Focus for Teacher Assessment of Primary Science (Focus4TAPS)

Evaluation Report

December 2022

Tamjid Mujtaba, Richard Sheldrake, Jeremy Hodgen,
Michael J. Reiss

Funded by

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 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 re-endowed with an additional £137m, 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
SW1P 4QP

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The Focus for Teacher Assessment of Primary Science (Focus4TAPS) trial was independently evaluated by a team from IOE, UCL's Faculty of Education and Society.

IOE is UCL's Faculty of Education and Society. IOE was previously called the UCL Institute of Education. The project was independently evaluated by a team from UCL (Dr Tamjid Mujtaba, Dr Richard Sheldrake, Professor Jeremy Hodgen, and Professor Michael J. Reiss).

The principal investigator for this trial was Dr Tamjid Mujtaba.

Contact details:

Name: Dr Tamjid Mujtaba
Address: IOE, UCL's Faculty of Education and Society,
University College London,
20 Bedford Way,
London, WC1H 0AL,
United Kingdom
Email: t.mujtaba@ucl.ac.uk

Acknowledgements

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

Executive summary

The project

The Focus for Teacher Assessment of Primary Science (Focus4TAPS) programme aims to improve science attainment for pupils in primary schools, by improving teaching approaches and assessment in science. It aims to support teachers to gain an enhanced understanding of progression in science, to apply formative assessment strategies, and to adapt levels of support and challenge during lessons across an academic year. In this trial, science subject leaders and Year 5 teachers were trained, and the outcomes of Year 5 pupils (aged around 9–10 years old) were analysed.

The programme was delivered during the 2020/2021 academic year, and involved six training sessions spread over the year. Each session was delivered remotely to groups of teachers and had a follow-up video. Additional content discussed how to carry out practical work given COVID-19 arrangements and restrictions in schools. Teachers were expected to deliver science lessons guided by the Teacher Assessment in Primary Science (TAPS) materials.

The evaluation was an efficacy randomised controlled trial involving 121 schools and 2,882 pupils. The process evaluation included teacher questionnaires and teacher and pupil interviews. This project and its evaluation were affected by the 2020 and 2021 partial school closures caused by the COVID-19 pandemic. The trial originally started in the academic year 2019/2020 but had to stop due to partial school closures. The trial restarted with a new cohort of Year 5 pupils in academic year 2020/2021. Pupils completed tests and questionnaires during June 2021 and July 2021.

The trial was developed and delivered by Bath Spa University and draws together research from the TAPS project, which was funded by the Primary Science Teaching Trust. The trial was co-funded by Wellcome as part of the Education Endowment Foundation (EEF) themed round on improving science education.

Table 1: Key conclusions

Key Conclusions
Pupils in Focus for Teacher Assessment of Primary Science (Focus4TAPS) schools made the equivalent of two additional months' progress in science, on average, compared to pupils in other schools. This result has a high security rating.
Pupils eligible for Free School Meals (FSM) in Focus4TAPS schools made the equivalent of two additional months' progress in science, on average, compared to pupils eligible for FSM in other schools. This result is based on smaller numbers of pupils and may need to be interpreted with caution.
Pupils in Focus4TAPS schools had similar, and positive, attitudes and orientations towards science to pupils in other schools.
The programme ran as expected, with 80% of teachers attending the minimum number of training sessions and 63% of teachers conducting at least four 'Focused Assessment' lessons using Focus4TAPS materials. Greater compliance with the programme (e.g. attending more training sessions and conducting more lessons) was associated with higher science test scores.
Teachers in Focus4TAPS schools reported higher confidence for some aspects of teaching and assessing science than teachers in other schools, and believed that various benefits followed from the programme, e.g. changing how they taught Working Scientifically and how they applied formative assessment.

EEF security rating

These findings have a high security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools.

Additional findings

Pupils in Focus4TAPS schools made, on average, two additional months' progress in science compared to pupils in other schools. This is our best estimate of impact, which has a high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes no additional progress and positive effects of up to 4 months of additional progress.

The programme assumed that delivering more and refined practical science may be positively received by pupils. Surveys showed that teachers believe that pupils tended to enjoy practical work and science in general, although there were no observable differences across the pupils within the intervention and control schools, with both displaying positive attitudes and orientations towards science.

The programme broadly ran as intended following an adapted design of online session due to the pandemic. Teachers from 80% of schools attended at least three of the six training sessions, which was the minimum expected. Almost two-thirds of schools (63%) completed at least four 'Focused Assessment' class lessons using Focus4TAPS lesson plans. Many Year 5 teachers within the intervention group provided narrative responses in the survey that showed an understanding of the Focus4TAPS principles. Analysis found that higher levels of compliance with the Focus4TAPS programme (e.g. attending more training sessions and conducting more lessons) was associated with higher science test scores.

When surveyed at the end of the academic year, the teachers in the intervention group reported, on average, higher confidence than those in the control group for many aspects of science assessment and Working Scientifically. Majority of Year 5 teachers within the intervention schools conveyed that the Focus4TAPS programme had specifically changed how they taught Working Scientifically and how they applied formative assessment. This supports the hypothesis that the intervention enhances teachers' practices, as reflected through their reported confidence.

The programme intended to refine science teaching rather than increase the extent of teaching/learning. However, pupils and teachers within intervention schools reported experiencing or delivering more science teaching and learning compared to those in control schools. It is possible that the Focus4TAPS programme may have indirectly inspired more science teaching and learning and directly facilitated this through aspects of the training. Further analysis revealed that pupils in Focus4TAPS schools had higher attainment even when they received the same amount of science teaching as pupils in other schools. This suggests that the difference in test scores between Focus4TAPS schools and other schools is due to the different teaching and learning that happened in Focus4TAPS schools rather than the extent of teaching and learning.

The circumstances of the trial may complicate how the findings may generalise to other situations. Majority of Year 5 teachers in intervention schools were the same in 2019/2020 and 2020/2021, and so may have gained experience and familiarity with some elements of the Focus4TAPS programme during 2019/2020. This experience may have increased the potential to achieve an impact on Year 5 pupils during 2020/2021. On the other hand, some continuing disruption following from the COVID-19 pandemic may have limited the potential to achieve an impact through this trial.

Cost

The average cost of the Focus4TAPS programme was calculated as £1,355.10 per school over 3 years, or £18.82 per pupil per year when averaged over 3 years. The cost follows from 15 hours of teaching cover for a Year 5 teacher, 15 hours of teaching cover for a science leader, and £150.00 for online training costs, all occurring within the first year as start-up costs. The cost calculations assume 24 pupils per class and assume costs of £40.17 per hour for teaching cover. Purchasing additional or specific equipment or resources is not formally required for delivering the programme.

Impact

Table 2: Summary of impact on primary outcome(s)

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No of pupils	P Value	EEF cost rating
Science	0.17 [0.03, 0.32]	2 months' progress	4 of 5 padlocks	2,513	.021	£ £ £ £ £ (1 of 5)
Science, FSM-eligible children	0.15 [-0.06, 0.36]	2 months' progress	N/A	582	.156	N/A

Introduction

Background

Focus for Teacher Assessment of Primary Science (Focus4TAPS)

The Focus4TAPS intervention has developed from the Teacher Assessment in Primary Science (TAPS) project, which emphasises the active participation of pupils and the responsiveness of teachers as they utilise assessment for learning (Davies, *et al.*, 2017; Earle, *et al.*, 2016). The intervention uses practical science activities and informal and frequent assessments and feedback practices to promote scientific learning. The TAPS assessment model broadly aims to help classroom practices and formative assessment to inform summative assessment and reporting across the context of an entire school (Earle, *et al.*, 2016). The model promotes teachers and pupils having a shared understanding of progression in science, including learning objectives and criteria for success, and active pupil involvement within teaching and learning; for example, the model involves pupils identifying their existing ideas and learning needs, engaging in self-assessment and peer assessment, and receiving and acting on feedback (Earle, *et al.*, 2016). More practically, the model is supported by resources that provide specific assessment lesson plans and activities for covering Working Scientifically at primary school, which describe learning objectives and criteria for success, together with guidance that broadly promotes a regulatory cycle of planning, applying, recording, and assessing (Primary Science Teaching Trust, 2016).

Assessment for learning

The TAPS assessment model aims to support teachers to use assessment for learning, focused on the teaching and assessment of Working Scientifically. Within education, 'assessment for learning' and 'formative assessment' are often used synonymously, although various different definitions exist (Wiliam, 2011b). Assessment for learning often refers to the use of information from assessment to adapt teaching to meet pupils' needs, and the use of information from assessment to help regulate and inform pupils' learning; this may inherently involve the integration of assessment and instruction, for example, and the use of feedback to pupils (Black & Wiliam, 1998; Black & Wiliam, 2009; Wiliam, 2011b).

Considered in review across various studies, learning outcomes have beneficially associated with aspects of formative assessment and feedback (Hattie & Timperley, 2007), and with metacognition and self-regulation in learning (Richardson, *et al.*, 2012). Recently, the 'Embedding Formative Assessment' intervention, which aimed to enhance the use of formative assessment for pupils in secondary school, had a positive effect (an overall effect of 0.10) on pupils' General Certificate of Secondary Education (GCSE) Attainment 8 scores considered as their grades across eight subjects including English, mathematics, and science subjects (Speckesser, *et al.*, 2018).

Working Scientifically

Within the National Curriculum at Year 5, Working Scientifically broadly encompasses planning and undertaking practical scientific enquiries, recording and interpreting results, and forming conclusions and explanations (Department for Education, 2014). More specifically, Working Scientifically at Year 5 encompasses:

> Planning different types of scientific enquiries to answer questions, including recognising and controlling variables where necessary; taking measurements, using a range of scientific equipment, with increasing accuracy and precision, and taking repeat readings when appropriate; recording data and results of increasing complexity using scientific diagrams and labels, classification keys, tables, scatter graphs, and bar and line graphs; using test results to make predictions to set up further comparative and fair tests; reporting and presenting findings from enquiries, including conclusions, causal relationships and explanations of and degree of trust in results, in oral and written forms such as displays and other presentations; and identifying scientific evidence that has been used to support or refute ideas or arguments. (Department for Education, 2014, p. 190)

Concurrently, the National Curriculum at Year 5 covers the following science topics and content: living things and their habitats; animals including humans; properties and changes of materials; Earth and space; and forces (Department for Education, 2014, pp. 192-195).

Recently, the 'Thinking, Doing, Talking Science' intervention aimed to enhance science teaching and learning for Year 5 pupils through supporting teachers to convey the principles of scientific enquiry (such as how to ask scientific questions and design experiments to find out the answers) and supporting pupils' scientific thinking around 'big questions', while also directly or indirectly fostering pupils' motivation and wider attitudes towards science (Hanley, *et al.*, 2015; Kitmitto, *et al.*, 2018). An initial efficacy evaluation found positive effects on science attainment test scores (an overall effect of 0.22) and appeared to associate with pupils' holding positive views towards science (Hanley, *et al.*, 2015); nevertheless, a subsequent effectiveness evaluation found no clear effects on science test scores, but small increases in pupils' confidence and interest in science (Kitmitto, *et al.*, 2018).

Science attainment, attitudes, and beliefs

Within science education, pupils' attainment and also their wider orientations towards science remain important, encompassing aspects such as thinking scientifically (reflected through Working Scientifically within the wider curriculum) together with attitudes and aspirations towards science (Royal Society, 2008, 2010, 2014). Science-related careers are often promoted as potential pathways to personal success in life, while science-related fields increase national prosperity through industry and innovation (EngineeringUK, 2017; Institute of Physics, 2012). Problematically, disadvantaged children often show lower attainment than other children, including in science subjects (Shaw, *et al.*, 2016; Social Mobility Commission, 2016). Additionally, and more generally, primary and secondary school pupils have often enjoyed science but have not necessarily exhibited strong identities and aspirations towards science (Archer, *et al.*, 2015; Archer, *et al.*, 2010). Essentially, science remains less accessible to many children, which may limit attempts to mitigate disadvantage and foster social mobility.

More generally within education and learning, motivational beliefs and attainment have positively associated (Freund & Kasten, 2012; Richardson, *et al.*, 2012; Valentine, *et al.*, 2004). Nevertheless, supporting pupils' motivational beliefs such as their confidence while concurrently supporting their attainment is more realistic than expecting that pupils' motivational beliefs (without any further support) can enhance their attainment (O'Mara, *et al.*, 2006). More specifically for science, pupils' attitudes such as their interest in science, and their motivational beliefs such as their confidence in their own abilities, together with their own attainment, have often associated with their aspirations towards science (Bøe & Henriksen, 2015; Regan & DeWitt, 2015). Interventions specifically aiming at enhancing pupils' attitudes and motivations towards science have entailed variable, but often positive results (Rosenzweig & Wigfield, 2016). Essentially, pupils' science attainment, attitudes, and beliefs often associate, although it remains possible that associations may vary in magnitudes, for different students, and/or at different circumstances or times; nevertheless, these remain important aspects within science education due to their wider implications on (and frequent associations with) pupils' aspirations and potential wider progression within science, especially within wider contexts of mitigating inaccessibility to science and disadvantage.

Summary

The Focus4TAPS intervention includes resources covering lesson plans and activities for scientific enquiry (Working Scientifically), each with an identified focus for assessment, learning objectives and criteria for success, and guidance on how to interpret children's responses; wider guidance aims to foster and support assessment for learning within science and across the wider school context (Davies, *et al.*, 2014; Earle, *et al.*, 2016; Primary Science Teaching Trust, 2016).

Concurrently, the focus on broadly enhancing the teaching of Working Scientifically (inherently involving practical work) may help address the need for support within this area, and this may be (more generally) also enjoyed and appreciated by pupils (Gatsby Charitable Foundation, 2017; Hanley, *et al.*, 2015; Kitmitto, *et al.*, 2018; National Foundation for Educational Research, 2011). Nevertheless, it remains possible that attainment gains may be unclear or vary.

TAPS places a particular emphasis on the focused teaching and assessment of Working Scientifically (where teachers select a focus for teaching, learning, and assessment for each of their practical science lessons, supported by resources such as example activities). Concurrently, the Education Endowment Foundation (EEF) aims to support pupils' educational attainment, and particularly for those pupils facing disadvantage. Accordingly, the trial focused on (as a primary outcome) pupils' science attainment while also considering (as secondary outcomes) pupils' attitudes, beliefs, and wider views concerning science and Working Scientifically at school. The trial also applied an implementation and process evaluation (IPE) in order to gain greater understanding and to help inform the generalised theory of change/logic model.

The Focus4TAPS trial design was a two-arm (intervention and control), cluster randomised efficacy trial (with schools as the units of clustering/randomisation). Intervention schools received the Focus4TAPS programme; control schools undertook their usual teaching/learning ('business as usual'). The trial recruited schools during the 2018/2019 academic year, and applied the Focus4TAPS programme with Year 5 pupils across some (but not all) of the 2019/2020 academic year before pausing because of the COVID-19 pandemic. The trial then recommenced in order to apply the Focus4TAPS programme with a new cohort of Year 5 pupils across the 2020/2021 academic year.

The various implications or changes that followed from the pandemic are highlighted as and when relevant across this report.

Intervention

The Focus4TAPS intervention operationalises assessment for learning applied to Working Scientifically at primary school, including teachers (and pupils) gaining enhanced understanding of progression in Working Scientifically, applying formative assessment strategies, and adapting levels of support and challenge during lessons across an academic year.

The Focus4TAPS trial was planned and prepared from August 2018. The intervention commenced with Year 5 pupils (aged around 9–10 years old) during the 2019/2020 academic year. From March 2020, the trial was unable to continue given disruption following from the COVID-19 pandemic. The Focus4TAPS trial recommenced with Year 5 pupils during the academic year of 2020/2021. During the 2020/2021 academic year, there was a national lockdown between January and March 2021 during which time schools had to manage teaching most of their pupils using online methods. As the report will outline there were disruptions to the delivery of the programme. The programme was originally funded to be delivered face to face; the original face to face trial was paused in March 2020 and restarted in September 2020 where every aspect of the programme and data collection took place online.

The intervention as originally conceptualised for delivery in 2019/2020 prior to the COVID-19 pandemic involved the following aspects.

1. Brief name: Focus for Teacher Assessment of Primary Science (Focus4TAPS).

2. Why (rationale/theory): Science assessment needs to support learning, by, for example, identifying starting points, utilising focused practical teaching activities and a choice of multimodal recording. Schools need support to develop such assessment practices because of: the lack of centralised guidance after the removal of levels; the inherent difficulty of assessing practical work; and the emphasis in primary schools on English and mathematics, leaving little time for science teaching and Continuing Professional Development (CPD).

Since 2013, the TAPS project has been developing support for teachers to use assessment for learning, with particular emphasis on the focused teaching and assessment of Working Scientifically. TAPS operationalised a model of assessment put forward by an expert group, led by Professor Wynne Harlen (Nuffield Foundation, 2012), whereby the rich information gathered during classroom formative assessment was utilised for summative purposes, in an attempt to stop 'teaching to the test' and a narrowing of the curriculum. TAPS worked in collaboration with local primary schools and Primary Science Teaching Trust (PSTT) Teacher Fellows to develop and exemplify a pyramid-shaped model that emphasised the active participation of pupils and the responsiveness of teachers as they utilise assessment for learning (Davies, *et al.*, 2017; Earle, *et al.*, 2016).

