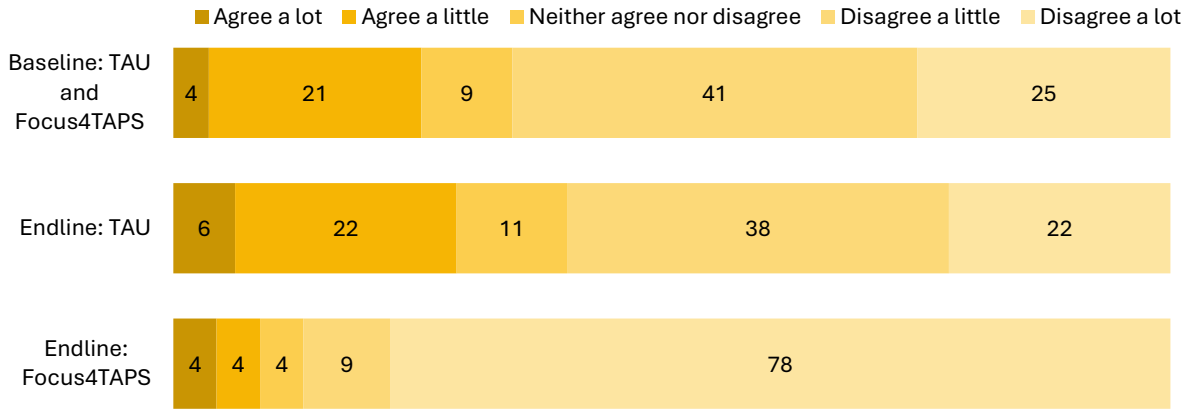


Base: 197 baseline respondents, 104 endline respondents

At baseline, just over two-thirds of teachers (64%, n=126) reported that they had a good understanding of formative assessment in Working Scientifically lessons. This proportion remained similar for the teaching as usual group at endline (65%, n=52). However, among Focus4TAPS teachers, almost all respondents (96%, n=22) 'agree' with this statement at endline. A similar pattern was observed for the question relating to formative assessment in science more broadly, with Focus4TAPS teachers showing a shift towards greater agreement and a reduction in neutral responses between baseline and endline.

Overall, these survey findings indicated that Focus4TAPS was associated with improved understanding of formative assessment, both in relation to Working Scientifically and to science lessons more broadly. This suggests that the programme supported teachers in developing a clearer and more applied understanding of formative assessment, consistent with the intended short-term outcomes in the logic model.

Figure 9: Whether teachers agree with the statement: 'I do not have a good understanding of formative assessment in science lessons' (baseline survey, endline survey is split by teaching as usual and Focus4TAPS group), %



Base: 197 baseline respondents, 104 endline respondents

Qualitative data reinforced these findings. Teachers reported that Focus4TAPS encouraged them to see formative assessment as part of everyday science teaching rather than as an additional or separate activity. In focus groups, teachers described how the programme gave them time and structure to plan and reflect on formative assessment within science lessons. They reported that tools such as the Working Scientifically wheel and Focused Assessment supported deliberate planning and provided 'quick wins', as they were easy to use and adaptable to lesson needs, so they made the process of formative assessment easier for teachers. One view among teachers was that this enabled them to focus lessons on specific enquiry skills where pupils needed more support.

Teachers also described developing a clearer understanding of how assessment strategies could be used flexibly. In focus groups, teachers noted a shift away from an emphasis on full written experiment reports, recognising that assessment did not always require extensive written reports. Instead, in their feedback to trainers and in focus groups, they described using other methods of assessment like observation, discussion, and focused recording to gather evidence of pupils' understanding and progression. Trainer interviews supported these findings, with trainers reporting that teachers had learned strategies for ongoing assessment during lessons and had begun to move beyond formal testing.

The programme developer felt that a key difference of the Focus4TAPS approach was that it placed more emphasis on 'thinking and doing' rather than writing. Teachers mentioned that this was a distinction from their usual approach and Focus4TAPS made the recording element of practical science lessons feel more purposeful.

> *We've always put emphasis on practical work but sometimes the recording side felt forced. I've found the TAPS tasks helpful because they narrow down what you're actually looking for, which makes lessons feel more purposeful.* (Teacher 6, Teacher focus group)

Although teachers described greater understanding and use of formative assessment strategies, the data provided limited direct insight into how this translated into explicit understanding of progression in science.

Short-term outcome: Teachers use their enhanced understanding to adapt the level of support and challenge and improve their practice

This outcome built directly on the preceding short-term outcome, where teachers developed a clearer understanding of formative assessment.

In interviews, trainers described how, during the programme, teachers moved away from rigid lesson delivery and began observing pupils more closely in lessons. Trainers reported that teachers reflected during training workshops on how formative assessment helped them identify when pupils needed more or less support and adjust their teaching accordingly.

Teachers' accounts in focus groups supported this view. Teachers reported that, in Focus4TAPS lessons, they used formative assessment to gain insight into pupils' understanding and progress and adapted lesson delivery in response. One view among teachers was that they were shifting their focus from checking whether pupils had the 'right answer' to assessing the development of scientific skills. This was described as making assessment feel more meaningful.

> *Personally, I feel more confident about what I'm assessing—it's not just, 'Did they get the right answer?' but 'Can they show me a scientific skill?' which feels more meaningful.* (Teacher 4, Teacher focus group)

Teachers also described using formative assessment findings to plan next steps for individual pupils and for the class as a whole. This linked to the outcome of adjusting support and challenge within lessons where formative assessments were used. These examples mainly related to teachers' experiences during the programme and early implementation of Focus4TAPS.

> *Seeing what they could do helped me decide what to do next, for pupils and for the science lessons.* (Teacher 6, Teacher focus group)

Overall, the evidence suggested that teachers used formative assessment to make their science teaching more responsive to pupils' needs.

Short-term outcome: Teachers take time to plan and reflect on formative assessment within science lessons

In focus groups, teachers reported that Focus4TAPS gave them time to plan and reflect on formative assessment within science lessons, which was an intended short-term outcome of the intervention. This planning and reflection was most clearly described as taking place during the training period, where teachers were supported through 'gap tasks' to think about how formative assessment could be implemented in their classrooms.

Trainers reported that teachers shared reflections during training sessions about implementing Focused Assessments, which supported more reflective lesson design.

While teachers confirmed that they planned and reflected on Focused Assessments in their lessons, there was more limited evidence that structured planning and reflection continued beyond the training period. Overall, the endline survey, lesson reflections, and teacher focus group data suggested that Focus4TAPS enabled planning and reflection on formative assessment during the intervention, though longer-term continuation was not discussed.

IPE RQ3 (d) (ii) To what extent have teachers improved confidence and attitudes towards Working Scientifically and science more broadly?

In addition to improved understanding of formative assessment, teachers reported improvements in confidence and attitudes towards Working Scientifically and science more broadly. These changes reflected a longer-term outcome within the logic model.

Long-term outcome: Teachers have improved confidence and attitudes towards Working Scientifically and science more broadly

Teachers reported significant improvements in confidence across areas like Working Scientifically and science more broadly. Most (70%, n=16) of science leads and teachers reported increased confidence towards teaching science in general.

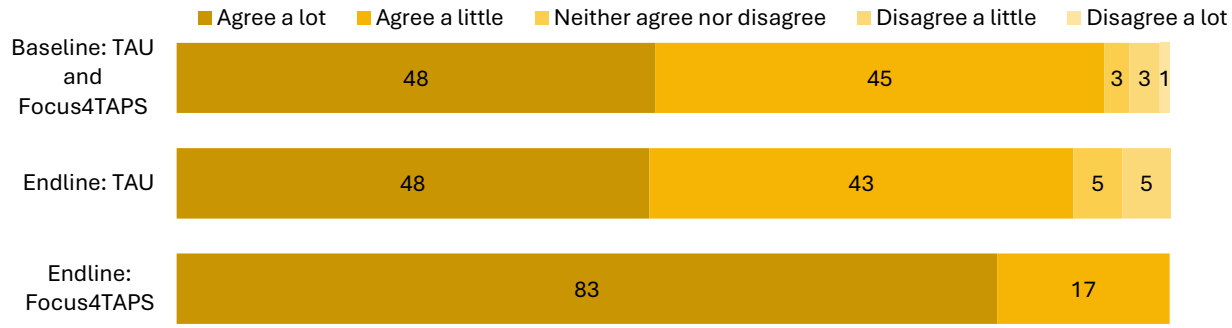
Survey data indicated that teachers already reported high levels of confidence in teaching science lessons to Year 5 pupils at baseline (Figure 12). At baseline, 93% (n=184) of teachers reported that they were confident in teaching science lessons to Year 5 pupils.

At endline, all Focus4TAPS group respondents (100%, n=23) reported that they were confident in teaching science lessons to Year 5 pupils, compared with 91% (n=74) of respondents in the teaching as usual group. While the increase among Focus4TAPS teachers was modest given the high baseline, this suggested that confidence in teaching science was sustained over the course of the programme.

Similarly, endline survey data showed that most science leads and teachers from the Focus4TAPS group (74%, n=17) reported increased confidence towards Working Scientifically.