Focus4TAPS draws together the TAPS research on formative assessment into a CPD package, with teachers attending three training days and completing activities in the gaps between the training days ('gap tasks'). The training aims to develop teacher understanding of the Focused Assessment approach, in which teacher attention to children's action, talk, and recording within a 'normal' class science lesson is directed to focus on a particular aspect of scientific learning. The TAPS Focused Assessment lesson plans contain guidance for assessing a Working Scientifically objective (e.g. recording results or drawing conclusions, as outlined by the National Curriculum; Department for Education, 2014) within a meaningful, and often practical, conceptual context. By exploring the Focused Assessment activities on the training, selecting some to try out as gap tasks, and linking these to the broader principles in the TAPS pyramid, the teachers will be enabled to develop their teaching of Working Scientifically and their use of formative assessment.

The impact of assessment for learning approaches is well supported by research evidence (Black & Wiliam, 1998; Mansell & James, 2009; Wiliam, 2011a) and TAPS has developed particular examples of how to implement such approaches in Primary Science, which emphasise the elicitation of pupil ideas (Nuffield Foundation, 1998), together with the focused teaching and recording of Working Scientifically or scientific reasoning (Goldsworthy, *et al.*, 2000; McMahon & Davies, 2003; Nunes, *et al.*, 2017). Such focused recording was an element of the 'Thinking, Doing, Talking Science' project, where an initial evaluation found a positive impact on pupil science test scores (Hanley, *et al.*, 2015) while a subsequent effectiveness evaluation found no impact on pupil science test scores (Kitmitto, *et al.*, 2018).

3. Who (recipients): Year 5 pupils in the classes of teachers who have had the training (one class per school).

4. What (materials): Teachers attend CPD days and will be provided with example plans and activities for use in their classrooms. Teachers select five TAPS plans from the bank of 40+, to carry out as gap tasks during the year, using largely everyday materials, which will be found in school.

5. What (procedures): Teachers will be supported through the TAPS training and resources to select a focus for teaching, learning, and assessment for each of their practical science lessons. They will use their new understanding of progression in Working Scientifically and new formative assessment strategies to adapt the level of support and challenge during subsequent lessons. By focusing on a different element of Working Scientifically each lesson, they consider the full breadth of enquiry skills across the year. Additional sources of support for implementation of the approach will be: discussions at the CPD days with the trainers and other teachers; email contact with a trainer between CPD days; and in-school support via discussions with colleagues.

6. Who (implementers): Focus4TAPS is delivered by class teachers who attend CPD days. There will be two trainers for each region: Dr Sarah Earle and one other experienced TAPS practitioner. A senior school leader will be invited to attend the introductory session, so that they understand the trial and can enable the attending teachers to carry out the intervention.

7. How (mode of delivery): The TAPS approach should be implemented within the normal class science lessons. Since the approach is designed to modify teaching practice within lessons, there would not be large-scale changes to the amount of science taught, unless the school is currently teaching less than the recommended 2 hours per week (Leonardi, *et al.*, 2017).

8. Where (setting): Regular classrooms in participating schools.

9. When and how much (dosage): A minimum of five Focused Assessment class lessons will be carried out. Ideally, the strategies should also be integrated into the class science lessons, with a clear focus for teaching, learning, and assessment developed in all science lessons.

10. Tailoring: A wide range of Focused Assessment lesson plans and exemplars are provided (169 Word and PDF files are currently on the TAPS website), with the expectation that teachers will begin by following some of these closely, then begin to adapt and broaden their teaching and assessment of science as they become confident with the approach.

The Focus4TAPS trial recommenced with Year 5 pupils during the academic year of 2020/2021. The following aspects of the intervention were subsequently refined for delivery in 2020/2021 given the COVID-19 pandemic.

11. Mode of delivery: During 2019/2020 the programme was delivered through three training sessions, each delivered in-person to regional groups of teachers. During 2020/2021, the programme was delivered through six training sessions, each delivered remotely (online) and with a follow-up video, delivered to groups of teachers. Teachers were also able to view a summary video for each session, so that they could 'catch up' if they were unable to attend a live session.

12. Timing, duration, and frequency of delivery: During 2019/2020 the programme was delivered through three training sessions, each delivered in-person to regional groups of teachers. During 2020/2021, the programme was delivered through six training sessions.

13. Content: The same content was delivered in 2020/2021, but was less 'hands on' so teachers did not directly experience the activities within the sessions. Additional content discussed how to carry out practical work given COVID-19 arrangements/restrictions in schools.

14. Recipients and deliverers: There were some changes of teachers within schools from 2019/2020 to 2020/2021. A new cohort of Year 5 pupils was the recipient of the intervention: Year 5 pupils as of 2020/2021.

15. Aim and intended outcomes of the programme: The aims and outcomes remained the same.

Figure 1 presents the logic model of the trial.

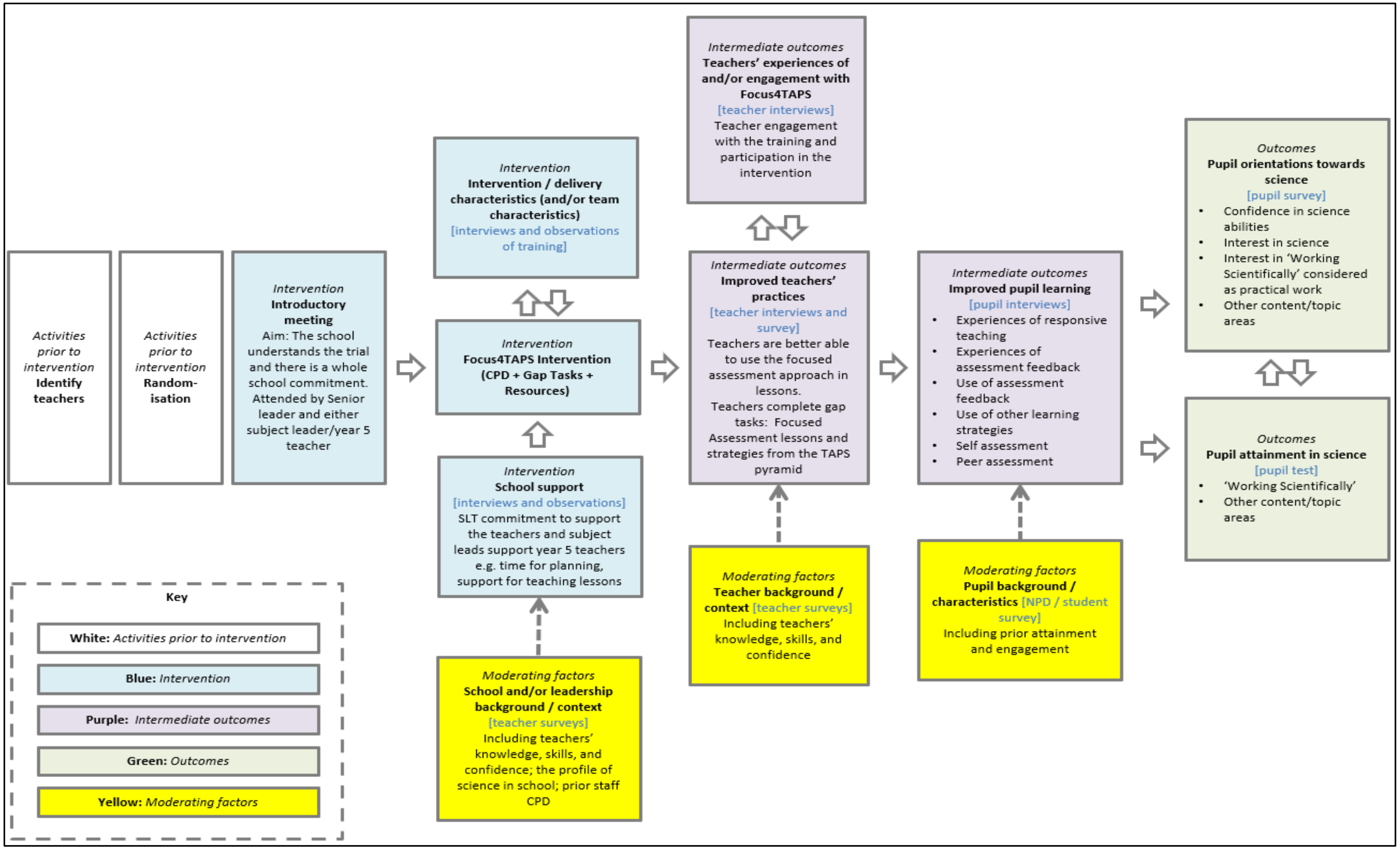

Figure 1: Logic model. CPD, Continuing Professional Development; Focus4TAPS, Focus for Teacher Assessment of Primary Science; NPD, National Pupil Database; SLT, Senior Leadership Team; TAPS, Teacher Assessment in Primary Science.

Evaluation objectives

The evaluation aimed to determine the impact of Focus4TAPS on pupils' science attainment and attitudes towards science. The overall evaluation research questions were as follows:

1. What impact does Focus4TAPS have on pupils' science attainment?
2. What impact does Focus4TAPS have on pupils' attitudes towards science?

The impact evaluation also considered the following research questions:

1. What impact does Focus4TAPS have in the presence of compliance/non-compliance with the intervention programme?
2. What impact does Focus4TAPS have when missing data has been accommodated?
3. Does any impact vary across subgroups of pupils (including across ever-eligible/not ever-eligible for Free School Meals [FSM] status)?
4. Does any impact vary when checking for robustness (including when accounting for further information and across different methods)?

The IPE also considered the following research questions:

1. To what extent do teachers, schools (and trainers) implement the intervention with fidelity and what factors affect fidelity?
2. To what extent do teachers, schools (and trainers) adhere to the intervention tasks and requirements and what factors affect compliance?
3. How does dosage vary across classrooms, what factors affect dosage, and to what extent does dosage impact on outcomes?
4. How well are different elements of the intervention delivered and what factors affect this quality of implementation?
5. To what extent does the intervention reach all (intended) pupils and what factors affect this?
6. How does the degree to which teachers and schools respond to the intervention (and CPD) vary and what factors affect this?
7. To what extent can the intervention (and its different elements) be distinguished from existing practice and what factors affect the degree of differentiation from normal practice?
8. To what extent does practice in control schools impact pupil outcomes? Is there a relationship between practice and how pupils can work scientifically?
9. (a) In what ways do teachers and schools adapt the intervention and how do these adaptations affect implementation? (b) What adaptations have teachers had to make to implement Focus4TAPS during school or class closures during the pandemic? Have the pandemic-related adaptations impacted implementation?
10. How practical is it for schools to implement the Focus4TAPS programme designed by Bath Spa?
11. Could teachers who attended training and delivered the Focus4TAPS programme continue to implement it without any further support or training?
12. How acceptable do teachers and school leaders feel it is for pupils to engage in the programme for the purpose of boosting science attainment?
13. Do teachers think there will be a sustained impact of the intervention on pupils' outcomes (in particular, attainment)? To what extent does this vary by pupil characteristic (e.g. FSM status, gender)?
14. How (if at all) do teachers think the intervention will impact on pupils in other ways (e.g. interest in other subjects)?
15. What do teachers think will be enablers and barriers to impact?
16. How scalable do teachers think the Focus4TAPS programme is and what are their suggestions for change if the programme is to be implemented more widely?
17. What impact does Focus4TAPS have on pupils' attitudes towards science?
18. What are the views of the intervention pupils about the teaching and learning methods applied by Focus4TAPS, and do such views vary by pupil characteristics (e.g. FSM status, gender)?
19. To what extent is Focus4TAPS distinguishable from control schools' model of learning?
20. Does the logic model suitably address the factors that are related to both the primary outcome and secondary outcomes?

21. What impact does Focus4TAPS have on teachers' attitudes towards science and confidence to teach/assess Working Scientifically in science?
22. What impact does Focus4TAPS have on teachers' assessment (and other) practices?
23. Does Focus4TAPS have an impact on teachers' perceptions of their job satisfaction and workload?

The trial Protocol and Statistical Analysis Plan can be found on the EEF website.[1]

Ethics and trial registration

The evaluation has been reviewed and approved by the University College London (UCL) Institute of Education Research Ethics Committee (with reference REC1146; see Appendix C). Changes to the evaluation processes following from the COVID-19 pandemic, such as the need to undertake remote/online interviews rather than in-person interviews, were also approved by the UCL Institute of Education Research Ethics Committee (as an amendment under the same reference). The British Educational Research Association (BERA) professional code of ethics was followed throughout the evaluation.

Invitation information sheets for schools (reproduced in Appendix D) explained the evaluation and allowed schools to make an informed decision about whether to participate. Schools were able to ask for additional information prior to making a decision. Schools then completed a Memorandum of Understanding (MoU) (reproduced in Appendix E), which explained the roles and responsibilities of the school, the developer team, and the evaluation team. After disruption following from the COVID-19 pandemic, schools affirmed their engagement in the recommenced trial through completing a second MoU (also reproduced in Appendix E).

After completing the trial, schools were provided with financial recompense as a 'thank you' (which is standard for the EEF evaluations), as explained on the school invitation materials and MoUs.

Information sheets for parents/guardians to discuss with their children were provided as part of the school recruitment process (reproduced in Appendix F), where schools decided to participate and then disseminated materials to parents and gathered any cases of withdrawal. To help ensure that children and parents/guardians could make informed and voluntary choices, the information sheets for parents/guardians explained the project in simple terms, provided an opportunity (with contact details) for parents/guardians to ask additional questions, and clearly allowed their child/children to withdraw from the project. The sheet and form also clarified that children and parents/guardians could withdraw at that point or at any point during the research by communicating these requirements directly. For the pupil interviews, specific opt-in consent was sought via information sheets and consent forms.

The evaluation is registered in the International Standard Randomised Control Trial Number (ISRCTN) Registry (with reference number ISRCTN17920547).

Data protection

Personal data were collected and processed in accordance with General Data Protection Regulation (GDPR) principles. Prior to the commencement of recruitment of schools, the trial was registered with the UCL Data Protection Officer (with registration number Z6364106/2018/11/83). Both UCL and Bath Spa University have conducted an assessment of the legal basis for processing data for their own organisations.

Personal data were collected and processed in accordance with GDPR principles under the 'public task' basis (i.e. 'processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller'). Special category personal data were collected and processed in accordance with GDPR principles under the scientific and historical research or statistical purposes.

UCL Institute of Education was a data controller for the purposes of the evaluation and research; Bath Spa was a data controller for the purposes of recruitment, administering the trial, and for research.

---

[1] Education Endowment Foundation. *Focus4TAPS* (accessed 20 December 2021):
https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/focus4taps

Information sheets for parents/guardians, pupils, and teachers contained clear and accessible information regarding the purpose of the study and data processing (reproduced in Appendix F). This information also assured all concerned that their data can be withdrawn up until the point of publication; contact details were provided so that any queries could be raised and clarified. The information sheets also contained a link to further information on how UCL uses participant information.[2]

Project team

Bath Spa University developed and delivered the Focus4TAPS programme, which draws together research from the TAPS project, which was funded by the PSTT and, which emphasises the active participation of pupils and the responsiveness of teachers as they utilise assessment for learning (Davies, *et al.*, 2017; Earle, *et al.*, 2016).

The Bath Spa University delivery team comprised of the following people:

- Dr Sarah Earle, Focus4TAPS Project Lead, responsible for recruitment and delivery;
- Dr Kendra McMahon, Focus4TAPS Adviser, initially supporting programme and delivery; and
- experienced TAPS practitioners (six during 2019/2020 and three during 2020/2021), supporting delivery of training.

The UCL Institute of Education evaluation team comprised of the following people:

- Dr Tamjid Mujtaba, Principal Investigator, responsible for the impact evaluation and statistical analysis alongside providing direction and undertaking the IPE;
- Dr Richard Sheldrake undertook the statistical analysis alongside quantitative data collection and contributed to other aspects of the evaluation;
- Professor Jeremy Hodgen provided management direction and advice;
- Professor Michael Reiss responsible for general advice; and
- research associates and research assistants, contributed to qualitative and quantitative data collection.

[2] UCL. *UCL General Research Participant Privacy Notice* (accessed 20 December 2021):
https://www.ucl.ac.uk/legal-services/privacy/ucl-general-research-participant-privacy-notice

Methods

Trial design

Table 3: Trial design

Trial design, including number of arms		Two-arm, cluster randomised efficacy trial
Unit of randomisation		Schools
Stratification variable (s) (if applicable)		Geographic region and school-level historical Key Stage 1 attainment
Primary outcome	Variable	Science attainment (test scores)
	Measure (instrument, scale, source)	A science attainment test developed by the University of York for the Education Endowment Foundation (EEF) (Joshi, *et al.*, 2021), providing a continuous single scale (0–45 marks)
Secondary outcome(s)	Variable(s)	Science attitudes (questionnaire responses)
	Measure(s) (instrument, scale, source)	A pupil questionnaire developed by the Focus for Teacher Assessment of Primary Science (Focus4TAPS) evaluation team, informed by existing national and international instruments, providing separate factors measuring: • interest/enjoyment in science • confidence in science • perceptions of science teachers and practices • self-efficacy within Working Scientifically • wider views concerning Working Scientifically • wider views concerning science These factors have agreement/disagreement scales (1–4, reflecting 'Disagree a lot' to 'Agree a lot')
Baseline for primary outcome	Variable	Key Stage 1 reading and mathematics attainment
	Measure (instrument, scale, source)	Information from the National Pupil Database (KS1_READ_OUTCOME and KS1_MATH_OUTCOME), providing categories: BLW 'Below the standards of Pre-Key stage'; PKF 'Pre-Key stage'; WTS 'Working towards the expected standard'; EXS 'Working at the expected standard'; GDS 'Working at a greater depth within the expected standard'
Baseline for secondary outcome(s)	Variable	Key Stage 1 reading and mathematics attainment
	Measure (instrument, scale, source)	Information from the National Pupil Database (KS1_READ_OUTCOME and KS1_MATH_OUTCOME), providing categories: BLW 'Below the standards of Pre-Key stage'; PKF 'Pre-Key stage'; WTS 'Working towards the expected standard'; EXS 'Working at the expected standard'; GDS 'Working at a greater depth within the expected standard'

The Focus4TAPS intervention aimed to apply assessment for learning to Working Scientifically at primary school, including teachers (and pupils) gaining enhanced understanding of progression in Working Scientifically, applying formative assessment strategies, and adapting levels of support and challenge during lessons across an academic year.

The Focus4TAPS trial design was a two-arm (intervention and control), cluster randomised efficacy trial (with schools as the units of clustering/randomisation). Intervention schools received the Focus4TAPS programme; control schools undertook their usual teaching/learning ('business as usual').

The primary outcome was attainment on a science test, which included aspects of Working Scientifically, developed by the University of York for use in the EEF evaluations (Joshi, *et al.*, 2021). The secondary outcomes were measured through a questionnaire and covered a range of attitudes and beliefs contextualised to science, including interest/enjoyment in science, confidence in science, and views related to Working Scientifically.

The trial was planned and designed for the intervention to occur during the 2019/2020 academic year. The intervention delivery team recruited schools during the 2018/2019 academic year, from December 2018 to April 2019; the evaluation team then randomised schools to the intervention group or to the control group in June 2019. The Focus4TAPS trial then commenced with Year 5 pupils during the 2019/2020 academic year, from September 2019 onwards. However, from March 2020, the trial was unable to continue given teaching/learning disruption following from the COVID-19 pandemic. The trial then recommenced with a new cohort of Year 5 pupils during the academic year of 2020/2021, from September 2020 onwards. The schools retained their original assignment to the intervention group or to the control group. Essentially, the trial recommenced with the same schools within the same design but with a different cohort of pupils. During the 2020/2021 academic year schools had to manage teaching most of their pupils using online methods during a national lockdown between January 2021 and March 2021.

Some data collection methods were also adjusted given the wider implications of the COVID-19 pandemic, which are detailed when relevant in the following sections. For example, the impact evaluation originally intended to measure baseline attitudes and beliefs related to science through a questionnaire; however, this was considered to be unfeasible at the recommencement of the trial, given the potential burden to schools/pupils and the limited time frame for recommencement activities following disruption from the COVID-19 pandemic. Similarly, within the IPE, the trial was unable to survey or interview teachers towards the start of the academic year in order to clarify 'business as usual' or 'usual practices'; the trial was only able to survey and interview teachers towards the end of the academic year. The testing and surveying within the impact evaluation was administered by teachers, rather than by independent invigilators, given the pandemic circumstances. These various changes or implications are detailed as and when relevant in the various sections that follow.

Participant selection

During recruitment, schools were eligible for the intervention if:

- the school would have children in Year 5 during the 2019/2020 academic year, who were also available during the 2018/2019 academic year for pre-intervention/baseline surveying; and
- the school had not taken part in the 'Thinking, Doing Talking Science' trial, the Primary Science Quality Mark (from 2017 onwards), or the TAPS project.

The intervention delivery team recruited schools during the 2018/2019 academic year, from December 2018 to April 2019. Over 1,800 schools were initially approached by the intervention delivery team, and 324 schools expressed interest. MoUs were sent to 228 schools and received from 147 schools; schools also provided details of the relevant teachers and pupils to the evaluation team. Six schools then withdrew before randomisation, leaving 141 schools.

Randomisation was then undertaken by the evaluation team in June 2019 for 141 schools: 71 were assigned to the intervention group; and 70 were assigned to the control group. The randomisation process is described in more detail within a later section of the report and also within Appendix G. Two intervention group schools then withdrew before commencement (before receiving any training) due to staff changes.

The trial then commenced with Year 5 pupils during the 2019/2020 academic year. However, from March 2020, the trial was unable to continue given learning disruption following from the COVID-19 pandemic. During the summer of 2020, the intervention delivery team engaged with the 139 schools to reaffirm their involvement through updated MoUs; 121 schools remained, with 61 schools remaining in the intervention group and 60 schools remaining in the control group. The schools retained their original assignment to the intervention group or to the control group.

The trial then recommenced with Year 5 pupils during the academic year of 2020/2021. Essentially, the trial recommenced with some (but not all) of the same schools and with a different cohort of Year 5 pupils (Year 5 as of the 2020/2021 academic year). Schools provided updated details of the relevant teachers and pupils to the evaluation team.

From the 121 schools that recommenced, 112 schools had the same specified Year 5 teachers across 2019/2020 and 2020/2021 (58 of 61 intervention schools and 54 of 60 control schools), while 9 schools had different Year 5 teachers (3 of 61 intervention schools and 6 of 60 control schools). Similarly, 112 schools had the same science leaders across 2019/2020 and 2020/2021 (56 of 61 intervention schools and 56 of 60 control schools), while 9 schools had different science leaders (5 of 61 intervention schools and 4 of 60 control schools). From the nine schools with changes in Year 5 teachers, seven schools (three intervention schools and four control schools) had continuity in science leaders; only two schools (both control schools) had changes in specified Year 5 teachers and also changes in science leaders. From the 121 schools that recommenced, as of 2020/2021, 16 intervention schools and 15 control schools indicated that the Year 5 teacher was also the science leader.

The trial originally offered ‘thank you’ payments following completion of all of the various requirements such as surveying and testing pupils (£500 for intervention schools and £1,000 for control schools); control schools were also offered the opportunity to subsequently access the Focus4TAPS training/development after the trial for a reduced fee. Following disruption due to the COVID-19 pandemic, the recommenced arrangements offered a series of ‘thank you’ payments for completing various requirements of the trial (totalling £400 for intervention schools and £400 for control schools) and then offered the final ‘thank you’ payments as originally planned (£500 for intervention schools and £1,000 for control schools). Control schools were again also offered the opportunity to subsequently access the Focus4TAPS training/development after the trial for a reduced fee.

Outcome measures

Baseline measures

The pupils’ Key Stage 1 classifications for mathematics and reading were sourced from the National Pupil Database (NPD), in order to provide baseline measures that reflected the pupils’ prior attainment. At the end of Key Stage 1, pupils in Year 2 (aged around 6–7 years old) undertake national tests encompassing mathematics and English reading, with optional tests in English grammar/punctuation/spelling, which inform teacher assessments conveyed as classifications (Department for Education, 2014). The most common classifications are pupils working towards, working at, or working at a greater depth than the expected standard.

Considering these indicators as recorded within the NPD from the Department for Education helped ensure consistency across different trials undertaken by the EEF and reduced the burden on pupils and schools. Evaluators are encouraged to source measures of baseline attainment from the NPD (Education Endowment Foundation, 2018). Additionally, the science test that formed the primary outcome measure was developed during the course of the trial and was not available to be also applied as a baseline measure. Prior attainment in mathematics and reading may not necessarily reflect prior attainment in science (and attainment may not necessarily reflect pupils’ views such as their interest in science).

The pupils within the recommenced trial were in Year 5 as of the 2020/2021 academic year and were previously in Year 2 as of the 2017/2018 academic year. For the participating pupils within the trial, schools securely transferred information including names and Unique Pupil Numbers to the evaluation team; this information was then securely transferred to the Department for Education, who provided the evaluation team with the necessary information from the NPD via the Secure Research Service (SRS) from the Office for National Statistics (ONS).

Key Stage 1 information was not available for all pupils. Some information was missing/unavailable (i.e. blank). The provided categories of 'A' (Absent for long periods, recently arrived, and insufficient information for teachers to make judgements) and 'D' (Disapplied from the National Curriculum) were also re-classified as missing/unavailable data, given that these categories essentially reflect prior attainment information being unavailable. The remaining categories were: BLW 'Below the standards of Pre-Key stage'; PKF 'Pre-Key stage'; WTS 'Working towards the expected standard'; EXS 'Working at the expected standard'; and GDS 'Working at a greater depth within the expected standard'.

Overall, Key Stage 1 information were available for 2,781 of 2,882 pupils for reading (96.5% coverage; 101 missing), and for 2,782 of 2,882 pupils for mathematics (96.5% coverage; 100 missing).

The trial originally intended to also measure baseline attitudes and beliefs related to science through a questionnaire. However, this was considered to be unfeasible at the recommencement of the trial, given the potential burden to schools/pupils and the limited time frame for recommencement activities following disruption from the COVID-19 pandemic.

Primary outcome

The Focus4TAPS intervention includes resources covering lesson plans and activities for scientific enquiry (Working Scientifically), each with an identified focus for assessment, learning objectives and criteria for success, and guidance on how to interpret children's responses; wider guidance aims to foster and support assessment for learning within science and across the wider school context (Davies, *et al.*, 2014; Earle, *et al.*, 2016; Primary Science Teaching Trust, 2016). Refining the teaching and learning of Working Scientifically, and practical work in general, was theorised to foster and otherwise support pupils' attainment in science. For example, ensuring that teachers and pupils have a shared understanding of learning aims and criteria for success, providing feedback to pupils, and more generally fostering assessment for learning may possibly enhance attainment (Hattie & Timperley, 2007).

Accordingly, the primary outcome was pupils' science attainment, measured through a test developed by the University of York for the EEF (Joshi, *et al.*, 2021). The test was comprised of 15 questions (with 38 individual items or 'question parts' such as 1A, 1B, and 1C) with a total of 45 marks. The test was designed with balanced coverage across subject areas, topic areas, levels of challenge, and included coverage of Working Scientifically (Joshi, *et al.*, 2021). The test questions involved a range of question formats, including multiple-choice, free-text, and drawing graphical responses.

The science test is reproduced in Appendix H.

The Focus4TAPS trial originally intended for tests (and questionnaires) to be undertaken under external invigilation, although the COVID-19 pandemic meant that this was unfeasible. The tests (and questionnaires) were administered by teachers following instructions/guidance. Pupils completed tests during June 2021 and July 2021.

Tests were provided as paper booklets for pupils to complete. Schools were provided with instructions for test administration, as written by the test developers, and pupils had 45 minutes in which to complete the test. The instructions conveyed the importance of pupils working alone, without copying from each other or discussing their answers.

The completed tests were processed through data entry by an external company, which recorded the pupils' responses (such as particular options or boxes being ticked) and transcribed their free-text responses. The data entry company was unaware of the control/intervention status of any material.

Responses to the multiple-choice and other similarly structured questions within the data were scored through software (IBM® SPSS®) by the evaluation team, which classified the relevant response options as correct and other responses as incorrect following the mark scheme from the test developers, and where the team was unaware of the control/intervention status of any material. Other responses were scored by postgraduate teaching students from the UCL Institute of Education, following the mark scheme from the test developers, either through considering the transcribed data in electronic format or through considering the paper tests to score the graphical responses to questions 9A and 11C. The markers were unaware of the control/intervention status of any material.

The mark scheme accounted for the pupils potentially selecting multiple options. For example, if a question instructed pupils to tick one of four responses, anything other than the single correct response (such as ticking two, three, or all four responses) was scored as incorrect. The mark scheme also involved blank responses (where questions were not answered/attempted) being scored as incorrect (i.e. a blank response was something other than the correct response specified within the scheme and hence scored as incorrect).

The overall marking process involved initial double-marking by two independent markers for all of the manually marked test responses, then further review and marking of any discrepancies (i.e. triple-marking). This involved quality assurance by the evaluation team reviewing all responses/marks and identifying and resolving discrepancies across markers to ensure that the final scores consistently followed the marking scheme, while being unaware of the control/intervention status of any material.

Encompassing intervention and control pupils/schools, 2,600 out of 2,882 pupils completed tests (90.2% coverage; 282 missing). Some pupils may have been absent at the time of testing, may have been disinclined to complete the test, or may not have provided a name on the test.

Considering the data across all 2,600 pupils, the test showed acceptable reliability/internal consistency (38 items, Cronbach's Alpha = .874). Across the 38 test items, factor analysis revealed eight underlying factors with Eigenvalues greater than one (specifically, with Eigenvalues of 7.349, 1.391, 1.324, 1.156, 1.103, 1.060, 1.033, and 1.027). Almost all items had greatest associations with the first factor (excepting question 5D with a factor loading/association of 0.230 with the first factor and 0.238 with the eighth factor). Given the information available, the other potential factors did not clearly show interpretable distinctions that might reflect different objects of measurement; in this context, it is possible that multiple factors may be artefacts following from the diversity of question topics, subjects, and modes (which may intersect in complex ways). Overall, this provided sufficient evidence to affirm that the test items could be aggregated into an overall measure of science attainment with acceptable reliability/internal consistency.

The test developers intended for all of the test items to form an overall measure of science attainment and did not recommend the separation of test items into subscales (Joshi, *et al.*, 2021). Nevertheless, in order to potentially gain greater insight, the trial Statistical Analysis Plan (which was devised at the start of the trial before the test was finalised) intended to consider any test items as an overall measure of science attainment and also through separate 'Working Scientifically' and 'content/topic knowledge' subscales, if those subscales were supported by features of the test (the conceptual/theorised focus per test item) and/or empirical approaches such as factor analysis. The test developers assigned each test item to a subject/domain (biology, chemistry, or physics), and some items were also assigned as covering Working Scientifically (Joshi, *et al.*, 2021), which clarified the different conceptual/theorised foci. Accordingly, separate subscales were formed that focused on the science subjects/domains of biology (10 items; Cronbach's Alpha = .654), chemistry (14 items; Cronbach's Alpha = .733), and physics (14 items; Cronbach's Alpha = .717). Additionally, a Working Scientifically subscale was formed (16 items; Cronbach's Alpha = .774). The subject/domain subscales of biology, chemistry, and physics did not overlap, while the Working Scientifically subscale overlapped with the subject/domain subscales. Factor analysis revealed two to three underlying factors with Eigenvalues greater than one per subscale (varying per subscale), where the items generally (but not always) had greatest associations with the first factor. Essentially, there was some empirical support for considering subscales. Subscales within a test may not necessarily be balanced across question format or mode, level of difficulty/challenge, and other aspects of measurement (which may intersect in various ways); for example, it is possible that some topics within Working Scientifically might be unavoidably more (or less) challenging than other concepts within science, potentially due to the topics themselves and potentially due to how they might be measured. Nevertheless, considering Working Scientifically was meaningful and relevant in the context of Focus4TAPS given potential foci towards inquiry and practical areas of science, although there was no theoretical basis for Focus4TAPS to have different impacts across biology, chemistry, and physics. Overall, it was reasonable to infer that the subscales could be formed, and these were used as supplementary indicators in order to potentially provide further insight.

Secondary outcomes

Refining the teaching and learning of Working Scientifically, and practical work in general, may foster and otherwise support pupils' motivations and orientations towards science, including their interest and enjoyment of science at school, their confidence in doing science at school, and numerous other attitudes and beliefs. Pupils' attitudes and beliefs, such as their interest and confidence in doing science, are important influences on their overall educational progressions and trajectories towards or away from science (Archer, *et al.*, 2017; Mujtaba & Reiss, 2014; Mujtaba, *et al.*, 2018).

Accordingly, the secondary outcomes covered a range of pupils' attitudes and beliefs related to science, measured through a questionnaire. The questionnaire design was informed by existing national and international instruments to help ensure efficient and reliable measurement, which also facilitated comparability against national/international research for potential contextualisation.

The questionnaire is reproduced in Appendix I.

The Focus4TAPS trial originally intended for questionnaires (and tests) to be undertaken under external invigilation, although the COVID-19 pandemic entailed that this was unfeasible. The questionnaires (and tests) were administered by teachers following instructions/guidance. Pupils completed questionnaires during June 2021 and July 2021.

Questionnaires were provided as paper booklets for pupils to complete. Schools were provided with guidance explaining that, in order to reduce pupil fatigue, the test and survey could be administered a few hours apart, such as one first thing in the morning and one later in the morning or afternoon. If pupils were missing on the day, schools were encouraged to arrange for them to complete the test and survey as soon as possible when they returned to school.

The completed questionnaires were processed through data entry by an external company, which recorded the pupils' responses such as particular options or boxes being ticked. The data entry company was unaware of the control/intervention status of any material.