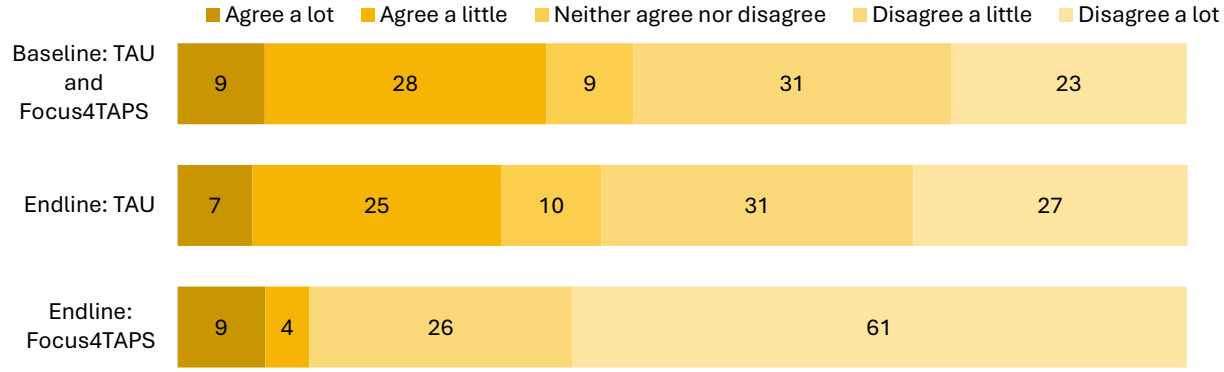
Confidence in teaching Working Scientifically was lower at baseline and therefore showed more scope for change. As shown in Figure 13, a higher proportion of Focus4TAPS teachers reported confidence in teaching Working Scientifically at endline compared with baseline and compared with the teaching as usual group.[47]

Figure 12:10 Whether teachers agree with the statement 'I am confident to teach science lessons to Year 5 pupils' (baseline survey, endline survey including teaching as usual and Focus4TAPS group), %



Base: 197 baseline respondents, 105 endline respondents

Figure 11: Whether teacher agree with the statement 'I am not confident to teach Working Scientifically to Year 5 pupils' (baseline survey, endline survey including teaching as usual and Focus4TAPS group), %



Base: 197 baseline respondents, 104 endline respondents

[47] It is important to caveat that the baseline and endline surveys were not completed by the same teachers in all cases, and differences in responses may therefore partly reflect changes in respondents rather than changes over time alone.

In focus groups, teachers described confidence in teaching and assessment of science as a result of the programme. One view among teachers was that they had become more confident in their subject knowledge and in making pedagogical decisions after accessing the training and Focus4TAPS resources as this had helped them build an understanding of formative assessment. This view was also shared by Focus4TAPS trainers in their interviews, where they reflected on teachers' ability and confidence in identifying pupil needs and applying best practice in primary science, which they said had increased over the duration of the programme.

Alongside increased confidence, some teachers also highlighted changes to their attitudes towards science. One view among teachers mentioned that they had a positive 'mindset shift' towards Working Scientifically and science more broadly. Teachers described moving away from a teacher-led approach towards facilitating pupil-led investigations, which they associated with a more enjoyable experience of teaching science.

Trainer interviews echoed these findings, with trainers reflecting that teachers' confidence in identifying pupil needs and applying best practice in primary science increased over the duration of the programme. Trainers also described a shift towards more pupil-centred approaches to teaching science.

Additionally, in focus groups, teachers described feeling more enthusiastic and excited about teaching science. One view among teachers was that having clear, ready-to-use resources for science teaching reduced planning demands, helping science feel more manageable within a busy curriculum. They described that the Focus4TAPS resources also reduced the time that they had to spend on lesson planning. Teachers linked this reduction in stress to more positive attitudes towards teaching the subject.

Teachers also described changes in how they approached science lessons, which they associated with increased enjoyment and engagement. One view in the focus group discussions was that engagement with Focus4TAPS encouraged greater creativity in science teaching, with teachers feeling more able to use practical activities in their lessons. Teachers reported that this supported increased curiosity among pupils and seeing pupils respond positively to these activities further increased teachers' enthusiasm for teaching science.

Overall, our data suggested that Focus4TAPS supported improvements in teachers' confidence and attitudes towards Working Scientifically and science more broadly.

IPE RQ3 (d) (iii) To what extent have teachers begun to apply the Focused Assessment approach across the science curriculum?

While the Focus4TAPS programme targeted Year 5 science, a longer-term outcome described in the logic model was for teachers to apply the Focused Assessment approach more consistently across their science teaching, with potential application to other year groups over time.

Long-term outcome: Teachers apply Focused Assessment approach across the science curriculum

Evidence from surveys and focus groups suggested that progress towards this outcome was mixed.

In the endline survey, we asked science leads and teachers whether they understood how to apply the Focus4TAPS approach to other science year groups. All respondents reported that they did: 72% (n=26) 'strongly agree'; and 28% (n=10) 'agree' with this statement.

However, fewer teachers reported having implemented the approach beyond Year 5. While 68% (n=134) of respondents anticipated at baseline that Focused Assessment would be applied across their school, less than half (39%, n=9) reported that this had occurred by endline.

Focus group data suggested that schools were at different stages of embedding Focus4TAPS. Some teachers reported that implementation remained limited to Year 5, while others described trialling the approach in additional year groups, such as Year 2 or Year 4, or using Focused Assessments across the school. These accounts indicated variation in how far schools had progressed towards using Focused Assessment across the science curriculum. This section examines the following question:

- IPE RQ3 (e) How do science leads in schools intend to support other teachers in their schools to implement the approach?

Evidence from surveys, focus groups, and trainer interviews indicated that science leads intended to support other teachers to implement the Focus4TAPS approach primarily through sharing information, resources, and planning practices within their schools. This aligned with a longer-term outcome in the logic model, which anticipated that science leads would support wider implementation of the approach.

In the endling survey, 94% (n=34) of Focus4TAPS group teachers and science leads reported that they planned to support other teachers in their school to implement Focus4TAPS. Teacher focus group participants said that it would be essential to share information with colleagues to embed the approach across classes and year groups and sustain practice over time.

Survey data also indicated that many teachers and science leads had already shared specific Focus4TAPS resources, most commonly guidance on how to access and use Focused Assessment lesson plans for any year group (96%, n=27), strategies for implementing formative assessment (64%, n=18), and approaches to focused recording (64%, n=18). The most commonly shared items are shown in Table 27.

Table 27: Sharing of Focus4TAPS resources and strategies[a]

Resource/strategy shared	%
How to access and use Focus4TAPS Focused Assessment lesson plans for any year group	96
How to implement formative assessment in lessons for any year group	64
Focus4TAPS strategies related to focused recording	64
How to focus on one element of Working Scientifically in Year 5 science lessons	61
Focus4TAPS strategies related to types of enquiries	54
Focus4TAPS strategies related to elicitation strategies	50
Focus4TAPS strategies related to self-assessment or peer assessment	32
Base (number of respondents)	*n=28*

[a] *We do not have additional data on the reasons why these resources were shared.*

Additionally, trainers reported that discussions with science leads had encouraged Year 5 teachers to trial Focus4TAPS principles in other contexts.

However, the available data does not allow firm conclusions to be drawn about the extent to which this sharing occurred across schools, or whether such sharing led to changes in practice. As such, while there is evidence that science leads were sharing their learning supported, these findings primarily reflect their intentions and early experiences. We are unable to comment on how this support will be implemented in the future.

Wider outcomes for pupils

This section reports on the perceived effects of Focus4TAPS on pupils, drawing on evidence from surveys, focus groups with pupils and teachers, and trainer interviews. Findings are structured around logic model outcomes, including pupil engagement, interest and enjoyment in science, pupil confidence, and pupils' understanding and attainment. The section aims to answer the following question:

- IPE RQ3 (c) What are the perceived effects of Focus4TAPS on pupils' engagement with learning/school more broadly, and does this differ by pupil characteristics?

Pupil engagement, interest, and enjoyment in science

This section examines pupils' engagement, interest, and enjoyment in science, drawing on teacher survey data and focus groups with pupils and teachers.

Teachers reported changes in pupils' attitudes and engagement. At endline, 79% (n=94) of Focus4TAPS teachers reported that pupils had positive attitudes towards science overall after being involved in the programme, while 66% (n=79) reported increased engagement with learning and school more broadly as a result of Focus4TAPS.

Baseline survey data indicated that pupils had already shown high levels of interest in Working Scientifically lessons prior to the intervention. At baseline, almost all teachers (93%, n=184) 'agree' that pupils generally looked forward to Working Scientifically lessons, and the same proportion (93%, n=183) 'agree' that pupils found these lessons interesting. This suggested that baseline engagement with Working Scientifically lessons was high.

At endline, teachers reported similarly high levels of interest in Working Scientifically lessons. Around 91% (n=21) of Focus4TAPS group teachers 'agree' that pupils generally looked forward to Working Scientifically lessons, and 96% (n=22) 'agree' that pupils found these lessons interesting. Given the high baseline, these findings suggested that Focus4TAPS helped to maintain high levels of pupil interest and engagement in Working Scientifically over the course of the programme.

In the teacher survey, the proportion of Focus4TAPS group teachers who 'disagree' with the statement: 'Year 5 pupils generally find science lessons boring' increased from 42% at baseline to 48% at endline. This suggested that, by the end of the Focus4TAPS intervention, a greater proportion of teachers perceived pupils as not finding science lessons boring. This finding matched the high levels of reported interest in science lessons and pupils looking forward to Working Scientifically lessons.

In focus groups, pupils from the Focus4TAPS group consistently described science, particularly practical activities within Focus4TAPS lessons, as 'exciting'. When we asked participants to describe science in one word, the responses included words such as 'fun', 'amazing', and 'I can't wait'. They explained that experiments and outdoor activities made Focus4TAPS lessons more enjoyable when compared to lessons without practical activities where they were 'just listening to the teacher talking'.

Similarly, teachers in focus groups noted that pupils were eager to participate in Focused Assessment science lessons and often asked them when science would take place. Teachers described pupils as motivated and keen to demonstrate their science learning:

> *They really want to do the science lessons and they like the way the science lessons are being done. They're very excited about how to show their science learning and like, 'Oh, what is our learning going to look like this week?'* (Teacher 5, Teacher focus group)

Teachers also reported that Focus4TAPS lessons allowed pupils time to explore ideas, discuss their reasoning and work at an appropriate pace without feeling rushed which increased pupils' engagement. Some teachers noted that parents/carers commented on children talking enthusiastically about experiments at home, suggesting that pupils' enjoyment of science extended beyond the classroom.

Pupil confidence

Two outcomes in the logic model relate to pupil confidence: in the short term, confidence in Working Scientifically; and in the longer term, confidence in pupils' science abilities.

Endline survey findings indicated improvements in pupils' science confidence more broadly. Around 82% (n=97) of teachers reported increased pupil confidence in their science abilities as a result of Focus4TAPS.

Additionally, in the baseline survey, teachers were asked to indicate the extent to which they 'agree' or 'disagree' with the statement: 'Pupils do not generally feel confident about their skills during lessons related to Working Scientifically'. At baseline, the proportion of respondents who 'agree' with this statement (39%, n=77) was slightly higher than the proportion who 'disagree' (36%, n=70), which indicated mixed perceptions of pupil confidence. A further quarter (n=25) of respondents expressed a neutral view, which suggested some uncertainty about pupils' confidence in Working Scientifically prior to the intervention.

We asked this question again at endline from the teaching as usual group respondents. At endline, higher proportions of respondents in the teaching as usual group reported disagreeing with the statement, indicating that pupil confidence was

perceived to have increased over this period for both groups. The proportion who 'agree' that pupils did not feel confident increased slightly from 39% (n=77) at baseline to 43% (n=35) at endline. The proportion who 'disagree' with the statement was 44% (n=36), representing an increase of eight percentage points from baseline. This suggested movement towards more positive perceptions of pupil confidence, alongside a slight increase in negative perceptions.

Qualitative data provided insight into how these changes were experienced in practice in Focus4TAPS schools. Teachers and trainers reported that pupils became more confident in planning investigations when lessons focused on developing one enquiry skill at a time. They described how the structured approach used in Focus4TAPS lessons allowed pupils to revisit and apply prior knowledge, which supported confidence during practical work.

Pupils echoed these views in focus groups, explaining that building understanding of a topic and the process for carrying out experiments in earlier lessons made them feel more confident when completing practical activities.

> *Building up the topic in non-practical lessons first made me confident before doing the experiment.* (Pupil 2, Pupil focus group)

Teachers also reported increased confidence among pupils who had previously been hesitant to participate. In focus groups, teachers described pupils showing greater willingness to answer questions, contribute to discussions, and share ideas. Teachers linked this to pupils having a stronger understanding of the science content because of Focused Assessment.

Teachers also noted that pupils were more confident using scientific vocabulary, which appeared to result from increased familiarity with key terms and more opportunities to practice using them in lessons. Pupils similarly reported feeling more confident when completing homework, linking this to improved understanding of lesson content because of carrying out practical activities.

Pupils reported some variation in confidence which was based on the science topics. Pupils commonly described feeling more confident when working on space-related topics, while plant-related topics were described as more challenging. The reasons for this variation were not explored during data collection, though it may reflect differing levels of familiarity with specific topic areas rather than the Focus4TAPS approach itself.

Pupil understanding and attainment

This section reports on perceived effects of Focus4TAPS on pupils' understanding and attainment in science. In line with the logic model, findings are structured to first consider short-term outcomes related to understanding and attainment in Working Scientifically, before examining longer-term and broader perceptions of science understanding and attainment.

Survey data indicated positive perceptions of pupils' understanding of Working Scientifically. In the endline survey, most (91%, n=108) Focus4TAPS group teachers reported that pupils had a good understanding of key concepts and skills associated with Working Scientifically.

Teachers and pupils reported improvements in pupils' scientific enquiry skills particularly when lessons focused explicitly on a single element of Working Scientifically. Narrowing the focus of science lessons allowed pupils more time and support to develop each specific skill in turn. Teachers also reported that focused recording supported pupils to demonstrate their understanding without excessive writing, which was seen as particularly supportive for some learners.

Teachers in focus groups reported that this approach appeared to be particularly beneficial for pupils who found elements of Working Scientifically challenging, especially recording and evaluation. Teachers described pupils demonstrating greater accuracy in recording results and increased confidence in evaluating investigations. In focus groups, teachers also reported improvements in pupils' scientific thinking, including decision-making, forming questions, and linking theory to practice.

> *The science work was much more purposeful, with evidence of thinking rather than just copied text.* (Teacher 4, Teacher focus groups)

Data from pupil focus groups supported this finding. In focus groups, pupils described how practical activities helped them understand scientific concepts and processes more clearly. Pupils explained that carrying out experiments linked to difficult concepts made learning clearer and easier to understand.

Overall, these findings suggested that Focus4TAPS was perceived to support improved understanding and attainment in Working Scientifically in the short term.

Beyond Working Scientifically, teachers and pupils also reported improvements in pupils' overall science understanding and attainment. In the endline survey, 77% (n=92) of Focus4TAPS group teachers reported that they had noticed an improvement in science attainment.

In focus groups, teachers reported that pupils' written work across science lessons became more purposeful, with greater emphasis on reasoning and explanation rather than copying information. Teachers also described pupils applying prior knowledge more confidently when encountering new topics, suggesting improvements in science understanding.

According to teachers in focus groups, pupils also demonstrated improved evaluation skills and greater accuracy in recording results.

> *Before, they'd rush through without much thought, but now they're more deliberate. For example, when we did a friction experiment, pupils actually debated what variables to keep constant and were able to justify their choices. That's not something I'd seen them do so confidently before.* (Teacher 4, Teacher focus group)

Pupils similarly described feeling more confident in their science learning overall. Pupils reported that engaging in practical activities helped them answer questions more confidently.

> *I think my confidence, it is up because when we finish topics, I have all that knowledge in my head, so when the questions come back to the topic, I can answer them.* (Pupil 6, Pupil focus group)

Teachers also reported that Focused Assessments supported pupils across a range of attainment levels. One view among teachers was that the approach encouraged pupils of different attainment levels to produce good quality work, as pupils had greater freedom to carry out experiments and record their findings in ways that suited them.

> *In terms of attainment, I've noticed the gap closing between my high attainers and middle group, because the focused tasks give everyone a clear route to success.* (Teacher 5, Teacher focus group)

Overall, the evidence suggested that Focus4TAPS was perceived to support improvements in pupils' understanding and attainment in Working Scientifically in the short term, with indications of broader improvements in science understanding and attainment over time.

Differences between pupils

In focus groups, we asked teachers to reflect on any variation in perceived effects of the programme between different groups of children. Some teachers identified differences in perceived outcomes for pupils with additional needs and in relation to gender. Other teachers reported that they had not observed noticeable differences in outcomes based on characteristics between pupils in their classes.

One view among teachers was that Focus4TAPS led to increased confidence and engagement for pupils with additional needs. This was because they were better able to engage and demonstrate their understanding through discussion, practical activities, and visual representations and did not have to rely solely on written work to demonstrate their understanding. Overall, this suggested that the approach supported engagement and access to the science curriculum for a wide range of learners.

Some teachers also reported perceived differences in engagement by gender. A view among teachers was that, compared with previous science lessons, boys appeared more engaged during Focus4TAPS lessons. Teachers attributed this to the practical nature of the activities, which encouraged hands-on involvement such as sourcing equipment and calculating averages by using experiments. This view also described girls as showing strengths in evaluation and explanation within Focus4TAPS lessons, which they felt created an 'encouraging balance' of participation within classes.

Cost

The endline survey questionnaire administered to teachers and science leads provided information on the time and costs involved in delivering Focus4TAPS. This was supplemented with information from the delivery partner (Bath Spa University) on the costs associated with training the trainers who delivered the Focus4TAPS training to teachers.

Costs reported by delivery partner

Bath Spa University reported that a total of 25 trainers delivered training across 13 groups (one trainer was in two groups). The training of trainers was one-off, with trainers expected to deliver training across multiple years. The training sessions for trainers were based on a hybrid model with trainers attending the CPD days in person, then doing online sessions with the developer. The current training model pairs newer trainers with more experienced trainers, reducing the need for in-person training. The total time spent by the developer on preparation for the training sessions was estimated at 27 hours, covering trainer recruitment and preparing training sessions.

Bath Spa University also incurred £2,600 in costs of training materials (£200 per group), which are reusable and so form a set-up cost for the programme.[48] The total time spent by the developer on training the trainers was estimated at 49.5 hours. The total costs incurred to the developer for training the trainers was estimated at £4,800. Trainers were paid £500 per person per training day that they delivered. Trainers spent 234 hours in total preparing training sessions for Focus4TAPS teachers—this included six hours per trainer for preparing training materials, and one hour per trainer in booking room hire, organising catering, and coordinating trainees.

The usual fee charged to schools for a year of Focus4TAPS training would be £600 per school. The Focus4TAPS schools accessed it at a subsidised rate of £150 instead. In line with the EEF cost guidance (EEF, 2023), costs were estimated using the market rate of £600 per school.