The questionnaire items covering interest and enjoyment in science, confidence in science, and perceptions of science teachers were sourced with permission from the Trends in International Mathematics and Science Study (TIMSS) 2015 for Grade 4 (Year 5) pupils (Martin, *et al.*, 2016b). Items covering pupils' self-regulation within their science learning followed from the reflection scale of the 'Self-Regulation of Learning Self-Report Scale' (Toering, *et al.*, 2012). Items considering pupils' self-efficacy to undertake aspects of Working Scientifically were devised by the evaluation team in order to cover relevant aspects of the Year 5 National Curriculum (Department for Education, 2014). Items covering pupils' wider views concerning Working Scientifically followed from the measurement of epistemic beliefs about science from the Programme for International Student Assessment 2015 (OECD, 2017). Items covering the wider benefits of science followed from the State of the Nation survey for Primary Science from the Wellcome Trust (Leonardi, *et al.*, 2017).

Further items within the questionnaire were devised by the evaluation team in order to inform the IPE, rather than acting as secondary outcomes within the impact evaluation.

The phrasing of some questionnaire items covering perceptions of science teachers were slightly adapted from TIMSS in order to have a clearer focus on science and hence to avoid ambiguity (e.g. 'I know what my teacher expects me to do' was refined into 'I know what my teacher expects me to do in science', 'My teacher is easy to understand' was refined into 'My teacher is easy to understand when teaching science', etc.).

Encompassing intervention and control pupils/schools, 2,501 out of 2,882 pupils completed questionnaires (86.8% coverage; 381 missing). Some pupils may have been absent, disinclined to complete the questionnaire, or may not have provided a name.

The questionnaire items were measured with response categories of 'Disagree a lot' (scored as 1), 'Disagree a little' (scored as 2), 'Agree a little' (scored as 3), and 'Agree a lot' (scored as 4). Factor analysis and reliability coefficients (Cronbach's Alpha values) affirmed that the relevant items could be aggregated together as theorised and documented in advance within the Statistical Analysis Plan (as summarised below). Each secondary outcome was then operationalised as the average of the relevant questionnaire items where pupils provided answers (the default handling of the COMPUTE MEANS functionality within IBM® SPSS® to aid potential comparability with wider research); for example, if a pupil answered eight of nine items, leaving one item blank, then the average would be calculated across the eight items with available information. The scoring of responses for any negatively orientated items (such as ‘Science is boring’) were reversed for consistency within this process. The outcomes could therefore be interpreted using the same agreement scale as the underlying items.

Table 4: Secondary outcomes and questionnaire items

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	.905
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)	.799
Perceptions of science teachers	• 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	.863
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	.771
Self-efficacy for Working Scientifically	• I can plan different science investigations • 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	.813
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	.765

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	.743

Sample size

The sample size was determined at the protocol stage in order to be able to detect an effect size of .200 in the primary outcome of science test scores. The calculations assumed: 25 pupils per class with 5 pupils per class ever-eligible for FSM; moderate correlation of .500 between earlier Key Stage 1 attainment (in mathematics and English) and contemporary science test scores, due to the potential for changes over time and differences across academic subjects; and an intracluster correlation of .150, which has been observed within other trials with Year 5 pupils (Hanley, *et al.*, 2015; Kitmitto, *et al.*, 2018).

The minimum detectable effect size (MDES) calculations at the protocol, randomisation, and recommencement stages considered FSM status as reported by schools (i.e. an initial indication provided as part of the submission of core pupil information). The updated calculations within this report considered the ever-eligible for FSM status from the NPD, which reflects a pupil having been eligible for FSM at any time in the last 6 years up to the pupil's current year (not including nursery) as of the spring 2021 school census (via the EVERFSM_6_P_SPR21 indicator from the NPD). The pupil numbers in the MDES table update and correct those reported within the published Statistical Analysis Plan (which contained a minor error around numbers).

The calculations at the analysis stage followed from pupil numbers as analysed within the statistical modelling of the primary outcome (science test scores), from observed intracluster correlation coefficients (ICCs), and from observed correlations between Key Stage 1 classifications and science test scores. Nevertheless, any correlation between Key Stage 1 classifications (measured as categories) and science test scores (measured as a scale) can only be an indication. The process used here involved forming the Key Stage 1 categories into 'working-scales' (1–5 scales; 1 = BLW; 2 = PKF; 3 = WTS; 4 = EXS; 5 = GDS). Across all Year 5 pupils, the observed correlations between the science test scores and the Key Stage 1 working-scales were moderate ($R = .523$, $p < .001$, for reading and $R = .535$, $p < .001$, for mathematics). The average of the two observed correlation coefficients (.523 + .535 = 1.058 / 2 = .529) was then used within the MDES calculations. The same approach was applied with the subsample of pupils ever-eligible for FSM (KS1 reading: $R = .498$, $p < .001$; KS1 mathematics: $R = .504$, $p < .001$; average for MDES calculation = .501).

In summary, considering all Year 5 pupils, MDES were:

- .200 at the protocol stage (assuming 3,500 pupils across 140 schools);
- .199 at the randomisation stage (3,537 pupils across 141 schools);
- .216 at the recommencement stage (2,882 pupils across 121 schools); and
- .248 at the analysis stage (2,513 pupils across 120 schools).

Considering the subsample of Year 5 pupils ever-eligible for FSM, MDES were:

- .287 at the protocol stage (assuming 420 pupils across 140 schools);
- .241 at the randomisation stage (895 pupils across 141 schools);
- .271 at the recommencement stage (618 pupils across 121 schools); and
- .310 at the analysis stage (582 pupils across 120 schools).

The MDES at analysis was higher than at earlier stages, broadly following from higher observed ICCs and lower numbers of involved pupils/schools than earlier assumptions. At the analysis stage, keeping all other values constant, achieving an MDES lower than .200 would have required an ICC below .120 or would have required 180 or more schools.

Nevertheless, the sample sizes were potentially influenced by unknown factors during recommencement (where some schools did not continue with the trial), rather than only following from random allocation to the intervention group or to the control group, so MDES estimates should be interpreted with caution.

The trial recruitment aimed to achieve an MDES lower than .200 through covering a sufficiently large overall sample, encompassing intervention and control schools. The overall sample also encompassed subgroups of pupils with different characteristics such as gender and FSM status. The trial recruitment was not designed to achieve minimum sample sizes for any subgroups of pupils, including pupils who were ever-eligible for FSM. Following the EEF guidance for evaluations (Education Endowment Foundation, 2018b), any subgroup analysis is therefore considered to be exploratory.

Randomisation

Schools were recruited by the intervention delivery team, and randomisation was undertaken by the evaluation team. One member of the evaluation team generated random seeds (random numbers between 1 and 5,000,000 as generated from https://www.random.org) and another member of the evaluation team then ran the randomisation commands. The full details of the randomisation process and code were detailed in the Statistical Analysis Plan (and also reproduced in Appendix G).

The trial design applied school-level stratified randomisation; strata were geographic region and school-level historical Key Stage 1 attainment. Stratified randomisation (rather than simple randomisation) was chosen in order to facilitate the intervention training to be delivered to appropriately sized groups of schools within particular regions, and to increase the likelihood of the intervention and control groups having balanced prior Key Stage 1 attainment.

Geographic regions were defined and assigned per school by the intervention delivery team (Birmingham, Coventry, London, Plymouth, Reading, Somerset, or Swindon). Prior attainment information was sourced from publicly available information about schools from the Department for Education; the strata were implemented as quartiles based on the available school-level ‘Key Stage 1 average point score’ (for the cohort who undertook Key Stage 2 tests in 2017/2018), with an additional category for missing/unavailable data, across the 141 participating schools at the randomisation stage. Random allocation into the intervention group and the control groups was then implemented on a 1:1 basis within each geographic region within each prior attainment strata.

Analysis was not undertaken blind to the randomisation (intervention or control) allocation. Essentially, indicators/data need to be available and visible to users of statistical software otherwise they cannot be considered within any modelling. Mitigation/minimisation of bias was undertaken through the analysis involving the outcomes and models as specified within the Statistical Analysis Plan.

Statistical analysis

The analysis followed the Statistical Analysis Plan, which followed wider guidance (Education Endowment Foundation, 2018b). Analysis outside of the ONS SRS was undertaken with IBM® SPSS® Statistics 27 (IBM, Armonk, NY, USA), while analysis within the SRS was undertaken with IBM® SPSS® Statistics 24 (IBM, Armonk, NY, USA) and STATA SE 17 (StataCorp LLC, College Station, TX, USA).

Primary analysis

The analysis followed the ‘intention-to-treat’ principle: essentially, the analysis aimed to reveal the ‘effect’ on the outcome that can be inferred to follow from being assigned to the intervention group rather than the control group, calculated to account for any differences in prior attainment and across any other modelled factors (the randomisation strata/blocks).

The analysis of the primary outcome (science test scores) followed the Statistical Analysis Plan, which followed wider guidance (Education Endowment Foundation, 2018b). A multi-level model (which can be alternately referred to as a mixed model) was used to predict the pupil-level science test scores using the pupil-level prior attainment measures (Key Stage 1 reading category and Key Stage 1 mathematics category), a school-level intervention status indicator (whether the school was assigned to the intervention group or to the control group), and the school-level randomisation stratification indicators; the multi-level model included a ‘random effect’ on the school-level via the ‘random intercept’ archetype and used maximum-likelihood parameter estimation.

The core analysis code is detailed in Appendix J.

Secondary analysis

The same analytical modelling approach was used across the primary outcome and the secondary outcomes. A multi-level model was used to predict each pupil-level secondary outcome using the pupil-level prior attainment measures (Key Stage 1 reading category and Key Stage 1 mathematics category), a school-level intervention status indicator (whether the school was assigned to the intervention group or to the control group), and the school-level randomisation stratification indicators; the multi-level model included a 'random effect' on the school-level via the 'random intercept' archetype and used maximum-likelihood parameter estimation.

Analysis in the presence of non-compliance

Compliance essentially considers whether (and to what extent) schools delivered key aspects of the intervention as intended, and what implication(s) this might have to the inferred effect(s).

The primary outcome and the secondary outcomes analysis considered the 'effect' on the various outcomes that can be inferred to follow from being assigned to the intervention group rather than the control group. However, schools may have had different extents of engagement with the intervention; the 'intention-to-treat' effect following from being assigned to the intervention group could potentially differ to the effect of actually receiving the full intended scope of the intervention. If compliance was low, then the 'intention-to-treat' effect might underestimate the actual effect.

An 'instrumental variables' approach to analysis in the presence of non-compliance essentially aims to reveal the 'effect' of the intervention within schools that met all of the aspects of compliance, compared to the other intervention and control schools. Alternately, the approach aims to reveal the 'effect' of the intervention through increasing extents of compliance. The intervention is assumed to have an effect through (and only through) complete compliance (or alternately, has an effect through increasing compliance). Nevertheless, it remains possible that some extent of engagement, even if below the thresholds for compliance, may be sufficient to achieve an impact; it also remains possible that effects follow from different and multiple aspects of an intervention, which may not be included as indicators of compliance.

Indicators of compliance reflect aspects of the intervention (as per the theoretical model of change) that would entail an observed effect. The compliance indicators for the recommenced Focus4TAPS programme were detailed in the Statistical Analysis Plan. Specifically, compliance with the intervention involved:

- a school staff member attending the introductory meeting in 2019; the meeting was targeted at science leaders or coordinators, school leaders, or management such as assistant headteachers and headteachers, and those with equivalent roles/responsibilities (measured via attendance records collected by the developer) (minimum: attended by a relevant staff member);
- a school staff member who teaches Year 5 pupils attending the CPD days in 2020/2021 (measured via attendance records collected by the developer and teacher questionnaires) (minimum: three of six live sessions or online videos/materials for missed sessions, by Year 5 teachers); and
- overall use of Focus4TAPS lesson plans and work samples (measured via teacher questionnaires) (minimum: four assessment class lessons using Focus4TAPS).

If information was unavailable, for example through the Year 5 teacher not completing the teacher questionnaire and conveying the number of training days and lessons delivered, then the relevant aspect was considered not to be met. This was a conservative assumption that was intended to minimise the potential impact of missing data (i.e. mitigating against the potential risk of the 'instrumental variables' compliance analysis only considering a subset of schools). Nevertheless, this approach may underestimate the actual extent of compliance.

The final compliance indicator considering CPD attendance was operationalised from the teacher questionnaire reports as these allowed teachers to convey that they covered training material via live attendance and/or via video recordings (while attendance records from the intervention delivery team could only consider live attendance). For the considered Year 5 teachers, 49 schools met the compliance aspect around attending/viewing at least three of six live/recorded training sessions from teacher questionnaire reports (which could encompass covering material live or through recordings) compared to 33 from intervention delivery team records (which only considered live attendance).

A single binary school-level 'compliance' indicator (with values of 0 or 1) was created from these aspects (a value of 1 reflected that the school met the minimum on all three aspects; a value of 0 reflected that the school had not met the minimum on all three aspects). Additionally, a numeric school-level 'compliance' indicator (with values of 0–3) was calculated as the sum of how many aspects had been met. On the pupil-level, across all 2,882 pupils, the intervention/control indicator correlated to some extent with the binary school-level 'compliance' indicator ($R = .637$, $p < .001$) and with the numeric school-level 'compliance' indicator ($R = .901$, $p < .001$).

Analysis in the presence of non-compliance followed an 'instrumental variables' approach. The school-level 'compliance' indicator was predicted within a multi-level model using the pupil-level prior attainment measures, a school-level intervention status indicator (whether the school was assigned to the intervention group or to the control group), and the school-level randomisation stratification indicators (the school's geographical region and school-average historical attainment); the multi-level model also included a 'random effect' on the school-level via the 'random intercept' archetype. The 'predicted values' from this model were then saved, giving a 'predicted compliance' indicator (from the 'fixed' model parameters excluding the 'random' effects on the school-level). The primary outcome analysis was then reproduced, using this 'predicted compliance' indicator instead of the intervention status indicator.

The Statistical Analysis Plan intended for the compliance analysis to focus on the binary school-level 'compliance' indicator within multi-level modelling via SPSS MIXED software functions. Preliminary analysis considering the binary school-level 'compliance' indicator through logistic modelling encountered estimation/convergence problems, so the Statistical Analysis Plan was refined to involve the following:

- The 'instrumental variables' approach was undertaken with the binary school-level 'compliance' indicator via SPSS MIXED multi-level modelling.
- The 'instrumental variables' approach was also undertaken with the binary school-level 'compliance' indicator via STATA IVREGRESS (StataCorp LLC, College Station, TX, USA) modelling. This was additional unplanned exploratory analysis, undertaken as a confirmatory/robustness check and to also provide direct comparability with other evaluations that use this particular software. STATA IVREGRESS is a software function that performs 'instrumental variables' analysis, which has been applied in various other evaluations by the EEF (Pampaka, *et al.*, 2021).
- The 'instrumental variables' approach was also undertaken with the linear school-level 'compliance' indicator within SPSS MIXED multi-level modelling. There was additional unplanned exploratory analysis, undertaken in order to maximise insight and utilisation of the available information, and to also offer methodological comparability with other evaluations. Other evaluations by the EEF have undertaken unplanned exploratory analysis to undertake 'instrumental variables' analysis with linear compliance indicators (Roy, *et al.*, 2021).
- Additional unplanned exploratory analysis was undertaken through an alternate approach in order to potentially gain greater insight and understanding around the implications of different levels of compliance. This exploratory analysis involved applying the primary/secondary impact analysis modelling (using the same model specification as the core analysis) to compare pupils within control schools against pupils within intervention schools with different levels of compliance. Other evaluations by the EEF have undertaken exploratory analysis in this way (Culliney, *et al.*, 2021). Specifically, this analysis involved the following:

    - pupils within all of the control schools were compared against the pupils within intervention schools that met at least one of the three compliance aspects (omitting pupils within other intervention schools from the modelling);
    - pupils within all of the control schools were compared against the pupils within intervention schools that met at least two of the three compliance aspects (omitting pupils within other intervention schools from the modelling); and
    - pupils within all of the control schools were compared against the pupils within intervention schools that met all three of the compliance aspects (omitting pupils within other intervention schools from the modelling).

The default calculation of effect sizes as specified within the EEF guidance assumes a binary predictor: the unstandardised predictive coefficient of the intervention/control indicator is divided by the standard deviation (SD) of the outcome in order to produce a standardised effect size (Education Endowment Foundation, 2018b; Tymms, 2004). When considering an interval/continuous predictor (such as the numeric 'predicted compliance' indicator), the unstandardised predictive coefficient is multiplied by the SD of that predictor and then divided by the SD of the outcome in order to produce a standardised effect size (Tymms, 2004).

Missing data analysis

The prevalence/patterns of available or missing data were considered across pupils' characteristics through considering cross-tabulations and average proportions/values. The availability of test data and questionnaire data were also predicted through logistic regression models; this analysis helped to consider whether the likelihood of data being present or missing associated with other factors (such as gender and prior attainment), which would suggest the plausibility of data being 'Missing at Random'. Making inferences/estimates of any missing information assumes that missing data are 'Missing Completely at Random' or 'Missing at Random' (Rubin, 1976).

For efficiency/brevity, the analysis considered the availability or unavailability of test or questionnaire data (pupils being identified as completing a test or a questionnaire or not) rather than the question-level or item/factor-level response or non-response within the available test and questionnaire data.