Costs reported by teachers

There were 62 total responses to the teacher survey from Focus4TAPS schools from the Year 5 teacher or the science lead. Responses to questions have been summarised by role, from the perspective of either the Year 5 teacher (n=31 teachers) or the science lead (n=39 teachers), with some teachers having both the Year 5 teacher and science lead roles (n=8). In most cases, schools provided responses from one teacher with one school where both the Year 5 teacher and the science lead provided responses.

The survey asked teachers to report the time they spent in hours in various aspects of Focus4TAPS training and delivery. The teacher survey also asked about costs incurred by schools in terms of training materials or other costs, as well as the time required in paid teacher cover across all training sessions attended. Most teachers did not report additional costs incurred by schools from participating in the Focus4TAPS—only a few responses (n=2) reported additional costs such as

[48] The 25 trainers attended four to six training sessions, which are delivered every year, so they attended at different times. These training sessions consisted of three online sessions of 1.5 hours each, and three face-to-face sessions lasting six hours each—at total of 22.5 hours. The developer spent a total of 24 hours in training delivery, including time spent to and from the training venue. The developer also estimated travel costs of £50 per training group pair per day (£50 x 13 groups x 3 days = £1,950), room hire of £15 per person per day (£15 x 25 trainers x 3 days = £1,125), and catering costs of £15 per person per day (£15 x 25 trainers x 3 days = £1,125). This totals costs of £4,200 incurred by the developer to deliver the train-the-trainers sessions. The delivery of the training sessions to Focus4TAPS teachers was estimated to involve £50 per group of trainers to prepare training materials (£50 x 13 groups = £650).

with resources for investigations. However, as quantifiable amounts were not provided, these have not been included in these cost calculations.

- **Attending Focus4TAPS training**. Out of the 62 teachers who responded to the survey from Focus4TAPS schools, 44 responded that they attended the training in person on all three days, with seven responding that they attended some in person and watched training summaries otherwise. Five teachers reported spending an average of one hour additionally outside of working hours, watching the training summary videos. Around 23 teachers reported that paid cover was required for attending training days, with 12 teachers reporting that paid cover was not required.

- **Engaging with training materials outside of training.** Around 36 teachers (14 with Year 5 teacher roles, and 22 with science lead roles) reported spending additional time outside of training sessions to engage with training materials outside of training, either reviewing training materials or watching/rewatching training summary videos.

- **Using Focus4TAPS resources to develop teaching practice.** Around 35 teachers reported spending time outside their usual teaching responsibilities to develop their usual teaching practice. Out of these, five teachers reported spending additional time using Focus4TAPS resources to develop teaching practice on a regular basis, with the remaining using resources on a more ad hoc basis.

- **Delivering Focus4TAPS lessons**. Around 22 teachers with Year 5 teaching roles reported spending additional time delivering Focus4TAPS lessons beyond their usual teaching duties, which included attending lessons, running extra discussions, and setting up investigations.

- **Sharing Focus4TAPS learning and resources with other school staff.** Around 14 teachers reported not spending any additional time sharing Focus4TAPS learning with other school staff, with 37 teachers reporting doing so. This included activities outside of usual practice, such as formal or informal Focus4TAPS training sessions, staff discussions, or developing Focus4TAPS-related resources for the school. All teachers who reported spending additional time sharing Focus4TAPS learning did so on an ad hoc basis.

Table 28 outlines the costs reported by teachers with costs classified as start-up or recurring costs.

Table 28: Cost of delivering Focus4TAPS

Category	Role	Type of cost	No. of responses (n / N)	Reported time in hours – mean (SD)	Estimated total cost to school per year	Total cost to school over three years	Total cost per pupil per year over three years
Training costs for Focus4TAPS CPD	Non-staff cost	Start-up (Year one only)	Provided by delivery team		£600.00	£600.00	£8.70
Attending Focus4TAPS training: paid cover required for two teachers per school	All teachers	Start-up (Year one only)	22 / 62	33.4 (9.04)	X £33.71 = £1125.91	£1,125.91	£16.32
Engaging with training materials outside of training: watching video summaries and reviewing training materials)	Year 5 teachers	Start-up (Year one only)	14 / 31	3.29 (2.32)	X £29.51 = £97.09	£97.09	£1.41
	Science leads	Start-up (Year one only)	22 / 39	3.02 (2.20)	X £36.11 = £109.05	£109.05	£1.58
Using Focus4TAPS resources to	Year 5 teachers	Recurring (all three years)	16 / 31	3.34 (7.13)	X £29.51 = £98.56	£295.68	£4.29

develop teaching practice	Science leads	Recurring (all three years)	23 / 39	11.04 (27.52)[a]	X £36.11 = £398.65	£1,195.95	£17.33
Delivering Focus4TAPS lessons	Year 5 teachers	Recurring (all three years)	22 / 31	1.00 (0.60)	X £29.51 = £29.51	£88.53	£1.28
Sharing Focus4TAPS learning with other school staff	Year 5 teachers	Recurring (all three years)	17 / 31	1.29 (0.99)	X £29.51 = £127.14	£381.42	£5.53
	Science leads	Recurring (all three years)	26 / 39	1.77 (0.67)	X £36.11 = £63.91	£191.73	£2.78
Total					**£2,649.82**	**£4,085.36**	**£59.21**

Notes: Multipliers for the total costs to schools per year correspond to the average hourly pay for the teacher roles involved (see Table 11 in 'Methods' section). The total cost per pupil per year over three years is calculated assuming one participating Focus4TAPS Year 5 class of 23 pupils per school in each year. The programme may be delivered more efficiently where multiple Year 5 classes participate.

[a] Responses for this category were highly variable (median = 1.50 hours). In line with the EEF cost guidance (EEF, 2023), the mean is used in cost calculations.

The total costs of delivering Focus4TAPS in each of the three years are reported in Table 29 below. Overall, the per pupil per year cost of delivering Focus4TAPS is £59.21 (Table 28), which is a very low-cost programme (see Appendix A in this report).

Table 29: Cumulative costs of Focus4TAPS (assuming delivery over three years)

Programme	Year one	Year two	Year three
Focus4TAPS Cost per school per year	£2,649.82	£717.77	£717.77
Focus4TAPS Cost per pupil per year	£115.21	£31.21	£31.21

Conclusion

Table 30: Key conclusions

Key conclusions	
1.	Pupils in Focus4TAPS schools made no additional months' progress in science attainment, on average, compared to pupils in other schools. This result has a moderate to high security rating.
2.	Among pupils eligible for free school meals (FSM), those in Focus4TAPS schools made one month's additional progress in science attainment, on average, compared to those in other schools. However, there is uncertainty around these findings as the possible range of impact estimates include both a reduction and increase in pupil science attainment, and the results may have lower security than the overall findings because of the smaller number of pupils.
3.	Pupils in Focus4TAPS schools reported more positive attitudes and beliefs about science teaching and learning than pupils in other schools.
4.	Focus4TAPS training was implemented largely as intended, with consistent delivery of programme content. However, the number of lessons teachers delivered varied considerably, and, given the limited availability of data on this, it is not possible to determine whether more consistent lesson delivery would have led to improved attainment outcomes.
5.	Teachers viewed the programme positively and reported increased confidence in teaching science, particularly in relation to using formative assessment and developing pupils' Working Scientifically skills. Pupils were observed taking a more active role in scientific enquiry. However, the extent to which these practices were consistently embedded varied across schools.

Impact evaluation and IPE integration

In this section, we bring together the findings from the impact evalution and the IPE to provide an integrated interpretation of our results. We analyse the extent to which the emerging evidence aligned with the programme's logic model and reflect on what the findings suggested about how Focus4TAPS worked in practice. We also outline the limitations of the evaluation to support the interpretation of the results.

Evidence to support the logic model

We found mixed evidence to support the Focus4TAPS logic model. On one hand, the IPE provided strong evidence that many of the core inputs and activities, particularly the training, was implemented broadly as described in the logic model. The impact evaluation did not find conclusive evidence that Focus4TAPS improved science attainment but did find evidence of improvements in pupil attitudes to science. Perceptions of the short- and long-term benefits varied across schools. In the sections that follow, we will consider each stage of the logic model in turn.

Inputs and activities

The logic model outlined several core inputs for the delivery of Focus4TAPS, including three teacher training days, structured resources to support Working Scientifically, and ongoing support from both trainers and school science leads. These inputs were largely delivered as intended.

Observations and attendance data showed that the three training days took place as planned (see Table 26 for training attendance), and in their focus group discussions, teachers consistently described the training as helpful and relevant to their needs. The training covered core components of Working Scientifically, the structure and purpose of the Focused Assessment lessons, and approaches to formative assessment. In the survey and focus group data, many teachers reported that the training increased their confidence in planning and teaching types of enquiries within science lessons.

Training observations showed that trainers provided support to the teachers at the training days, and in the focus group discussions teachers reported that they valued opportunities during the training to ask questions and clarify elements of the Focus4TAPS approach. Trainers also provided ongoing support between training sessions. The logic model also suggested that science leads would support teachers in embedding Focus4TAPS in their schools. In practice, however, the level of involvement varied across schools. In interviews, trainers said that science leads in some schools took an active role—for example, attending training, coordinating implementation, and facilitating discussions—but that their involvement was less evident in other schools. As a result, while the core inputs were in place across participating schools,

the degree to which schools used the full range of Focus4TAPS support differed. The logic model also assumed that teachers would receive access to clear lesson plans and resources for delivering Focused Assessment lessons. We heard in the trainer interviews and teacher focus groups that these materials were provided as intended. Teachers said they made use of them when planning their lessons and described drawing on the example lesson structures and adapting activities to their own classrooms. The availability of lesson plans contributed to a better understanding of the Focus4TAPS approach, even though some variation remained in how the lessons were interpreted and implemented by different teachers. This was in line with the Focus4TAPS approach, which offered this flexibility and encouraged teachers to adapt lessons in ways that worked for their classes.