The approach to creating inferences/estimates of missing information detailed in the Statistical Analysis Plan assumed access to IBM® SPSS® Statistics with all available modules. However, the IBM® SPSS® Statistics 24 software within the ONS SRS environment lacked the 'complex samples and testing' module that allows comprehensive analysis of the prevalence/patterns of missing data, an empirical test of 'Missing Completely at Random' circumstances (Little's MCAR test), and multiple-imputation to create and analyse inferences/estimates of missing data. Multiple-imputation was therefore undertaken using STATA SE 17 through chained equations, creating five imputations. Ideally, multiple-imputation would create estimates of any/all missing data; however, preliminary work revealed that estimates could not be calculated for any missing Key Stage 1 information (Key Stage 1 categories were accommodated as input into the creation of inferences/estimates of other missing information; however, the creation process was unable to converge on solutions when attempting to also create inferences/estimates of missing Key Stage 1 categories).

The implementation of multiple-imputation created inferences/estimates of missing information for the primary and secondary outcomes:

- science attainment test scores;
- interest and enjoyment in science;
- confidence in science;
- perceptions of science teachers;
- self-regulation of learning in science;
- self-efficacy for Working Scientifically;
- working Scientifically beliefs; and
- wider benefits of science.

The implementation considered the primary and secondary outcomes, together with:

- school intervention or control status;
- school randomisation stratification indicators;
- age in months (as of the start of the 2020/2021 academic year);
- gender as reported within the NPD or gender as reported within the NPD supplemented/updated by questionnaire reports (both entailed the same subsequent findings);
- disadvantaged/advantaged status via the pupils being ever-eligible for FSM;
- Key Stage 1 reading category;
- Key Stage 1 mathematics category;
- pupils having parents who attended university or not (as reported by pupils via the questionnaire);
- pupils having a family member who works within a science-related job or not (as reported by pupils via the questionnaire); and
- pupils thinking that their parents or guardians are interested in science or not (as reported by pupils via the questionnaire).

The final analysis considered the implications of missing data on the magnitude and significance of effect sizes and involved the following:

- The primary and secondary outcomes were predicted using the main model specification but where the Key Stage 1 indicators had additional categories to encompass missing/blank information: essentially, a category for subject-specific 'A / D / blank / missing information' was added alongside the existing subject-specific categories of BLW, PKF, WTS, EXS, and GDS. This essentially allowed pupils with missing Key Stage 1 information to be encompassed within the analysis.
- The primary and secondary outcomes were predicted using the main model specification, where the data included inferences/estimates of missing information for the primary outcome and secondary outcomes via multiple-imputation, and also through using categories to encompass missing/blank instances of Key Stage 1 information. This essentially allowed the analysis to encompass pupils with missing Key Stage 1 information, and also to encompass some (but not all) pupils where previously missing outcome information was replaced by estimates/inferences.

Subgroup analyses

Subgroup analysis considered whether the intervention can be inferred to have a different effect on different groups of pupils. Subgroup analysis was undertaken via interaction models; essentially, the primary outcome analysis was repeated with the addition of the subgroup indicator and the interaction between the subgroup indicator and the intervention indicator. A significant interaction term can then be interpreted as reflecting the intervention having a different effect across the subgroups.

The following groups were considered:

- disadvantaged/advantaged status via the pupils being ever-eligible for FSM (ever-eligible compared to not ever-eligible);
- gender (boys compared to girls);
- pupils having parents who attended university or not, as reported by pupils via the questionnaire ('Yes' compared to 'No' and 'I don't know');
- pupils having a family member who worked within a science-related job or not, as reported by pupils via the questionnaire ('Yes' compared to 'No' and 'I don't know'); and
- pupils thinking that their parents or guardians are interested in science or not, as reported by pupils via the questionnaire ('Yes' compared to 'No' and 'I don't know').

Additionally, given circumstances of potential disruption arising from the pandemic, the teacher questionnaire asked for a list of pupils who remained in school throughout this year (pupils who did not undertake any remote learning at home). This information was used to also compare listed pupils with all other pupils. Considering this additional subgroup was exploratory and not specified within the Statistical Analysis Plan.

The trial recruitment was not designed to achieve minimum sample sizes for any subgroups of pupils, including pupils who were ever-eligible for FSM. Following the EEF guidance for evaluations (Education Endowment Foundation, 2018b), any subgroup analysis is therefore considered to be exploratory.

Gender was considered as reported through the NPD (as of the spring 2021 school census), as reported by the pupils through the questionnaire, and as reported through the NPD supplemented/updated by the questionnaire reports (i.e. the latest questionnaire reports would then replace/supersede any legacy information within the NPD). Preliminary sensitivity analysis highlighted no differences in findings across these indicators.

Following the established processes and standards from the EEF (2018b), pupils ever-eligible for FSM were also considered as a separate subsample, regardless of the results from the interaction model. The primary analysis model was therefore reproduced for this subsample alone.

Additional analyses and robustness checks

Additional analysis involved various checks to consider the robustness of the findings.

Analysis with additional pupil-level information

Further analysis supplemented the main analytical model (predicting science test scores) with additional pupil-level information (age, gender, and ever-eligible for FSM status), in order to encompass other potential characteristics and aspects of life that might influence and otherwise associate with science attainment. The additional pupil-level information was sourced from the NPD as of the spring 2021 school census: age in months (linear); gender (binary); ever-eligible for FSM status (binary). The analysis was also repeated with the students' contemporary questionnaire responses replacing older NPD information regarding gender if/as relevant, as a sensitivity check.

Analysis with Key Stage 1 detailed scores

Further analysis adapted the main analytical model (predicting science test scores), where the Key Stage 1 categories (from the NPD) were replaced by Key Stage 1 detailed test scores (provided by schools). However, relatively few Key Stage 1 test scores were provided by schools (many schools may not have retained and have had access to this information), so this analysis could only be indicative. English scores were provided for 855 pupils (29.7% of 2,882) and mathematics scores were provided for 820 pupils (28.5% of 2,882). The Statistical Analysis Plan originally intended to produce estimates/inferences for any missing Key Stage 1 test scores and then undertake the further analysis; however, this would have involved the analysis considering more estimates/inferences than original data. The further analysis therefore only considered the subset of pupils where detailed test scores were available.

Science attainment test subscales

Additional analysis considered the subscales from the science test; the primary outcome analysis model was reproduced with each test subscale score in place of the overall test score.

Additional unplanned (exploratory) analysis

All analysis was originally planned to be consistently undertaken with SPSS; however, circumstances unavoidably entailed that analysis involving multiple-imputation needed to be undertaken with STATA rather than SPSS. In order to consider whether different software (potentially implementing estimation approaches in slightly different ways) might potentially influence findings, the main analysis of the primary outcome and secondary outcomes was replicated across SPSS and STATA.

Given emergent findings from the IPE, exploratory analysis (not covered within the Statistical Analysis Plan) also considered the potential impact/influence of the pupils' reported frequency of undertaking practical work and science on the primary outcome.

Estimation of effect sizes

An effect size refers to a magnitude of impact or difference, which is usually interpreted together as an indicator of statistical significance (calculated through particular statistical tests or approaches such as predictive modelling). Statistical significance is shown through p-values, which broadly convey the extent of statistical uncertainty; the standard threshold for 'statistical significance' is a p-value below 0.05.

Research within educational and other social science fields usually involves considering and interpreting effect sizes and statistical significances together, rather than only focusing on the effect size in isolation or focusing on the statistical significance in isolation.

Observed differences

Observed differences in means (averages), such as across the pupils assigned to the intervention group and the pupils assigned to the control group, were conveyed through via Cohen's D values. The calculation of these values followed the formula detailed in the Statistical Analysis Plan, which followed wider guidance (Education Endowment Foundation, 2018b). Within the Focus4TAPS evaluation, these values quantified magnitudes of observed differences, such as across pupils assigned to the intervention group and pupils assigned to the control group for the baseline indicators.

Predictive modelling

The impact evaluation aimed to reveal effect sizes associated with the Focus4TAPS programme through predictive modelling. Essentially, this revealed predicted differences across those pupils assigned to the intervention group and those assigned to the control group, while accounting for other information such as their Key Stage 1 information.

The effect sizes from predictive modelling are conveyed through Hedge's g values. The calculation of these values followed the process and formulae detailed in the Statistical Analysis Plan (where the process is also briefly summarised below), which followed wider guidance (Education Endowment Foundation, 2018b; Tymms, 2004). Within the Focus4TAPS evaluation, these values provided the final 'effect sizes' that were inferred to be associated with the intervention programme.

Initially, 'unconditional' multi-level models (with no predictors) were used to reveal the relevant school-level variance and the pupil-level variance, and to therefore reveal the total variance of each outcome, which was then used to calculate the SD of each outcome. For each outcome, the analysis model revealed the unstandardised predictive coefficient related to the intervention status indicator, with associated confidence intervals (CIs) and statistical significance, all of which were directly produced by the statistical software (IBM® SPSS® Statistics or STATA). The effect size was then calculated by dividing the unstandardised predictive coefficient by the SD of the outcome, and the same process was followed for the CIs.

Within this process, the intervention status indicator was coded either as '0' reflecting membership of the comparison group or as '1' reflecting membership of the intervention group (i.e. as a binary indicator).

The effect size calculation process is slightly different when numeric (interval/continuous) indicators are predictors (Tymms, 2004), although this only occurred within the compliance analysis. When the numeric 'predicted compliance' indicator was modelled as a predictor, the unstandardised predictive coefficient was multiplied by the SD of that predictor and then divided by the SD of the outcome in order to produce the standardised effect size.

The EEF interprets effect sizes from predictive modelling into 'months of additional progress' for primary outcomes that involve attainment such as test scores. Effect sizes from -0.04 to 0.04 are interpreted as reflecting 0 months of additional progress, effect sizes from 0.05 to 0.09 as reflecting 1 month of additional progress, effect sizes from 0.10 to 0.18 as reflecting 2 months of additional progress, effect sizes from 0.19 to 0.26 as reflecting 3 months of additional progress, effect sizes from 0.27 to 0.35 as reflecting 4 months of additional progress, effect sizes from 0.36 to 0.44 as reflecting 5 months of additional progress, and with further thresholds and interpretations for higher magnitudes.

Estimation of ICC

The calculation of ICCs (or ρ) followed the Statistical Analysis Plan, which followed wider guidance (Education Endowment Foundation, 2018b). 'Unconditional' multi-level models revealed the relevant school-level variance and pupil-level variance, and therefore revealed the total variance, for each outcome. The ICC was calculated by dividing the school-level variance by the total variance.

IPE

The IPE was informed by established guidance (Education Endowment Foundation, 2019b; Humphrey, *et al.*, 2016), and has considered eight established dimensions of implementation: fidelity; dosage; quality; reach; responsiveness; programme differentiation; monitoring of control groups; and adaptations. The process evaluation has considered further factors that may impact implementation: preplanning and foundations; the implementation support system; implementation environment; implementer factors; and intervention characteristics.

The process evaluation broadly has two purposes: assessing the fidelity, reliability, and other aspects related to delivery of the intervention; and, occurring alongside the impact evaluation, to help understand why any intervention effects were observed or not observed. Quantitative and qualitative data were collected to examine: feasibility; fidelity; reach/coverage of pupils; context (including facilitators and barriers to feasibility); and acceptability to various stakeholders.

The range of methods (surveys, interviews, observations of the training) has helped to produce comprehensive data against the various dimensions underlying the TAPS model in order to help inform 'theories of change'. The process evaluation has triangulated between the qualitative data and synthesis with the quantitative work helps inform theory verification and generation. The IPE included considering what happened at each phase of the intervention, whether compliance was achieved, and involved considering whether findings cohered with the logic model.

Research methods

Pupil interviews

In October 2020 all schools were invited to take part in the qualitative element. For the IPE, we conducted semi-structured group interviews of students from 10 schools that were audio-recorded. The project team were aware that the pandemic was having additional impact on teachers' time and that not all schools would have time to participate. We routinely invited schools to take part in the qualitative element until we had 10 case study schools (as agreed in the original proposal we would work with 10 schools).

Before we could proceed with the selection of students all schools were sent information sheets and consent letters to hand out to parents. Of those who returned consent forms, we asked schools to select students who were confident speakers and as far as was possible a mix of students from different backgrounds (gender, ethnicity, and FSM eligibility). Each school had students bunched into two groups of five students where possible. Group interviews were then undertaken with the pupil volunteers, in order to efficiently gather their views. The interviewing process facilitated each pupil to convey their views, to mitigate against imbalanced contributions. In total, 100 pupils (with parental consent) agreed to participate in the interviews in Autumn Term 2020; however, on the day of one of the interviews one student was unavailable; we were unable to find a replacement as the school went into a second national lockdown (January 2021) and so 99 interviews with pupils across 10 schools took place. The final set of interviews of pupils took place in June 2021 and July 2021 and they were interviewed within their same groups. Interviews took place while teachers undertook the training but prior to the intervention being delivered in classrooms. Interviewing pupils towards the start and towards the end of the academic year was intended to help explore potential changes in reported views and experiences.

The IPE involved the methods illustrated in Table 5.

Table 5: IPE methods overview

Research methods	Data collection methods	Participants/ data sources	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
Teacher questionnaires (July 2021 to October 2021)	Online questionnaires with various multiple-choice, agreement scale, and free-text questions	Year 5 teachers and science leaders	Quantitative analysis including quantification and comparison of averages and frequencies across intervention and control groups Qualitative analysis through thematic summarisation of free-text responses	RQ1, RQ2, RQ3, RQ4, RQ5, RQ6, RQ7, RQ9, RQ11, RQ12, RQ13, RQ14, RQ15, RQ16	Teacher background / context, teachers' practices, teachers' experiences, and engagement with Focus4TAPS
Pupil questionnaires (June 2021 and July 2021)	Paper questionnaires with various multiple-choice and agreement scale questions; primarily designed for the impact	Year 5 pupils	Quantitative analysis including quantification and comparison of averages and frequencies across	RQ18	Pupil learning and orientations to science

	evaluation with some IPE items		intervention and control groups		
Teacher interviews (October 2021 and November 2021; June 2021 to January 2022)	Semi-structured individual online interviews that were audio-recorded	Year 5 teachers	Qualitative analysis through thematic summarisation	RQ1, RQ2, RQ3, RQ4, RQ5, RQ6, RQ7, RQ9, RQ11, RQ12, RQ13, RQ14, RQ15, RQ16	Teacher background / context, teachers' practices, teachers' experiences, and engagement with Focus4TAPS
Pupil interviews (October 2020 to December 2020; June 2021 and July 2021)	Semi-structured group online interviews that were audio-recorded	Year 5 pupils	Qualitative analysis through thematic summarisation	RQ1, RQ2, RQ8	Pupil learning and orientations to science; pupil experiences, and engagement with Focus4TAPS
Trainer interviews and observation (across 2021/2022)	Online observations of training sessions and interview with training provider	Year 5 teachers and science leaders / training provider	Qualitative analysis through thematic summarisation	RQ1, RQ2, RQ3, RQ4, RQ5, RQ6, RQ7, RQ9, RQ10	Intervention / delivery characteristics

IPE, implementation and process evaluation; Focus4TAPS, Focus for Teacher Assessment of Primary Science; RQ, research question.

Teacher questionnaires

Views from teachers can help affirm whether and how (and which particular) teaching practices were amended through the Focus4TAPS programme, while also providing another perspective onto potential effects and other implications of the programme. Views from teachers can also help clarify aspects related to the delivery of the programme, such as facilitating factors and barriers to how the programme was applied, which may have influenced compliance and fidelity to the programme.

Year 5 teachers and science leaders within control schools and intervention schools completed teacher questionnaires around the end of the 2020/2021 academic year, from July 2021 to October 2021.

The teacher questionnaire was provided through an online format, with one version for teachers within control schools and another version with additional intervention-specific questions for teachers within intervention schools. The questionnaire involved a range of multiple-choice items, agreement/disagreement scale items, and free-text response items. The questionnaire broadly gathered information about the teachers and their views, including their confidence in aspects of teaching and learning that might be relevant to (and influenced by) the Focus4TAPS programme.

Some items from the teacher questionnaire can be aggregated together into summary indicators, similarly to how items from the pupil questionnaire were aggregated into summary factors covering aspects such as confidence and interest in science. Nevertheless, the items from the teacher questionnaire were ultimately intended to be considered in detail, as items, in order to help clarify particular insights. Preliminary analysis considered item-level and factor-level information to consider/affirm that findings did not depend on analysis involving items rather than factors or vice versa.

The teacher questionnaire covered areas including the following (with indicators of reliability considered across the responses from the 111 Year 5 teachers):

- frequencies of applying diverse practices in science education (e.g. 'Relate the lesson to students daily lives', 'Ask students to explain their answers'; eight items; Cronbach's Alpha = .610);
- confidence in teaching aspects of science education including inquiry methods (e.g. 'Inspiring students to learn science', 'Explaining science concepts or principles by doing science experiments'; 10 items; Cronbach's Alpha = .905);
- confidence in covering aspects of assessment of science including Working Scientifically (e.g. 'Understand the learning objectives and criteria for success', 'Undertake pedagogical planning to elicit students science knowledge and skills'; 13 items; Cronbach's Alpha = .932); and

- reflections on teaching science to Year 5 pupils during the year covering perceptions of pupil benefits and aspects/changes in teaching approaches (e.g. 'My students have enjoyed their science lessons', 'My students have made good progress in science'; 15 items; Cronbach's Alpha = .849).