Outputs

The logic model specified several teacher-level outputs: attending and engaging with training; following and adapting Focused Assessment lesson plans; and delivering at least five lessons with each lesson focusing on a different element of Working Scientifically. Evidence suggested partial achievement of these outputs.

We observed that trainers delivered the Focus4TAPS training as intended. As shown in Table 26, teacher attendance remained high across all scheduled training days. In the survey data and focus group discussions, teachers reported engaging positively with the training content. Lesson observations indicated that teachers applied several elements of the approach in their classrooms, particularly in relation to using formative assessment and developing pupils' Working Scientifically skills. Many teachers used the materials and resources to plan lessons, and some adapted them to suit their pupils' needs or to integrate with existing schemes of work at their schools.

However, delivery of the minimum of five Focused Assessment lessons was less consistent. The Day 2 training survey, conducted at the midpoint of the intervention, indicated that a high number of teachers were beginning to implement Focused Assessment lessons in their schools. However, dosage data was not provided by all teachers at endline. Among the teachers who provided dosage data in the endline survey, only around half met this expectation. There were variations both in the number of lessons delivered and in teachers' interpretations of what constituted a Focused Assessment lesson. This suggested that, while teachers understood and valued the programme, they were not consistently able to deliver at the minimum expected level.

The IPE also found that several features of the training encouraged teachers to engage, including the quality of the face-to-face sessions, skilled trainers, and opportunities for peer support. Teachers' participation was at times limited by school-based pressures such as cover constraints, workload, and variable SLT buy-in. These conditions contributed to variation in how the programme was delivered across schools.

At the pupil level, the logic model suggested that pupils would participate in Focused Assessment lessons and engage more actively in enquiry. They would also increase their involvement in self-assessment and peer assessment. These outputs were observed in classrooms. Observations showed pupils recording their own data, discussing how to conduct investigations, and reflecting on their understanding. In the endline survey, teachers also reported an increase in pupils' self-assessment, though peer assessment was used less often. In the Focused Assessment lessons we observed, pupils appeared familiar with enquiry routines and confident in contributing ideas during science lessons. Teachers commonly identified focusing lessons on a single element of Working Scientifically, or having pupils record just one part of a practical activity, as key facilitators of pupil engagement.

The IPE evidence indicated that while most inputs were delivered and many intended outputs were achieved, there was variation in the extent of implementation across schools. This provides important context when interpreting the short-term and long-term outcomes of the Focus4TAPS programme.

Outcomes

There was mixed evidence for the outcomes in the Focus4TAPS logic model. The logic model proposed that Focus4TAPS would lead to improvements in teachers' planning for formative assessment, their understanding of progression in Working Scientifically, and their ability to adapt instruction based on pupils' needs. There is evidence that these short-term outcomes were achieved. In teacher focus groups and through trainer interviews, teachers reported feeling more confident in using formative assessment and in structuring lessons around a specific enquiry skill. Survey findings indicated that most

Year 5 teachers and science leads 'agree' that they used formative assessment to plan future lessons and adjust the level of support and challenge. Lesson observations supported this. Teachers were observed prompting pupils to reflect on their thinking, involving them in decisions about investigations, and using questioning to improve their understanding.

For pupils, the logic model expected greater engagement in enquiry, increased self-assessment and peer assessment, and improved confidence in Working Scientifically. The IPE found evidence of greater engagement among pupils where pupils were actively recording their own data, planning aspects of investigations, and contributing ideas during discussions. Our survey data indicated that there was an increase in the frequency of self-assessment among pupils. Comparatively, peer assessment was used less frequently. The impact evaluation found positive effects on pupils' engagement with science across most secondary outcomes capturing pupils' attitudes towards science. Specifically, there was evidence of positive intervention effects on six subscales around attitudes towards science, including: interest and enjoyment in science; confidence in science; perceptions of science teachers' teaching; Working scientifically beliefs, self-regulation of learning in science; and self-efficacy in Working Scientifically; though no evidence was found for attitudes related to wider benefits of science. These patterns suggested that some short-term changes to pupils' engagement and attitudes to science occurred.

While the evaluation found some evidence of progress against short-term logic model outcomes, this did not extend to the longer-term outcome of improved science attainment. The impact evaluation found no conclusive evidence that Focus4TAPS had an impact on science attainment. The associated effect size was very small (0.03 SD; 95% CI: -0.11 to 0.17), equivalent to zero months' additional progress. The estimated range of effects included zero (no impact) as well as small negative and small positive values. As a result, we did not find conclusive evidence that participation in Focus4TAPS improved science attainment scores. Given baseline imbalances between groups in FSM eligibility and school-level Key Stage 1 reading and mathematics attainment, a robustness check that adjusted for these differences and included an alternative specification of Key Stage 1 reading and maths attainment, in line with the efficacy trial, produced a larger effect estimate equivalent to approximately one month's additional progress in science attainment. The range of possible impacts was consistent with both negative and positive effects, and the findings were inconclusive.

For pupils eligible for FSM, the impact evaluation did not find conclusive evidence that Focus4TAPS improved science attainment scores. The estimated effect size was small (0.07 SD; 95% CI: -0.12 to 0.25), which equated to one month's additional progress. The range of effects included zero (no impact) as well as small negative and small positive values. As a result, there was no conclusive evidence that FSM eligibility moderated the impact of Focus4TAPS on science attainment. The impact evaluation findings did not provide conclusive evidence of a variation in the effect of Focus4TAPS on the secondary outcomes for any of the other subgroups considered in this evaluation.

The impact evaluation findings aligned with the IPE evidence. Teachers reported integrating several core elements of Focus4TAPS into their lessons, with IPE findings suggesting that these were generally implemented as intended at the classroom level. However, implementation varied more widely in relation to lesson-level dosage and the extent to which practices were embedded beyond the Focused Assessment lessons and/or across the wider curriculum. While teachers and pupils described positive changes in classroom interactions and learning, these changes did not translate into measurable improvements in science attainment during the trial period. The lack of variation in impact across subgroups (except where pupils had high prior interest in science) further suggests that the conditions necessary for long-term impact were not met.

Overall, the findings provide mixed support for the outcomes outlined in the logic model. Several short-term outcomes were achieved, including increases in teachers' confidence, use of formative assessment, and pupils' active engagement in scientific enquiry. However, longer-term anticipated outcomes like improved attainment in science were not observed. One possible interpretation is that the inputs and activities specified in the Focus4TAPS logic model do not, in themselves, lead to the intended longer-term impacts. The evidence in the IPE data of short-term changes that align with the logic model suggests, however, that the underlying theory of change may apply. Stronger conditions (including clearer expectations of how lessons should be delivered and embedded in practice) and more sustained delivery may be required for longer-term anticipated effects to be achieved.

Contextual moderating factors

Several contextual factors influenced the extent to which Focus4TAPS was implemented and the degree to which outcomes were achieved. Differences in schools' existing approaches to science teaching shaped how readily Focus4TAPS practices could be integrated. Schools with more established structures for planning and teaching science (such as regular science meetings or clear links to the National Curriculum) appeared better able to embed the Focused Assessment approach. In such settings, teachers described stronger alignment between Focus4TAPS and their existing routines. Teachers in schools with more protected time for science also said they found it easier to deliver the intended sequence of Focused Assessment lessons. The role of the science lead also varied across schools and functioned as an important contextual moderator. In schools where science leads attended training and actively supported teachers, teachers reported greater confidence in applying the approach. In some schools, staffing changes also contributed to a lack of continuity in delivery as some teachers who attended the delivery were unable to continue with the programme and implement the Focus4TAPS lessons because of staff changes.

Teacher focus groups also indicated that engaging with Focus4TAPS improved teachers' understanding of the Working Scientifically curriculum, and that teachers who observed positive changes in pupil engagement during science lessons reported feeling more motivated to continue using the approach. This suggests that early, visible impacts on pupils may help sustain teacher engagement with Focus4TAPS over time.

Interpretation

This evaluation aimed to determine whether Focus4TAPS improved Year 5 pupils' science attainment and attitudes, and to understand how the programme was delivered in schools. Interpreting the findings required integrating evidence from both the impact evaluation and the IPE and linking the results within the wider evidence base on formative assessment and enquiry-led primary science.

The primary hypothesis for this evaluation was that participation in Focus4TAPS would lead to measurable gains in pupils' science attainment. The impact evaluation did not find evidence of such gains. The effect size was very small, and the CI included both negative and positive values, so that a true 'null' impact could not be ruled out. As a result, the findings do not support the conclusion that Focus4TAPS improves science attainment. Small improvements were observed in pupils' engagement with science across several secondary outcomes, although these changes were modest. The evaluation did not find conclusive evidence of an impact on science attainment or pupils' attitudes to science among pupils eligible for FSM.

To understand why these outcomes occurred, it is important to consider the intended mechanisms within the logic model. The logic model assumed that teachers' participation in high-quality training would lead to changes in planning, assessment, and lesson structure. This would then give pupils more opportunities to develop Working Scientifically skills. These opportunities would build over time and when delivered consistently across classrooms, were expected to generate improvements in pupils' understanding and science attainment.

The IPE findings showed that several early outcomes were achieved. Teachers reported increased confidence in formative assessment and Working Scientifically, and many teachers adopted practices such as focusing lessons on single enquiry skills and supporting pupils to record and interpret data. Pupils were observed taking a more active role in enquiry, especially in self-assessment and planning aspects of investigations. These findings suggested that the programme successfully influenced elements of science teaching.

However, these changes did not appear to operate at a level which could influence learning outcomes within the timescale of the trial. Two issues in particular can help explain the limited impact:

- **Variation in lesson dosage.** The logic model assumed that pupils would experience a minimum of five structured Focused Assessment lessons, each concentrating on a different enquiry skill. Only around half of teachers who reported dosage reached this threshold. Moreover, teachers varied in how they interpreted what constituted Focused Assessment lessons. Without repeated opportunities to try out their enquiry skills, improved attainment is likely to have been weakened.

- **Overlap with usual practice.** Another factor that might help explain the lack of measurable gains is the degree of overlap between Focus4TAPS and usual practice in many schools. The IPE indicated that several elements of the approach such as encouraging pupils to record their own data, engaging them in aspects of enquiry, and using questioning to assess understanding were already features of science teaching in some Focus4TAPS schools before the programme began. Additionally, some schools were already using PSTT resources (which includes TAPS resources) in their science teaching prior to the trial. This reduced contrast between Focus4TAPS and usual practice schools, which was not present to the same extent in the previous efficacy trial, may also help explain the absence of detectable impacts on attainment within the trial period. Where teachers delivered fewer Focused Assessment lessons, the distinctiveness of Focus4TAPS relative to existing practice might have been further reduced.

Overall, the IPE data suggest that the programme began to influence teacher behaviour and pupil engagement, but that these changes were not sufficient to produce measurable improvements in attainment in the time period of the evaluation, which may be explained by limited or partial implementation in classrooms. This interpretation is consistent with the impact evaluation findings, which found no evidence of an effect on science attainment but identified positive changes in several secondary outcomes related to pupils' engagement with science.

The findings from this evaluation matched the wider evidence base. The Teaching and Learning Toolkit identified formative assessment as a high-impact approach (EEF, 2021). However, studies emphasised that effects depend heavily on consistent implementation and teacher expertise (Black and Wiliam, 2009; Heritage, 2010). Similarly, evidence on enquiry-based science teaching suggested that conceptual gains arise when enquiry routines are used regularly and with clear instructions (Minner *et al.,* 2010; Furtak *et al.,* 2012; Lazonder and Harmsen, 2016). The findings from Focus4TAPS aligned with this wider evidence. Teachers engaged positively with the approach, but variation in implementation, with some teachers delivering less than the intended number of lessons, could have limited the programme's impact on attainment.

The previous Focus4TAPS efficacy trial reported positive impacts on pupil attainment. Efficacy trials typically involve more controlled conditions and more intensive monitoring. In the efficacy trial, teachers received more structured guidance, dosage was more consistent, and science leads played a more active role in supporting delivery. Under these conditions, the mechanisms in the logic model were more fully achieved. In the present effectiveness trial, schools experienced greater autonomy, and the programme was delivered in a wider range of contexts, with more variability in teacher experience and science lead roles. These differences likely contributed to variation in implementation between schools, where some did not deliver the intended minimum dosage of lessons for the programme or did not make full use of the Focus4TAPS resources within their lessons. Further information on these differences is outlined in Appendix C in this report.

Although the trial did not find evidence of an impact on science attainment, the evaluation showed that Focus4TAPS has the capacity to support teachers in strengthening enquiry-led science teaching and formative assessment practice. Teachers responded positively and many were able to implement key elements of the approach, with pupils demonstrating increased engagement in enquiry and self-assessment. These early mechanisms provide a promising foundation for future development. With clearer expectations, more consistent dosage, and strengthened in-school support, Focus4TAPS may be well placed to achieve stronger and more sustained outcomes as it continues to evolve.

Limitations and lessons learned

As with any large-scale effectiveness trial, there are limitations to the methods and data, and lessons to be learned. In this section, we outline key limitations relating to attrition and statistical power, suitability of outcome measures, and missing data across elements of the evaluation, before summarising implications for future delivery and research.

The trial experienced a high level of attrition, exacerbated by school withdrawals and the high number of pupils that did not complete the endline science attainment test. Pupil-level attrition from randomisation to analysis was 21.95% for the Focus4TAPS group and 20.35% for the teaching as usual group. The overall attrition across both groups was 21.13%, representing more than 20% of the randomisation sample not included in analysis. This was higher than in the previous Focus4TAPS efficacy trial, where overall attrition was 12.8%. The cost evaluation also faced limitations due to a relatively low response rate, with some item non-response (and low response rates). Estimates of staff time and cover costs should

therefore be interpreted with some caution, as responses may have been influenced by schools' internal capacity to supply financial information or by research burden during the delivery period.

As with other primary science assessments, the outcome measure may not fully capture subtle improvements in pupils' enquiry processes or reasoning skills targeted by Focus4TAPS. Furthermore, the implementation period was limited to a single academic year, and formative assessment approaches are known to require consistent use to translate into measurable gains in attainment. It is therefore, possible that impacts may have emerged over a longer time horizon.

Some teachers also reported that the science attainment measure used in the evaluation was challenging for many pupils and focused on science very broadly, which might have affected how well it captured the specific enquiry skills targeted by the programme.

The IPE also had limitations. There was missing data for lesson dosage across all schools, which reduced the certainty with which patterns in dosage and lesson delivery could be interpreted. Compliance with the intervention could only be partially assessed due to low response rates to the endline teacher survey. This limited the extent to which definitive conclusions could be made about whether higher levels of compliance with Focus4TAPS were associated with higher science attainment scores. In addition, elements of the IPE such as focus groups, surveys, and lesson reflections relied on self-reported data, which may be subject to bias. Although the qualitative fieldwork gathered valuable insights, the research team were only able to speak with a selected number of teachers and science leads. It is therefore, possible that the perspectives captured reflect those more engaged with the programme, and that other views, especially from teachers with lower levels of engagement, were underrepresented.

Despite these limitations, the evaluation generated important learning about both the programme and its implementation. First, the findings highlighted the importance of clearer delivery expectations and a need for more explicit guidance on dosage and lesson structure. Second, the evaluation indicated that monitoring dosage more systematically for the lessons would help ensure pupils receive the intended sequence of science learning. Finally, the findings underline the need to minimise research burden for schools wherever possible to improve response rates and ensure that a broad range of perspectives contribute to both quantitative and qualitative evidence.

Future research and publications

The evaluation highlights several areas for future research. Studies could examine whether Focus4TAPS has a greater impact when implemented over a longer period, as formative assessment and enquiry routines may require sustained use before influencing attainment. Future research could draw on Key Stage 2 science outcomes from the NPD to explore whether impacts on science attainment emerge over time. Research could explore how clearer delivery expectations and stronger science lead involvement affects implementation quality, given their potential role as moderators of impact in this trial.

While the evaluation captured pupils' attitudes and beliefs about Working Scientifically as secondary outcomes, future evaluations could consider analysing the Working Scientifically elements of the science attainment assessment specifically, where possible, to provide a more direct measure of scientific enquiry skills.

There are no planned additional publications for this trial.

References

Archer, L. Moote, J., Francis, B., DeWitt, J. and Yeomans, L. (2017) 'The "Exceptional" Physics Girl: A Sociological Analysis of Multimethod Data From Young Women Aged 10–16 to Explore Gendered Patterns of Post-16 Participation'. *American Educational Research Journal*, 54: 1, 88–126. https://doi.org/10.3102/0002831216678379

Black, P. and Wiliam, D. (2009) 'Developing the Theory of Formative Assessment'. *Educational Assessment, Evaluation and Accountability*, 21: 1, 5–31. https://doi.org/10.1007/s11092-008-9068-5

Carpenter, J.R., Kenward, M.G. and White, I.R. (2007) 'Sensitivity Analysis After Multiple Imputation Under Missing at Random: A Weighting Approach'. *Statistical Methods in Medical Research*, 16: 3, 259–275. https://doi.org/10.1177/0962280206075303

Davies, D.J., Earle, S., McMahon, K., Howe, A. and Collier, C. (2017) 'Development and Exemplification of a Model for Teacher Assessment in Primary Science'. *International Journal of Science Education*, 39: 14, 1869–1890. https://doi.org/10.1080/09500693.2017.1356942

Davies, D., and McMahon, K. (2003) 'Assessment for Enquiry: Supporting Teaching and Learning in Primary Science'. *Science Educational International,* 14: 4, 29–39. www.icaseonline.net/sei/14-4-2003/14-4-2003-29_39.pdf

Dong, N. and Maynard, R. (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

Earle, S. (2021) 'Principles and Purposes of Assessment in the Classroom'. *Impact*, 12. 20–23. https://my.chartered.college/impact_article/principles-and-purposes-of-assessment-in-the-classroom/

Education Endowment Foundation (EEF). (2019) *'Classification of the Security of Findings From EEF Evaluations'.* London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/peer-review-process/Classifying_the_security_of_EEF_findings_2019.pdf?v=1786005272 (Last accessed 27 08 2026).

Education Endowment Foundation (EEF). (2022) '*Statistical Analysis Guidance for EEF Evaluations'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1786005272 (Last accessed 27 08 2026).