Teacher questionnaires were completed by 111 Year 5 teachers (i.e. the main teacher of the relevant class of pupils, who engages with the programme in intervention schools, with one main teacher per school). This sample involved 56 teachers from intervention schools (i.e. covering 56 of 61 intervention schools) and 55 teachers from control schools (i.e. covering 55 of 60 control schools). Teacher questionnaires were also completed by 112 science leaders, where some also had Year 5 teaching responsibilities (such that the Year 5 teacher and science leader samples overlap). The science leaders covered 58 from intervention schools (i.e. covering 58 of 61 intervention schools) and 54 from control schools (i.e. covering 54 of 60 control schools). The teachers/leaders who completed the questionnaires may not have answered every question.

Interviews with pupils within intervention schools

Pupil volunteers for interviews from intervention schools were sought via disseminating information sheets and consent forms to parents/families. We asked schools to select students who were confident speakers and as far as was possible a mix of students from different backgrounds (gender, ethnicity, and FSM eligibility). Each school had students bunched into two groups of five students where possible. Group interviews were then undertaken with the pupil volunteers, in order to efficiently gather their views. The interviewing process facilitated each pupil to convey their views, to mitigate against imbalanced contributions.

Pupil questionnaire

The pupil questionnaire was applied for the impact evaluation in order to measure the secondary outcomes via encompassing a range of items covering attitudes and orientations towards science (considered through the impact evaluation). The pupil questionnaire also included a small number of other items that were intended to help consider aspects of the wider Focus4TAPS approach to help inform the IPE (such as 'My teacher plans and discusses science lessons with us' and 'I check my own work to find out what I have learned in science').

Pupils completed the questionnaire during June 2021 and July 2021.

Interviews with teachers who delivered the Focus4TAPS training

Two sets of interviews were conducted with the teaching staff that took part in the delivery of Focus4TAPS lessons. All schools were invited to take part in the qualitative element. We were aware that the pandemic was having additional impact on teachers' time and that not all schools would have time to participate. We routinely invited schools to take part until we had 10 case study schools. Ten teacher interviews (between 10 to 15 minutes) from 10 schools were conducted in the first term of Focus4TAPS between October 2021 and November 2021 prior to the delivery of any Focus4TAPS lessons. The second set of interviews (between 50 minutes to 1 hour and 20 minutes) took part at the end of the delivery of TAPS (six teacher interviews late June 2021); due to the impact of the pandemic on teachers' time we were not able to collect the remaining teacher interviews until January 2022 (four were collected). The 10th school had their interviews rescheduled a number of times but no data was collected. They were offered the opportunity to respond to the interview with written comments, which the teacher agreed to, but no document was returned.

The following areas were explored through the interviews:

- teachers background characteristics;
- teacher practices that Focus4TAPS programme is trying to encourage;
- examples of lessons where teachers were able to embed the principles of Focus4TAPS;
- effect of the Focus4TAPS programme on what pupils do in their science lessons;
- how often TAPS principles were taught to Year 5;
- what difference the Focus4TAPS programme made for pupils;
- how feasible teachers found the delivery of the Focus4TAPS programme;
- whether the implementation of the Focus4TAPS programme was constant over the year;

- impact of school closures on access to the programme and whether this was particularly an issue for certain groups of pupils;
- what adaptations teachers made to fit Focus4TAPS into teaching;
- importance of the professional development;
- impact of training on teachers' attitudes towards science e.g. confidence and Working Scientifically;
- would you yourself be able to continue to implement Focus4TAPS without any further support or training;
- Did you find whether TAPS had any other impact on your teaching practice?;
- Was there an impact on your workload by taking part in TAPS either positively or negatively?;
- Did taking part in TAPS have any influence on feelings around job satisfaction and the way you identify as being a teacher?
- Will you continue to implement Focus4TAPS without any further support or training?
- I realise this school year, and the end of the previous one, have been very different to usual because of COVID-19. What if any difference has COVID-19 made to your implementation of the Focus4TAPS programme? (Probe: e.g. online teaching, sickness of staff, school closures during lockdowns, and pupil absences);
- Do you think there will be a sustained impact of the Focus4TAPS programme on pupils' attainment in science? Do you have a feel whether the TAPS programme has had an impact on your children's attainment?;
- Do you think schools engage in the TAPS programme for the purpose of boosting science attainment?;
- Do you think the programme impacted pupils in other ways? (e.g. interest in other subjects or increased motivation or engagement);
- Do you think the Focus4TAPS programme works particularly well for certain pupils? (Probe: e.g., depending on their gender, socio-economic status, ethnicity, or prior attainment?);
- How scalable do you think the Focus4TAPS programme is? (If this doesn't seem to be understood, ask about whether they feel the Focus4TAPS programme could be rolled out across the country.); and
- If the Focus4TAPS programme was to be widely implemented, have you any suggestions as to any changes that would be needed?

Approaches to analysis

The responses from the teacher questionnaire were analysed through quantitative approaches focused around comparing average extents of agreement/disagreement across intervention schools and control schools, and through qualitative approaches focused around identifying common points/themes conveyed through the various free-text responses.

The narrative responses from teacher and pupil interviews were similarly analysed through identifying common points/themes. Thematic analysis was undertaken; themes were guided by the instruments themselves.

Compliance

Compliance was measured/considered in a number of ways. The Focus4TAPS intervention delivery team kept attendance records to monitor how many training sessions teachers attended. In addition, the teacher survey asked teachers to report their training attendance. In addition, the primary trainer/provider of the intervention delivery was interviewed to provide an additional perspective on compliance.

Fidelity and adherence

Attendance for training days was recorded by the Focus4TAPS intervention delivery team. Sessions were spread out through the calendar year and pre-recorded training was available for those who were unable to attend the training. We observed training events and recorded details about how the training was being handled both within and outside of school lockdowns. The data recorded in the observations of training alongside with interviews with teachers, the trainer and pupils helped to monitor fidelity. The fidelity assessed the degree to which elements of the Focus4TAPS programme were delivered as planned, which included the training itself alongside teachers' perceptions of how much of the Focus4TAPS programme they were able to use in the classroom, how it was received by pupils, and the quality of the Focus4TAPS programme in being able to provide the skills needed for teaching and assessment.

Instrument bias

As researchers we have been trained in mitigating bias both in terms of the way one might interpret data and respond to answers in an interview alongside the development of instruments. All instruments were developed with mitigating bias in mind and aimed to facilitate participants to easily and clearly convey their views. The sample of case study schools for the IPE covered different types of schools across different geographical locations. Responses were authenticated by triangulating data from various sources: teachers; pupils; observations; and the Focus4TAPS intervention delivery team alongside the findings from the quantitative and qualitative work.

Training programme with teachers across the intervention

Evidence from the trainer interview, observation of materials, and CPD events alongside the teacher data sources affirmed that the format, mode, timing, and focus of the training appeared to be consistent across all geographical regions and schools and that the programme was evaluated by teachers in a similar way. The logic model indicated that teachers were required to attend online training or access the pre-recorded videos. During 2019/2020 the programme was delivered through three training sessions, each delivered in-person to regional groups of teachers. During 2020/2021, the programme was delivered through six training sessions, each delivered remotely (online) and with a follow-up video, delivered to regional groups of teachers. Teachers were also able to view a summary video for each session, so that they could 'catch up' if they were unable to attend a live session.

Costs

Information about the time and costs associated with the Focus4TAPS programme were gathered through the teacher questionnaire and directly from the delivery team. Questions for teachers covered the time (in hours) spent: attending Focus4TAPS training; preparing Focus4TAPS lesson plans, tasks, and other aspects for delivery; delivering Focus4TAPS within science lessons; and applying other Focus4TAPS aspects within the school such as approaches to assessment and other aspects of the TAPS pyramid. Further questions for teachers considered what existing and what new equipment, materials, and resources were used in order to deliver the Focus4TAPS programme (and how much these costed).

Determining the overall cost per pupil per year involved applying various assumptions and standards (Education Endowment Foundation, 2019a).

Costs are estimated from the perspective of schools (Education Endowment Foundation, 2019a). This entails that time spent in training is assumed to require an additional cost to schools (where teaching/learning then needs to be covered by others, whether existing or external staff). The time spent by teachers in training is essentially counted once through this process (through reflecting the time required for additional teaching cover) rather than counted twice (the time required for additional teaching cover and also the time spent by teachers in training).

Furthermore, additional costs are expected to be presented as monetary units in addition to time units (Education Endowment Foundation, 2019a). This involves estimating an hourly/daily cost for teaching cover and staff. Teaching cover costs were calculated through the following assumptions/calculations: the maximum M6 spine point for salary[3]; 13.8% employer National Insurance contributions[4]; 23.7% employer pension contributions[5]; and 1,265 hours per year or 195 days per year.[6] Nevertheless, teaching cover may involve agency overheads and could be calculated at different salary points (e.g. within the upper pay range for classroom teachers and including additional payment from Teaching and Learning Responsibilities).

---

[3] NASUWT The Teachers' Union. *Teaching Salary Scales* (accessed 20 December 2021):
https://www.nasuwt.org.uk/advice/pay-pensions/pay-scales/england-pay-scales.html
[4] GOV.UK. *National Insurance Rates and Categories* (accessed 20/12/2021):
https://www.gov.uk/national-insurance-rates-letters
[5] Teachers' Pension Scheme. *Calculating Contributions* (accessed 20 December 2021):
https://www.teacherspensions.co.uk/employers/managing-members/contributions/calculating-contributions.aspx
[6] GOV.UK. *School Teachers' Pay and Conditions* (accessed 20 December 2021):
https://www.gov.uk/government/publications/school-teachers-pay-and-conditions

Table 6: Teaching salary cost calculations

Classroom teacher salary point	Salary per year	Employer National Insurance contributions		Employer pension contributions		Total costs		
		Rate	Amount per year	Rate	Amount per year	Total per year	Hourly cost (total / 1,265)	Daily cost (total / 195)
M6	£36,961	13.80%	£5,100.62	23.68%	£8,752.36	£50,813.98	£40.17	£260.58

M, England teacher main salary scale.

Costs are based on the resources needed to implement the programme in comparison to the counterfactual (Education Endowment Foundation, 2019a). First, this involved clarifying the resources required for the programme. Input from the delivery team and teacher questionnaire responses revealed that the Focus4TAPS programme relies on existing facilities, equipment, and materials within science departments and school contexts (such as science-specific equipment including beakers, thermometers, magnifying glasses, sieves, magnets, etc.), which would also be used within ‘business as usual’ science teaching/learning. Nevertheless, these facilities, equipment, and materials are pre-requisite costs for the programme. Second, this entailed that time spent in training is required for the programme (reflecting an additional cost to schools) while time spent delivering the programme does not entail additional costs; essentially, science teaching/learning (and related assessment activities) would be delivered in some way within schools regardless of the Focus4TAPS programme.

Costs are estimated for the programme as it was implemented within the trial (Education Endowment Foundation, 2019a). The recommenced programme of delivery (as implemented) involved online training rather than in-person training, which entailed that travel and subsistence costs related to attendance were not applicable/considered within this process.

Costs are estimated using market practices where possible (Education Endowment Foundation, 2019a). This entailed estimating what training costs and programme access costs would be if these were to be applied outside of a trial. This principle is assumed to overrule the earlier principle of estimating costs for the programme as implemented (within the trial as implemented, schools were not charged costs for training or access to TAPS materials). Nevertheless, the delivery team affirmed that schools would not be charged for programme/material access, although in-person training may incur costs outside of a trial (in order to cover venue, catering, and other costs).

Costs were then estimated for the programme as if it was implemented over 3 years (Education Endowment Foundation, 2019a), essentially as if the same teachers continued to subsequently apply the programme to successive cohorts of Year 5 pupils. The cost per pupil per year was then calculated by dividing the cost per year by the number of involved pupils.

Timeline

Table 7: Timeline

Dates	Activity	Staff responsible/leading
December 2018 to April 2019	Recruitment	Bath Spa University
January 2019 to April 2019	Recruitment data collection (to include pupil names and information) prior to randomisation (2019/2020 Year 5 cohort)	University College London (UCL) Institute of Education
January 2019	Develop pupil survey	UCL Institute of Education
April 2019 to May 2019	Pupil survey data collection prior to randomisation (2019/2020 Year 5 cohort) (data not considered within this report)	UCL Institute of Education
June 2019	Randomisation completed, and schools notified of their randomisation status	UCL Institute of Education

Dates	Activity	Staff responsible/leading
June 2019 to July 2019	Focus for Teacher Assessment of Primary Science (Focus4TAPS) Introduction to Training and Project (2019/2020 Year 5 cohort delivery) (data not considered within this report)	Bath Spa University
September 2019	Focus4TAPS Training Day 1 – observation of the training (2019/2020 Year 5 cohort delivery) (data not considered within this report)	Training delivery: Bath Spa University Observation: UCL Institute of Education
September 2019 to October 2019	Pupil interviews data collection (2019/2020 Year 5 cohort) (data not considered within this report)	UCL Institute of Education
January 2020	Focus4TAPS Training Day 2 – observation of the training (2019/2020 Year 5 cohort delivery) (data not considered within this report)	Training delivery: Bath Spa University Observation: UCL Institute of Education
March 2020	Trial paused due to the COVID-19 pandemic	The Education Endowment Foundation (EEF)
June 2020	Re-affirming school participation (recommencement)	Bath Spa University
September 2020 to October 2020	Recommencement data collection (to include pupil names and information) (2020/2021 Year 5 cohort)	UCL Institute of Education
November 2020 to December 2020	Pupil interviews data collection (2020/2021 Year 5 cohort)	UCL Institute of Education
October 2020	Focus4TAPS Update to Training and Project arrangements (2020/2021 Year 5 cohort delivery)	Bath Spa University
November 2020	Focus4TAPS Training Day 1 – observation of the training (2020/2021 Year 5 cohort delivery)	Training delivery: Bath Spa University Observation: UCL Institute of Education
October 2021 and November 2021	Teacher interview data collection	UCL Institute of Education
January 2021	Focus4TAPS Training Day 2 – observation of the training (2020/2021 Year 5 cohort delivery)	Training delivery: Bath Spa University Observation: UCL Institute of Education
March 2021	Focus4TAPS Training Day 3 – observation of the training (2020/2021 Year 5 cohort delivery)	Training delivery: Bath Spa University Observation: UCL Institute of Education
June 2021 to July 2021	Science tests and pupil survey (2020/2021 Year 5 cohort)	UCL Institute of Education
May 2021 to July 2021	Pupil interview data collection (2020/2021 Year 5 cohort)	UCL Institute of Education
July 2021 to October 2021	Teacher survey data collection	UCL Institute of Education
June 2021/July 2021 and January 2022	Teacher interview data collection	UCL Institute of Education
March 2022	Submit initial draft report	UCL Institute of Education
March 2022 to October 2022	Peer review and finalisation of report	UCL Institute of Education

Impact evaluation

Participant flow including losses and exclusions

Figure 2: Participant flow diagram (two-arms)

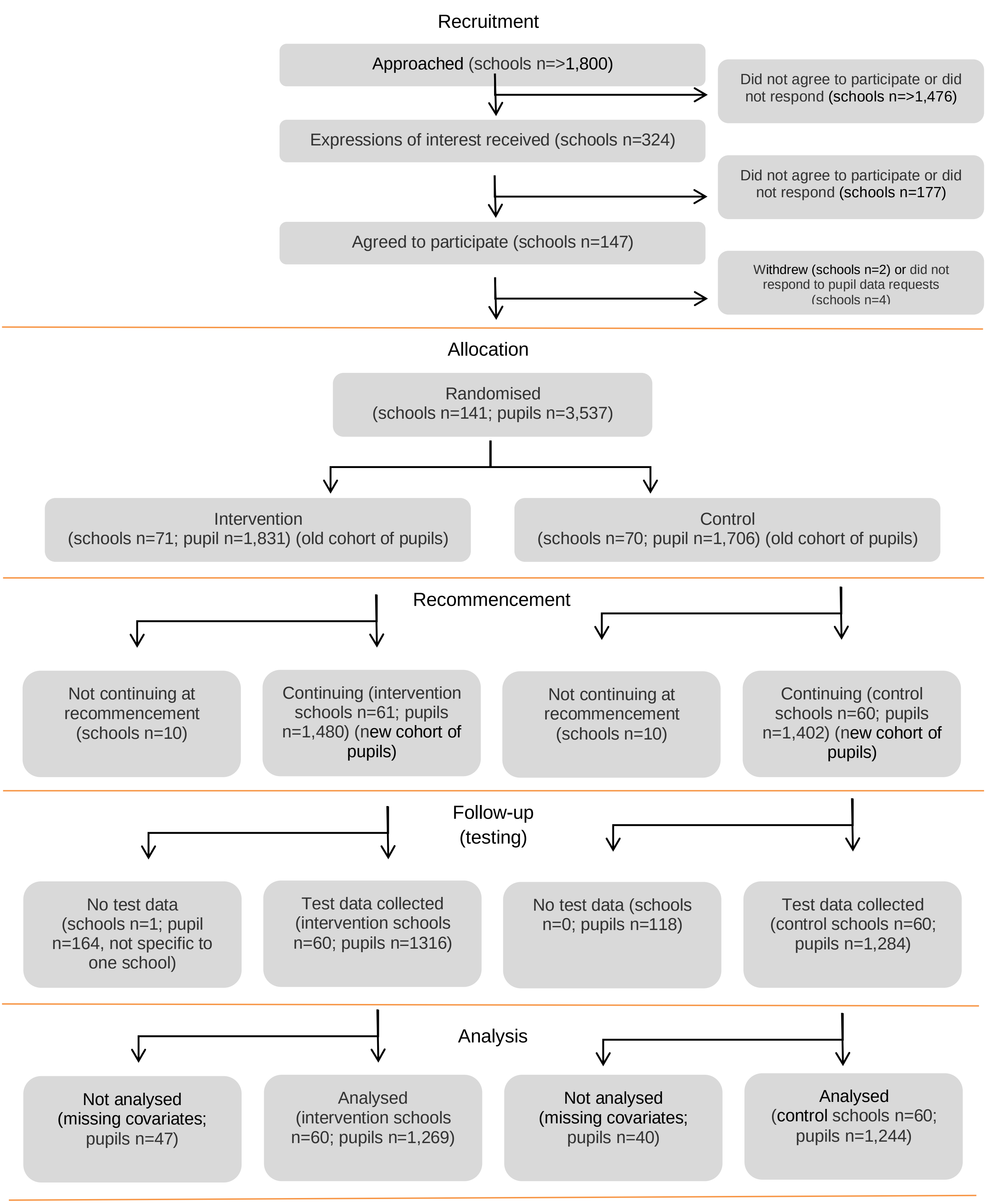

Figure 2 presents details of the participant flow through each stage of the trial. As previously conveyed, the Focus4TAPS trial design was a two-arm (intervention and control), cluster randomised efficacy trial where schools were the units of clustering/randomisation. The trial recruited schools during the 2018/2019 academic year and started to apply the Focus4TAPS programme with Year 5 pupils across some of the 2019/2020 academic year before pausing because of the COVID-19 pandemic. The trial then recommenced with some (but not all) of the same schools (with the same intervention and control allocations) but with a different cohort of Year 5 pupils across the 2020/2021 academic year.

Table 8: Minimum detectable effect size (MDES) at different stages

		Protocol		Randomisation		Recommencement		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.200	0.287	0.199	0.241	0.216	0.271	0.248	0.310
Pre-test/post-test correlations	Level 1 (pupil)	0.500	0.500	0.500	0.500	0.500	0.500	0.529	0.501
	Level 2 (class)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Level 3 (school)	0	0	0	0	0	0	0	0
Intracluster correlations (ICCs)	Level 2 (class)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Level 3 (school)	0.150	0.150	0.150	0.150	0.150	0.150	0.204	0.248
Alpha		0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Power		0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		25	3	25	6	24	5	21	5
Number of schools	Intervention	70	70	71	71	61	61	60	60
	Control	70	70	70	70	60	60	60	60
	Total:	140	140	141	141	121	121	120	120
Number of pupils	Intervention	1,750	210	1,831	461	1,480	325	1,269	366
	Control	1,750	210	1,706	434	1,402	293	1,244	342
	Total:	3,500	420	3,537	895	2,882	618	2,513	582

FSM, Free School Meals; N/A, not applicable.

Attrition

Table 9: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Recommenced	1,480	1,402	2,882
	Analysed	1,269	1,244	2,513
Pupil attrition (from recommencement to analysis)	Number	211	158	369
	Percentage	14.3%	11.3%	12.8%

As of randomisation, 3,537 pupils in Year 5 as of the 2019/2020 academic year were considered within the trial (1,831 pupils in 71 intervention schools and 1,706 pupils in 70 control schools). Following disruption due to the COVID-19 pandemic in 2019/2020, the trial recommenced with a new cohort of Year 5 pupils as of the 2020/2021 academic year within the schools that were willing and able to continue in the trial. As of recommencement, 2,882 pupil were considered within the trial (1,480 pupils in 61 intervention schools and 1,402 pupils in 60 control schools).

Almost all schools were able to test and survey their pupils in spring/summer 2021, except for one intervention school where the pupils were required to isolate due to the COVID-19 pandemic before testing/surveying could be undertaken. Test data were available for 2,600 pupils (1,316 in intervention schools and 1,284 in control schools) reflecting 90.2% coverage and 9.8% attrition at the stage of data collection. Questionnaire data were available for 2,501 pupils (1,261 in intervention schools and 1,240 in control schools) reflecting 86.8% coverage and 13.2% attrition at the stage of data collection. Other pupils may have been absent, may not have been inclined to complete a test, or may not have provided a name on their test such that the test could not be matched against any other information.

The analysis of the primary outcome (science test scores) considered 2,513 pupils (1,269 in intervention and 1,244 in control schools), reflecting 87.2% coverage and 12.8% attrition at the stage of analysis. Essentially, Key Stage 1 information was missing/unavailable for some pupils, so they could not be considered within the analysis even if they had completed a test. Nevertheless, sensitivity analysis, undertaken as part of the planned missing data analysis, allowed all of the 2,600 pupils who completed tests to be considered, which reflected 90.2% coverage and 9.8% attrition, and revealed the same findings as the main analysis. Essentially, because the Key Stage 1 information was categorical, sensitivity analysis could easily encompass the ‘missing/unavailable’ status as another discrete category.

Missing/unavailable data can potentially be considered in different ways, which may depend on different circumstances, which may influence the observed extent of attrition as of analysis. For example, when baseline attainment indicators are continuous (such as test scores or numeric grades), any missing/unavailable information usually entails that the relevant pupils cannot be considered within the main analysis. When baseline attainment indicators are categorical, it is possible that some trials might plan their main analysis such that missing/unavailable information entails that the relevant pupils cannot be considered within the analysis, while other trials might plan their main analysis such that ‘missing/unavailable’ is another discrete category and inherently allow more pupils to be considered within the analysis.

The Focus4TAPS trial intended for the main analysis to be comparable with the most commonly encountered approach of handling missing/unavailable information, where any missing/unavailable information entails that the relevant pupils cannot be considered (i.e. an analytical model only considers those pupils with available information for every indicator that is used within the model). This entailed that attrition as of analysis was 12.8%, which was greater than a threshold of 10%. Nevertheless, alternate analysis entailed that attrition as of (alternate) analysis was 9.8%, which was less than a threshold of 10%.

Pupil and school characteristics

School-level information was sourced from publicly available information from the Department for Education as of the summer of 2021 (Department for Education, 2021). This information was also used to provide a national-level context, through summarising information about all open 'Primary' and 'All-through' schools (16,916 schools across England), which encompasses the schools within the trial. This school-level information was not considered within the analysis of the primary outcome and secondary outcomes and is only provided to help readers consider how the sample may relate to other contexts.

Pupil-level information was sourced from the NPD. This pupil-level information was considered within the analysis of the primary outcome and secondary outcomes. Pupil-level 'ever-eligible' for FSM status reflected a pupil having been eligible for FSM at any time in the last 6 years up to the pupil's current year (not including nursery) as of the spring 2021 school census (via the EVERFSM_6_P_SPR21 indicator from the NPD).

Considering cross-tabulations and considering average proportions per school-level and pupil-level category provided equivalent results; the magnitudes of the observed differences (effect sizes) can be enumerated either as V (Cramer's V via chi-square tests) from cross-tabulations or as D (Cohen's D) from comparisons of average percentages or proportions. For example, a cross-tabulation of gender by intervention or control status can be alternately analysed through comparing the average proportion of girls (or boys) by intervention or control status.

The ONS disclosure controls prevent the reporting of school-level counts lower than 3 and pupil-level counts lower than 10, which occurred for a pupil-level Key Stage 1 category for one subject within one group in the recommenced sample. Given that a low count for one category requires the suppression of at least two categories (otherwise the numbers of pupils within one particular category can be calculated from the other categories and the total), the counts/percentages for the Key Stage 1 categories of BLW 'Below the standards of Pre-Key stage' and PKF 'Pre-Key stage' have been aggregated or suppressed as/when necessary (in order to mitigate against the potential possibility of low counts being identifiable through cross-referencing different tables within the report and appendices). This approach did not impact the reporting of any substantive findings for balance.

Randomisation

At randomisation, the sample of Year 5 pupils as of the 2019/2020 academic year consisted of 3,537 pupils within 141 schools (1,831 pupils within 71 intervention schools; 1,706 pupils within 70 control schools). Cross-tabulating and considering average proportions for the various categories where information/data was available revealed no differences in school-level or pupil-level numbers across the intervention and control schools for the considered categories, except for the reading Key Stage 1 category of WTS 'Working towards the expected standard': more pupils in the control group had the WTS outcome in reading at Key Stage 1 than the intervention group (intervention = 14.9% with WTS category (263 of 1,767 pupils with information available); control = 18.7% with WTS category (310 of 1,661 pupils with information available); $V = .051$ or $D = .101$, $p = .003$).

The full details are provided in Appendix K (Table 1 and Table 2 within Appendix K).

Overall, it was possible to infer that the intervention and control groups appeared to be comparable for the considered indicators at randomisation, with the partial exception of Key Stage 1 reading. Nevertheless, a difference across the intervention and control groups such as more (or less) pupils with one Key Stage 1 category such as WTS might suggest the presence of further differences involving less (or more) pupils with other categories, although this was not clearly observed.

Randomisation to recommencement changes

Most schools recommenced the trial following disruption due to the COVID-19 pandemic. Considering cross-tabulations to compare the 121 schools that recommenced with the other 20 schools revealed no differences in numbers across the school-level characteristics (intervention/control status, school type, urban/rural location, and Office for Standards in Education [Ofsted] rating).

The full details are provided in Appendix K (Table 3 within Appendix K).

The original cohort of pupils as of randomisation (Year 5 pupils as of the 2019/2020 academic year) was unavoidably replaced by a new cohort of pupils as of recommencement (Year 5 pupils as of the 2020/2021 academic year). There were no differences in pupil-level characteristics (the calendar month aspect of age, gender, ever-eligible for FSM status, Key Stage 1 reading, and Key Stage 1 mathematics) across the old and new pupils when considered within control schools and when considered within intervention schools, except for the reading Key Stage 1 category of EXS 'Working at the expected standard' for pupils within intervention schools (old intervention cohort = 54.4% with EXS category [961 of 1,767 pupils with information available]; new intervention cohort = 50.5% with EXS category [722 of 1,429 pupils with information available]; $V = .038$ or $D = .077$, $p = .030$).

The full details are provided in Appendix K (Table 4 and Table 5 within Appendix K).

Overall, it was possible to infer that the trial recommencement had minimal impact on the profile of schools and pupils within the trial.

Recommencement

After recommencement, the sample of Year 5 pupils as of the 2020/2021 academic year consisted of 2,882 pupils within 121 schools (1,480 pupils within 61 intervention schools; 1,402 pupils within 60 control schools). Cross-tabulating and considering average proportions for the various categories where information/data was available revealed no differences in school-level or pupil-level numbers across the intervention and control schools for the various categories, except for the reading Key Stage 1 category of WTS 'Working towards the expected standard': more pupils in the control group had the WTS outcome in reading at Key Stage 1 than the intervention group (intervention = 16.5% [236 of 1,429 pupils with information available]; control = 20.3% [275 of 1,352 pupils with information available]; $V = .049$ or $D = .099$, $p = .009$).

The full details are shown in Table 10, which focuses on observed numbers per indicator or category. Table 11, then shows the mean (average) proportion for each pupil-level indicator or category across the control and intervention schools and shows the magnitude of any observed differences in proportion.

Considered from an additional perspective, cross-tabulating the three non-suppressed categories (WTS, EXS, and GDS 'Working at a greater depth within the expected standard') across the intervention and control schools revealed an overall difference across the various numbers for reading ($V = .052$, $p = .028$) but no difference for mathematics ($V = .026$, $p = .419$). Statistical tests from cross-tabulations across all five Key Stage 1 categories cannot be reported because at least one category count was suppressed.

Overall, it was possible to infer that the intervention and control groups appeared to be comparable for the considered indicators, with the partial exception of Key Stage 1 reading. More pupils in the control group had the WTS outcome in reading at Key Stage 1 than the intervention group. While this suggests that fewer pupils in the control group would have other categories for reading (such as EXS or GDS), this was not definitively shown through comparing the intervention and control groups for each of these other categories.

The primary and secondary analysis mitigated any potential baseline imbalance through including the five considered Key Stage 1 categories within the analytical modelling, which would have been undertaken regardless of balance or imbalance. Further sensitivity analysis additionally encompassed those pupils with missing Key Stage 1 information to mitigate against balance/imbalance potentially relating in some way to missing/available information. The various information such as school location and Ofsted rating are summarised to help readers consider how the sample may relate to other contexts, and this school-level information is not considered within any of the analytical approaches. It remains unclear why some information such as Ofsted ratings was available for some but not all schools.

Table 10: Baseline characteristics of groups at recommencement

School-level (categorical) via publicly available indicators	National-level mean	Intervention group		Control group	
		n/N (missing)	Count (%)	n/N (missing)	Count (%)
School type: Academies	6,204 / 16,916 (0) = 36.7%	24 / 61 (0)	39.3%	32 / 60 (0)	53.3%
School type: Local Authority maintained schools	10,448 / 16,916 (0) = 61.8%	37 / 61 (0)	60.7%	28 / 60 (0)	46.7%
School type: Free schools	264 / 16,916 (0) = 1.6%	0 / 61 (0)	.0%	0 / 60 (0)	.0%
School location: Rural	4,879 / 16,916 (0) = 28.8%	18 / 61 (0)	29.5%	12 / 60 (0)	20.0%
School location: Urban	12,037 / 16,916 (0) = 71.2%	43 / 61 (0)	70.5%	48 / 60 (0)	80.0%
Ofsted rating: Good	10,620 / 14,044 (2,872) = 75.6%	42 / 55 (6)	76.4%	32 / 44 (16)	72.7%
Ofsted rating: Outstanding	2,068 / 14,044 (2872) = 14.7%	7 / 55 (6)	12.7%	6 / 44 (16)	13.6%
Ofsted rating: Requires improvement	1,291 / 14,044 (2872) = 9.2%	6 / 55 (6)	10.9%	6 / 44 (16)	13.6%
Ofsted rating: Serious Weaknesses	26 / 14,044 (2872) = .2%	0 / 55 (6)	.0%	0 / 44 (16)	.0%
Ofsted rating: Special measures	39 / 14,044 (2872) = .3%	0 / 55 (6)	.0%	0 / 44 (16)	.0%
		Intervention group		Control group	
School-level (continuous) via publicly available indicators		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)
FSM: school-level percentage	21.29 (14.53)	60 / 60 (1)	19.58 (13.05)	60 / 60 (0)	20.22 (13.86)
		Intervention group		Control group	
Pupil-level (categorical) via NDP indicators		n/N (missing)	Count (%)	n/N (missing)	Count (%)
FSM: ever-eligible: Yes	Not available	366 / 1,475 (5)	24.8%	342 / 1,395 (7)	24.5%
Gender: Girls	Not available	726 / 1,475 (5)	49.2%	716 / 1,395 (7)	51.3%
Gender: Boys	Not available	749 / 1,475 (5)	50.8%	679 / 1,395 (7)	48.7%
KS1 Reading category 1 BLW: Yes and category 2 PKF: Yes	Not available	67 / 1,429 (51)	4.7%	56 / 1,352 (50)	4.1%

KS1 Reading category 3 WTS: Yes	Not available	236 / 1,429 (51)	16.5%	275 / 1,352 (50)	20.3%	
KS1 Reading category 4 EXS: Yes	Not available	722 / 1,429 (51)	50.5%	672 / 1,352 (50)	49.7%	
KS1 Reading category 5 GDS: Yes	Not available	404 / 1,429 (51)	28.3%	349 / 1,352 (50)	25.8%	
KS1 Maths category 1 BLW: Yes and category 2 PKF: Yes	Not available	64 / 1,429 (51)	4.5%	50 / 1,353 (49)	3.7%	
KS1 Maths category 3 WTS: Yes	Not available	254 / 1,429 (51)	17.8%	253 / 1,353 (49)	18.7%	
KS1 Maths category 4 EXS: Yes	Not available	778 / 1,429 (51)	54.4%	760 / 1,353 (49)	56.2%	
KS1 Maths category 5 GDS: Yes	Not available	333 / 1,429 (51)	23.3%	290 / 1,353 (49)	21.4%	
		Intervention group		Control group		
Pupil-level (continuous) via NPD indicators		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Difference
Age (NPD) in months at start of 2020/2021 academic year	Not available	1,475 / 1475 (5)	113.40 (3.56)	1,395 / 1,395 (7)	113.44 (3.53)	D = .010, p = .785

BLW, 'Below the standards of Pre-Key stage'; EXS, 'Working at the expected standard'; FSM, Free School Meals; GDS, 'Working at a greater depth within the expected standard'; KS, Key Stage; NPD, National Pupil Database; PKF, 'Pre-Key stage'; SD, standard deviation; WTS, 'Working towards the expected standard'.