Education Endowment Foundation (EEF). (2023) *'Cost Evaluation Guidance for EEF Evaluations'.* London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1765786924 (Last accessed 27 08 2026).

Education Endowment Foundation (EEF). (2021) 'Teaching and Learning Toolkit: Feedback'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/feedback (Last accessed 27 08 2026).

Estyn. (2022) *'Effective Approaches to Assessment that Improve Teaching and Learning'.* Cardiff: Estyn. Available at www.estyn.gov.wales/system/files/2022-10/Effective%20approaches%20to%20assessment%20that%20improve%20teaching%20and%20learning.pdf (Last accessed 27 08 2026).

Furtak, E.M., Seidel, T., Iverson, H. and Briggs, D.C. (2012) 'Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis'. *Review of Educational Research*, 82: 3, 300–329. https://doi.org/10.3102/0034654312457206

General Data Protection Regulation (GDPR). (2016) *'Council Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (United Kingdom General Data Protection Regulation) (Text with EEA relevance)'.* Available at: www.legislation.gov.uk/eur/2016/679 (Last accessed 27 08 2026).

Hedges, L.V. (2007) 'Effect Sizes in Cluster-Randomized Designs'. *Journal of Educational and Behavioural Statistics*, 32: 4, 341–370. https://doi.org/10.3102/107699860629804

Hedges, L.V. (2011) 'Effect Sizes in Three-Level Cluster-Randomized Experiments'. *Journal of Educational and Behavioral Statistics*, 36: 3, 346–380. https://doi.org/10.3102/1076998610376617

Hemming, K. and Taljaard, M. (2021) 'Why Proper Understanding of Confidence Intervals and Statistical Significance is Important'. *Medical Journal of Australia*, 215: 4, 191–191.e1. https://doi.org/10.5694/mja2.51186

Heritage, M. (2010) 'Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity?'. Washington, DC: Council of Chief State School Officers. Available at: https://eric.ed.gov/?id=ED543063 (Last accessed 27 08 2026).

Hondrich, A.L., Hertel, S., Adl-Amini, K. and Klieme, E.(2016) 'Implementing Curriculum-Embedded Formative Assessment in Primary School Science Classrooms'. *Assessment in Education: Principles, Policy and Practice*, 23: 3, 353–376. https://doi.org/10.1080/0969594X.2015.1049113

Imbens, G.W. and Angrist, J.D. (1994) 'Notes and Comments: Identification and Estimation of Local Average Treatment Effects'. *Econometrica,* 62; 2, 467–475. www.jstor.org/stable/2951620

Joshi, K. Burke, L., Elliott, L., Fountain, I., Hanley, P., Haynes, A., Keding, A., Parvin, J., Waller, N. and Winter, J. (2021) *'Development of a Year 5 Science Assessment for the EEF'*. London: Education Endowment Foundation. Available at: (Last accessed 27 08 2026).

Lazonder, A.W. and Harmsen, R. (2016) 'Meta-Analysis of Inquiry-Based Learning: Effects of Guidance'. *Review of Educational Research*, 86: 3, 681–718. https://doi.org/10.3102/0034654315627366

Leonard, H., Poquillon, G., Bhatti, S., Morgan, H., Benny, L. and Wadsworth, K. (2024) *'Focus4Taps: Evaluation Protocol'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/focus4taps-2024-25-trial (Last accessed 27 08 2026).

Leonard, H., Poquillon, G., Bhatti, S., Parker, R., Benny, L. and Wadsworth, K. (2025) *'Evaluation of Focus4TAPS: Statistical Analysis Plan'.* Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/focus4taps-2024-25-trial (Last accessed 27 08 2026).

Luxton, K. and Pritchard, B. (2023) 'Improving Primary Science: Guidance Report'. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/primary-science-ks1-ks2 (Last accessed 27 08 2026).

McCormick, B. (2025) *'The State of Primary Science in the UK (2025)'*. Bristol: Primary Science Teaching Trust. Available at: https://pstt.org.uk/theory-practice/state-of-primary-science-in-the-uk-2025/ (Last accessed 27 08 2026).

Minner, D.D., Levy, A.J. and Century, J. (2010) 'Inquiry-Based Science Instruction—What Is It and Does It Matter? Results From a Research Synthesis Years 1984 to 2002'. *Journal of Research in Science Teaching*, 47: 4, 474–496. https://doi.org/10.1002/tea.20347

Mujtaba, T. and Reiss, M.J. (2014) 'A Survey of Psychological, Motivational, Family and Perceptions of Physics Education Factors that Explain 15-Year-Old Students' Aspirations to Study Physics in Post-Compulsory English Schools'. *International Journal of Science and Mathematics Education*, 12: 2, 371–393. https://doi.org/10.1007/s10763-013-9404-1

Mujtaba, T., Sheldrake, R., Reiss, M.J. and Simon, S. (2018) 'Students' Science Attitudes, Beliefs, and Context: Associations with Science and Chemistry Aspirations', *International Journal of Science Education*, 40: 6, 644–667. https://doi.org/10.1080/09500693.2018.1433896

Mutjaba, T., Sheldrake, R., Hodgen, J. and Reiss, M.J. (2022) *'Focus for Teacher Assessment of Primary Science (Focus4TAPS) Evaluation Report'.* Available at https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Focus4TAPS-Report.pdf?v=1689488901 (Last accessed 27 08 2026).

Office for Standards in Education, Children's Services and Skills (Ofsted). (2021) *Research Review Series: Science*. GOV.UK. Available at: www.gov.uk/government/publications/research-review-series-science/research-review-series-science (Last accessed 27 08 2026).

Primary Science Teaching Trust (PSTT). (2022) 'Working Scientifically Progression'. Bristol: Primary Science Teaching Trust. Available at: https://pstt.org.uk/resources/enquiry-skills(Last accessed 27 08 2026).

Richardson, M., Isaacs, T., Barnes, I., Swensson, C., Wilkinson, D. and Golding, J. (2020) *'Trends in International Mathematics and Science Study (TIMSS) 2019: National Report for England'.* London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/5fca467ad3bf7f5d09db26ae/TIMSS_2019_National_Report.pdf (Last accessed 27 08 2026).

Ritchie, J., Lewis, J., Nicholls, C.M. and Ormston, R. (2014) *'Qualitative Research Practice: A Guide for Social Science Students and Researchers. Second Edition'.* London: SAGE Publications Ltd.

Romano, J.P. and Wolf, M. (2005) 'Stepwise Multiple Testing as Formalized Data Snooping'. *Econometrica*, 73: 4, 1237–1282. https://doi.org/10.1111/j.1468-0262.2005.00615.x

Romano, J. P., & Wolf, M. (2016). 'Efficient computation of adjusted p-values for resampling-based stepdown multiple testing'. *Statistics & Probability Letters*, 113, 38-40. https://doi.org/10.1016/j.spl.2016.02.012

Wellcome Trust. (2017) *'The Science Content in Primary Initial Teacher Training (ITT)'.* London: Wellcome Trust. Available at https://cms.wellcome.org/sites/default/files/science-content-in-primary-initial-teacher-training.pdf(Last accessed 27 08 2026).

Wiliam, D. (2011) 'What is Assessment for Learning?'. *Studies in Educational Evaluation,* 37: 1, 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Appendix A: The EEF cost rating

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: Security classification of trial findings

OUTCOME: Year 5 *science attainment*

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%	**3**				**3**
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%			Adjustment for threats to internal validity **[0]**		
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	MODERATE	Randomisation at school-level with stratification by region and prior Key Stage 1 attainment. Baseline imbalances exceeding 0.05 SD were adequately controlled for via sensitivity analyses.
Threat 2: Concurrent interventions	MODERATE	Schools were excluded from participating if they were in recent science trials. However, there appears to be overlap in the approaches used by control schools and intervention schools, in that control schools used similar resources and formative assessments.
Threat 3: Experimental effects	LOW	Control schools practice broadly stable across the trial period.
Threat 4: Implementation fidelity	MODERATE	Only around half the teachers who reported dosage data (which was limited in itself—37% of intervention schools returned this) delivered the minimum five Focused Assessment lessons. Training attendance was high, but lesson delivery fidelity was insufficient and variable.
Threat 5: Missing data	MODERATE	Overall attrition was high—above 20% threshold but were similar between arms. Missingness was associated with baseline characteristics which means it is systematic (related to disadvantage-related characteristics)—but dealt with by authors appropriately using multiple imputation (and yielded comparable results to complete-case analysis).

Threat 6: Measurement of outcomes	LOW	The primary outcome of science attainment was previously validated and used in prior efficacy trial, with good internal reliability. Blinding and independent arbiter were used. Baseline of this measure was only a proxy (Key Stage 1 reading/maths) and measured also with a considerable gap in time.
Threat 7: Selective reporting	LOW	Trial was pre-registered on the Open Science Framework registry. SAP and evaluation protocol were published openly. The report transparently documents deviations from the efficacy trial and documents all pre-specified exploratory analyses.

- **Initial padlock score:** 3 padlocks. The design is a two-arm cluster randomised effectiveness trial with school-level randomisation, stratified by region and prior Key Stage 1 attainment, with analyses at the pupil level. The MDES at randomisation stage is 0.173, which falls within the ≤0.2 band. Overall attrition is 21.13%, which shifts the padlock score from 5 to 3.

- **Reason for adjustment for threats to validity**: No adjustments. 'Confounding factors' are classed as a moderate risk given baseline imbalances in pupil prior attainment between 0.05 and 0.1 SD. 'Concurrent interventions' is classed as a moderate risk based on evidence of comparable approaches in the control group. 'Implementation fidelity' is classed as a moderate risk as only ~half of reporting teachers met minimum dosage but extent of bias unknown given delivery data was missing for most intervention schools. 'Missing data' is classed as a moderate risk based on analysis that suggests missingness was associated with pupil-disadvantaged characteristics. The overall direction of bias for these four moderate threats to validity are unknown, so no further adjustments are applied.

- **Final padlock score:** 3 padlocks.

Appendix C: Changes since the previous evaluation

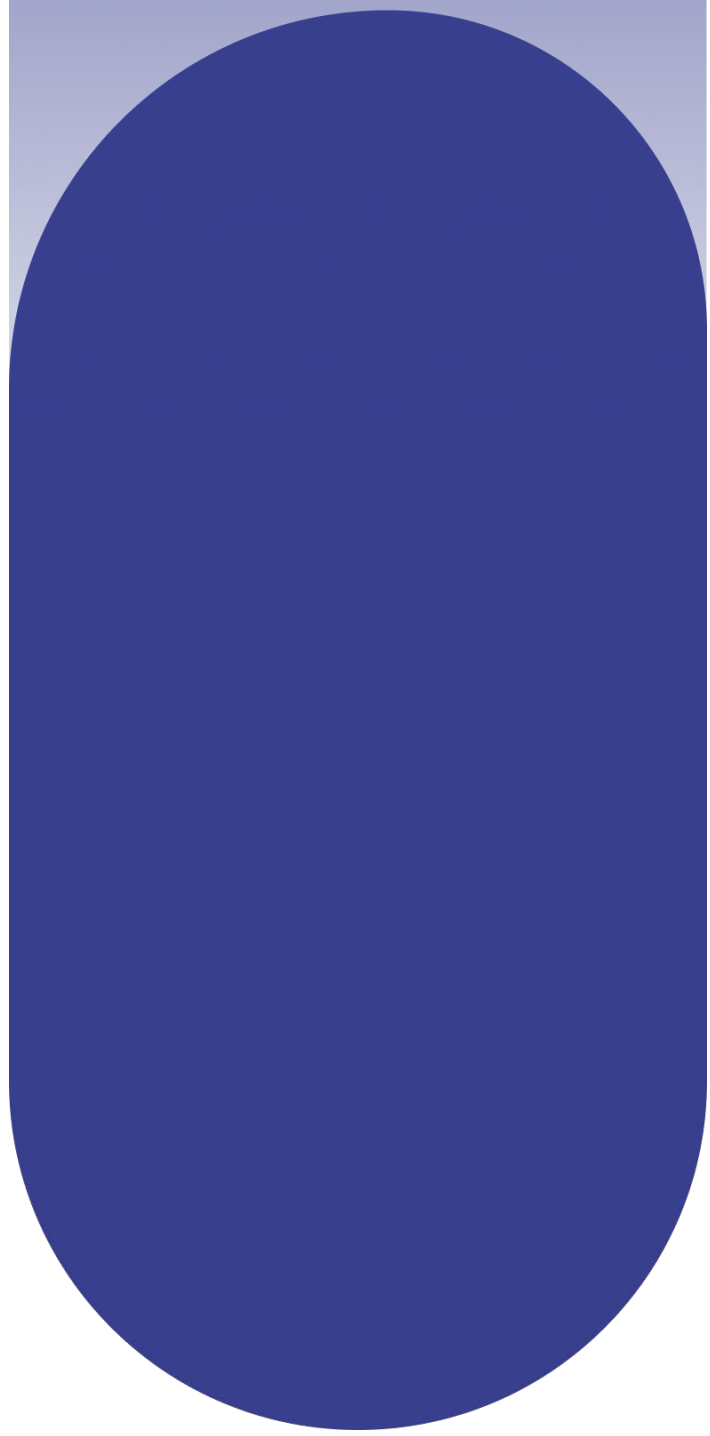
Table 1: Changes since the previous evaluation

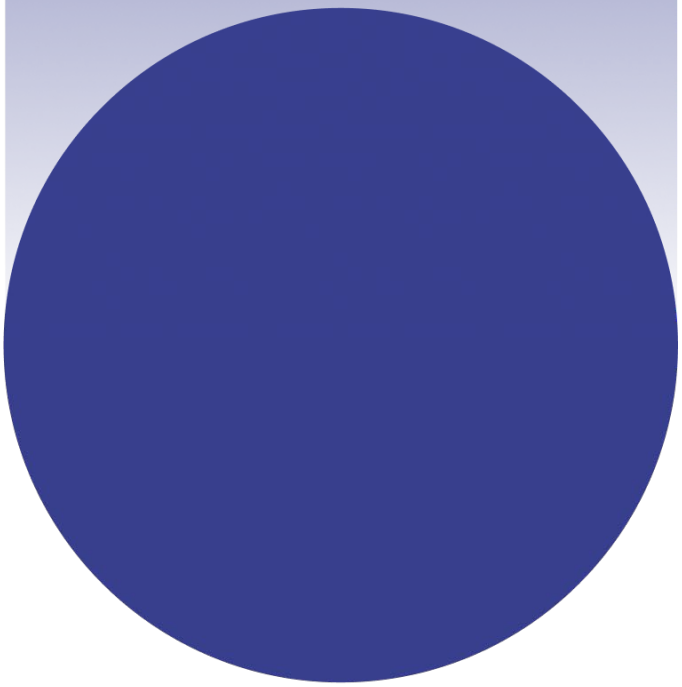
	Feature	Efficacy to effectiveness stage
Intervention	Intervention content	The content of the Focus4TAPS training workshops remained largely consistent between efficacy and effectiveness trials. Small changes were made to ensure discussions highlighted the programme's alignment with the most recent guidance for primary science teaching practice, however. This included, for example, government guidance such as the EEF's 'Improving Primary Science' report (Luxton and Pritchard, 2023).
	Delivery model	There was a move from direct delivery by the programme developer (in the efficacy trial) to a train-the-trainer model (in the effectiveness trial). Trainers were recruited and trained by the developer and then delivered the three workshop days across 13 regions. Training in the efficacy trial was originally delivered face-to-face, and most schools attended the first two days in person before Covid-19 restrictions closed schools, requiring the trial to be restarted the following year. At this point, training shifted to online 60-minute sessions delivered by the developer. In the effectiveness trial, training shifted to three full-day, in-person workshops with practical activities.
	Intervention duration	Both trials delivered Focus4TAPS over a single academic year, with teachers expected to attend three training sessions and deliver a series of Focused Assessment lessons during that same year.
Evaluation	Eligibility criteria	Both trials applied the same eligibility criteria. Schools were required to have a cohort of pupils progressing from Year 4 to Year 5 who would participate in the trial. Schools already participating in similar science initiatives were not eligible to take part. Teacher training shifted from 60-minute online sessions in the efficacy trial to three full-day, in-person workshops in the effectiveness trial. As a result, teachers' ability to attend in-person training in one of the participating regions was included in the eligibility criteria for the effectiveness trial.
	Level of randomisation	In both trials, randomisation was conducted at the school level.
	Outcomes and baseline	The primary outcome measure for both trials was science attainment, and the secondary outcome was pupils' attitudes and beliefs towards science endline subscale scores. For the primary analysis, baseline measures were the same for both trials: Key Stage 1 mathematics and reading attainment. In the efficacy trial, these were categorical (BLW, PFK, WTS, EXS, and GDS), whereas in the effectiveness trial they were binary indicators of whether pupils were working below or at/above the expected standard. For secondary analyses, baseline measures changed from Key Stage 1 reading and mathematics attainment to attitudes and beliefs towards science subscale scores at baseline.
	Control condition	In both trials, a teaching as usual condition was used.

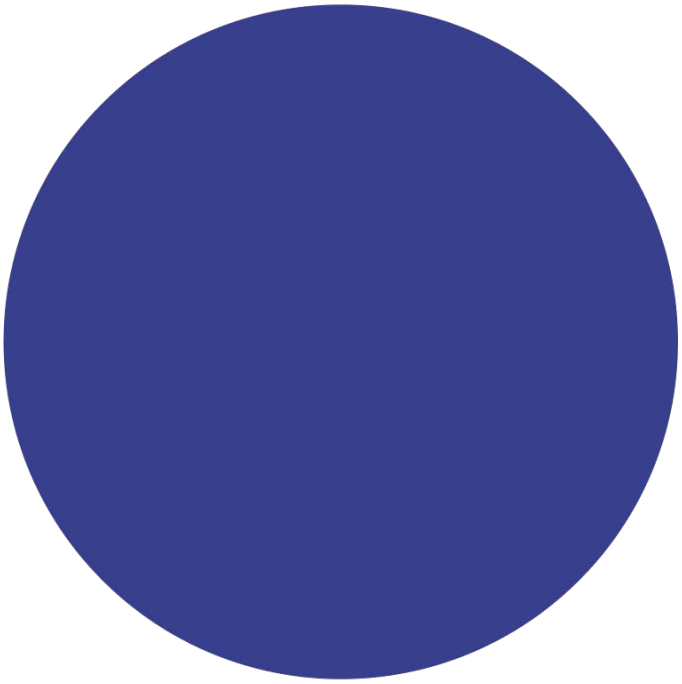
Further Appendices

Appendix D – M can be found published in a separate document on the project page named '*Further Appendices*'.

This document is available for download at https://educationendowmentfoundation.org.uk







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