Table 11: Baseline characteristics of groups at recommencement, comparison of pupil-level average proportions

	Intervention group			Control group			Difference	
Indicator	Mean	SD	N	Mean	SD	N	D	p-value
Pupil age (NPD) in months at start of 2020/2021 academic year	113.40	3.56	1,475	113.44	3.53	1,395	.010	.785
Pupil gender: 0=girls, 1=boys	.51	.50	1,475	.49	.50	1,395	.042	.260
Free School Meals: ever-eligible: 1=Yes	.25	.43	1,475	.25	.43	1,395	.007	.853
Pupil KS1 Reading category (as 1-5 scale)	4.01	.84	1,429	3.96	.81	1,352	.057	.136
Pupil KS1 Reading category 1 BLW and category 2 PKF: 1=Yes	.05	.21	1,429	.04	.20	1,352	.027	.483
Pupil KS1 Reading category 3 WTS: 1=Yes	.17	.37	1,429	.20	.40	1,352	.099	.009
Pupil KS1 Reading category 4 EXS: 1=Yes	.51	.50	1,429	.50	.50	1,352	.016	.665
Pupil KS1 Reading category 5 GDS: 1=Yes	.28	.45	1,429	.26	.44	1,352	.055	.145

Pupil KS1 Maths category (as 1-5 scale)	3.95	.80	1,429	3.95	.76	1,353	.009	.813
Pupil KS1 Maths category 1 BLW and category 2 PKF: 1=Yes	.04	.21	1,429	.04	.19	1,353	.039	.297
Pupil KS1 Maths category 3 WTS: 1=Yes	.18	.38	1,429	.19	.39	1,353	.024	.528
Pupil KS1 Maths category 4 EXS: 1=Yes	.54	.50	1,429	.56	.50	1,353	.035	.360
Pupil KS1 Maths category 5 GDS: 1=Yes	.23	.42	1,429	.21	.41	1,353	.045	.237

BLW, 'Below the standards of Pre-Key stage'; EXS, 'Working at the expected standard'; GDS, 'Working at a greater depth within the expected standard'; KS, Key Stage; NPD, National Pupil Database; PKF, 'Pre-Key stage'; SD, standard deviation; WTS, 'Working towards the expected standard'.

Teacher profile and characteristics

From the 121 schools that recommenced, teacher questionnaires were completed by 111 Year 5 teachers (Table 12) (i.e. the main teacher of the relevant class of pupils, who engages with the programme in intervention schools, with one main teacher per school). This sample involved 56 teachers from intervention schools (i.e. covering 56 of 61 intervention schools) and 55 teachers from control schools (i.e. covering 55 of 60 control schools). Teacher questionnaires were also completed by 112 science leaders (
Table 13), where some also had Year 5 teaching responsibilities (such that the Year 5 teacher and science leader samples overlap). The science leaders covered 58 from intervention schools (i.e. covering 58 of 61 intervention schools) and 54 from control schools (i.e. covering 54 of 60 control schools). The teachers/leaders who completed the questionnaires may not have answered every question.

The profile of Year 5 teachers who completed questionnaires appeared to be similar across the intervention and control groups (Table 12), including for their average number of years spent teaching and for the average proportion of those with an educational specialisation in science. Essentially, and within the context of the overall logic model, it is possible to infer that any observed findings (differences in pupil outcomes and differences in teaching practices) were less likely to follow from differences in teachers' background. Nevertheless, on average, more teachers in the intervention group compared to the control group conveyed that their post-secondary education (i.e. university degree or equivalent) involved STEM (science, technology, engineering, and mathematics) subjects; it remains unclear how/why such a difference may have arisen.

Table 12: Reported Year 5 teacher characteristics

	Intervention group			Control group			Observed difference	
Questionnaire item (scale)	Mean	SD	N	Mean	SD	N	D	p-value
By the end of this school year, how many years will you have been teaching altogether? (years)	10.93	8.39	56	12.31	8.03	55	.168	.378
If your major or main area of study was education, did you have a specialisation in science? (0=No, 1=Yes)	.13	.34	47	.06	.24	50	.232	.261
What is your gender? (0=women, 1=men)	.25	.44	56	.22	.42	55	.075	.695
What is your ethnicity? (categorised; 0=white/English, 1=BAME/diverse)	.04	.19	54	.05	.23	55	.083	.665
During your post-secondary education, what was your major or main area(s) of study? (categorised; 0=others, 1=science)	.18	.39	56	.04	.19	55	.466	.016

During your post-secondary education, what was your major or main area(s) of study? (categorised; 0=others, 1=STEM)	.23	.43	56	.05	.23	55	.518	.007
During your post-secondary education, what was your major or main area(s) of study? (categorised; 0=others, 1=STEM and psychology)	.29	.46	56	.11	.31	55	.450	.019

BAME, Black, Asian, and Minority Ethnic; SD, standard deviation; STEM, science, technology, engineering, and mathematics.

Table 13: Science leader reported characteristics

	Intervention group			Control group			Observed difference	
Questionnaire item (scale)	Mean	SD	N	Mean	SD	N	D	p-value
By the end of this school year, how many years will you have been teaching altogether? (years)	11.97	8.03	58	13.68	7.41	54	.221	.244
If your major or main area of study was education, did you have a specialisation in science? (0=No, 1=Yes)	.26	.44	50	.17	.38	47	.217	.286
What is your gender? (0=women, 1=men)	.14	.35	57	.20	.41	54	.167	.382
What is your ethnicity? (categorised; 0=white/English, 1=BAME/diverse)	.07	.26	56	.07	.26	54	.010	.958
During your post-secondary education, what was your major or main area(s) of study? (categorised; 0=others, 1=science)	.22	.42	58	.19	.39	54	.096	.613
During your post-secondary education, what was your major or main area(s) of study? (categorised; 0=others, 1=STEM)	.22	.42	58	.19	.39	54	.096	.613
During your post-secondary education, what was your major or main area(s) of study? (categorised; 0=others, 1=STEM and psychology)	.33	.47	58	.26	.44	54	.149	.431

BAME, Black, Asian, and Minority Ethnic; SD, standard deviation; STEM, science, technology, engineering, and mathematics.

Teacher changes or continuity

The trial recruitment occurred in the spring and summer of 2019 and schools were notified of their assignment to the intervention or control groups in June 2019; at that time, some schools may not have determined duties for teachers for 2019/2020 or staffing changes may have been occurring or may have subsequently occurred (e.g. schools may not have been able to assign teachers to classes at that time, schools may have been recruiting for a Year 5 teacher at that time, etc.).

The trial then commenced with Year 5 pupils during the 2019/2020 academic year. If necessary (if schools had been unable to specify particular teachers pre-randomisation or if there were other relevant staffing changes), schools provided any updated details of the new teacher and their class early in the academic year. From March 2020, the trial was unable to continue given learning disruption following from the COVID-19 pandemic. During the summer of 2020, the intervention delivery team engaged with the 139 schools to reaffirm their involvement through updated MoUs; 121 schools remained, with 61 schools remaining in the intervention group and 60 schools remaining in the control group. The schools retained their original assignment to the intervention group or to the control group. The trial then recommenced with a new cohort of Year 5 pupils during the academic year of 2020/2021. If necessary (if there were staffing changes), schools again provided any updated details of the relevant teacher and their class to the evaluation team early in the academic year.

Teacher questionnaires were completed around the end of the 2020/2021 academic year (from July 2021 to October 2021) by Year 5 teachers and science leaders within the schools that recommenced; teachers were not surveyed before or during the 2019/2020 academic year. It is not possible to consider whether/how teachers were similar or different in profile as of commencement and as of recommencement; teachers were not surveyed within schools that did not recommence the trial.

It is possible to consider the general extent of change or continuity in teachers, from the perspective of the 141 schools from randomisation and from the perspective of the 121 schools that recommenced. Similar patterns were seen from both perspectives. The details from across the 141 schools from randomisation are provided in Appendix K (Table 8 and Table 9 within Appendix K).

Pre-randomisation to 2019/2020 changes or continuity in teachers

From the 121 schools that recommenced (rather than from all of the 141 schools that were randomised): 47 schools had the same specified Year 5 teachers between pre-randomisation and 2019/2020 (20 of 61 intervention schools and 27 of 60 control schools); 37 schools had different Year 5 teachers (17 of 61 intervention schools and 20 of 60 control schools); and 37 were unable to specify a Year 5 teacher pre-randomisation (24 of 61 intervention schools and 13 of 60 control schools).

From the 121 schools that recommenced: 79 schools had the same specified science leader between pre-randomisation and 2019/2020 (40 of 61 intervention schools and 39 of 60 control schools); 40 schools had different science leader teachers (21 of 61 intervention schools and 19 of 60 control schools); and 2 were unable to specify a science leader teacher pre-randomisation (0 of 61 intervention schools and 2 of 60 control schools).

2019/2020 to 2020/2021 changes or continuity in teachers

From the 121 schools that recommenced, 112 schools had the same specified Year 5 teachers across 2019/2020 and 2020/2021 (58 of 61 intervention schools and 54 of 60 control schools), while 9 schools had different Year 5 teachers (3 of 61 intervention schools and 6 of 60 control schools).

Similarly, 112 schools had the same science leaders across 2019/2020 and 2020/2021 (56 of 61 intervention schools and 56 of 60 control schools), while 9 schools had different science leaders (5 of 61 intervention schools and 4 of 60 control schools).

From the nine schools with changes in Year 5 teachers across 2019/2020 to 2020/2021, seven schools (three intervention schools and four control schools) had continuity in science leaders; only two schools (both control schools) had changes in specified Year 5 teachers and also changes in science leaders.

The COVID-19 pandemic may have influenced many aspects of life, potentially including staffing and/or recruitment, which may have influenced the extent of continuity or change in staff or roles from 2019/2020 to 2020/2021.

Table 14: Changes or continuity in Year 5 teachers from the 121 schools that recommenced

Time period	Year 5 teachers	Intervention		Control		Total	
		Schools	%	Schools	%	Schools	%
Pre-randomisation to 2019/2020	Same	20 / 61	32.8%	27 / 60	45.0%	47 / 121	38.8%
	Different	17 / 61	27.9%	20 / 60	33.3%	37 / 121	30.6%
	Unable to specify a Year 5 teacher pre-randomisation	24 / 61	39.3%	13 / 60	21.7%	37 / 121	30.6%
	Same	58 / 61	95.1%	54 / 60	90.0%	112 / 121	92.6%

2019/2020 to 2020/2021	Different	3 / 61	4.9%	6 / 60	10.0%	9 / 121	7.4%

Table 15: Changes or continuity in science leaders from the 121 schools that recommenced

Time period	Science leaders	Intervention		Control		Total	
		Schools	%	Schools	%	Schools	%
Pre-randomisation to 2019/2020	Same	40 / 61	65.6%	39 / 60	65.0%	79 / 121	65.3%
	Different	21 / 61	34.4%	19 / 60	31.7%	40 / 121	33.1%
	Unable to specify a science leader pre-randomisation	0 / 61	.0%	2 / 60	3.3%	2 / 121	1.7%
2019/2020 to 2020/2021	Same	56 / 61	91.8%	56 / 60	93.3%	112 / 121	92.6%
	Different	5 / 61	8.2%	4 / 60	6.7%	9 / 121	7.4%

Outcomes and analysis

Primary analysis

Table 16: Variance and intracluster correlation coefficient (ICC)

	'Unconditional' multi-level model covariance parameters and derived values				
Outcome	Total modelled n (intervention; control)	Residual / other variance (σ^2)	School-level variance (τ^2)	Calculated ICC of outcome $= \tau^2 / (\sigma^2 + \tau^2)$	Calculated SD of outcome $= SQRT(\sigma^2 + \tau^2)$
Science attainment test scores (0–45 scale)	2,600 (1,316; 1,284)	58.771	15.087	.204	8.594

SD, standard deviation; SQRT, square root.

The primary outcome was science attainment, which was measured via science test scores. Science test scores were available for 2,600 pupils (1,316 intervention and 1,284 control). The ICC (or ρ) for the test scores was .204 across these 2,600 pupils (Table 16).

Considering the observed averages (unadjusted means), intervention pupils gained higher scores than control pupils (intervention: 25.31; control: 23.59; D = .201, p < .001) (Table 17).[7]

The primary analysis model considered 2,513 pupils where all relevant information was available; essentially, prior Key Stage 1 attainment information was unavailable for some pupils. When accounting for the pupils' prior Key Stage 1 attainment and the randomisation strata/blocks through the primary analysis model, intervention pupils were predicted to gain higher scores than comparison pupils (effect size: .174, p = .021) (Table 17).

[7] Magnitudes of observed difference can be calculated via different indicators. For the indicators and pupils considered here, the magnitudes were the same for Cohen's D and Hedges' g at three decimal places. The observed difference in science test scores across the intervention and control groups were: Cohen's D = .201 [95% CI: .124, .278], p < .001; Hedges' g = .201 [95% CI: .124, .278], p < .001.

The full model details and process of calculating the effect size is illustrated within Appendix L. According to thresholds from the EEF, the effect size on the primary outcome of science test scores via the modelling can be interpreted as reflecting 2 months of additional progress.

Table 17: Primary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n/N (missing)	Mean [95% CI] (SD)	n/N (missing)	Mean [95% CI] (SD)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (0–45 scale)	1,316 / 1,480 (164)	25.31 [24.84, 25.78] (8.69)	1,284 / 1,402 (118)	23.59 [23.13, 24.05] (8.38)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021

CI, confidence interval; SD, standard deviation.

The effect size as quantified through the primary analysis was smaller than the estimated MDES. Such circumstances have occurred within other education trials (Hanley, *et al.*, 2015; Speckesser, *et al.*, 2018). It is possible that the process and assumptions within MDES calculations (involving an algebraic formula with a relatively small number of specified inputs) may not necessarily reflect more complex data as analysed through particular forms of modelling. For example, the MDES calculations involved one single value that provided an approximation of the association between pupils' prior Key Stage 1 attainment and their contemporary attainment on the science test, while the primary analysis was able to consider the underlying detail of the pupils' Key Stage 1 attainment through particular categories for reading and for mathematics.

The test scores appeared to visually approximate normal distributions, although slightly negatively skewed (skewness statistics were -.400 for intervention pupils and -.244 for control pupils).

The histograms for the primary outcome are shown below; Figure 3 shows the histogram for the control group and Figure 4 shows the histogram for the intervention group.

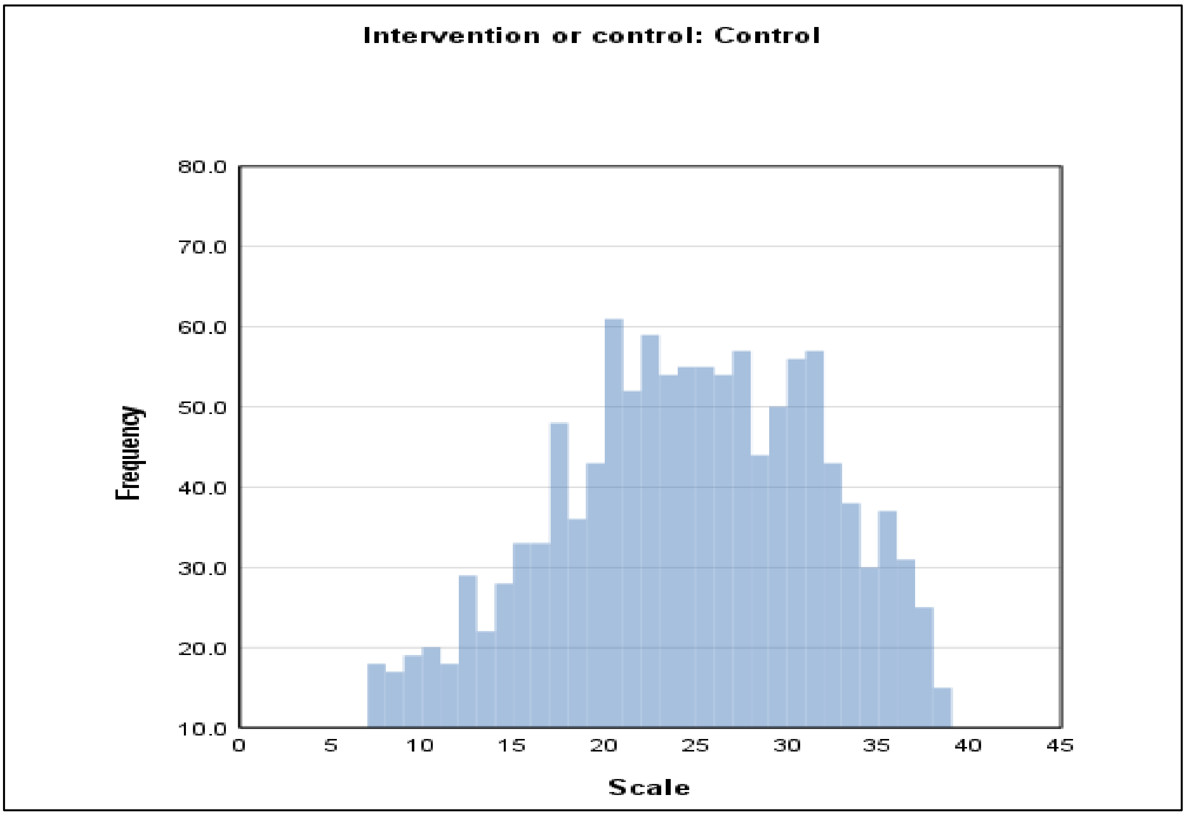

Figure 3: Science attainment test scores (0–45 scale) control group histogram

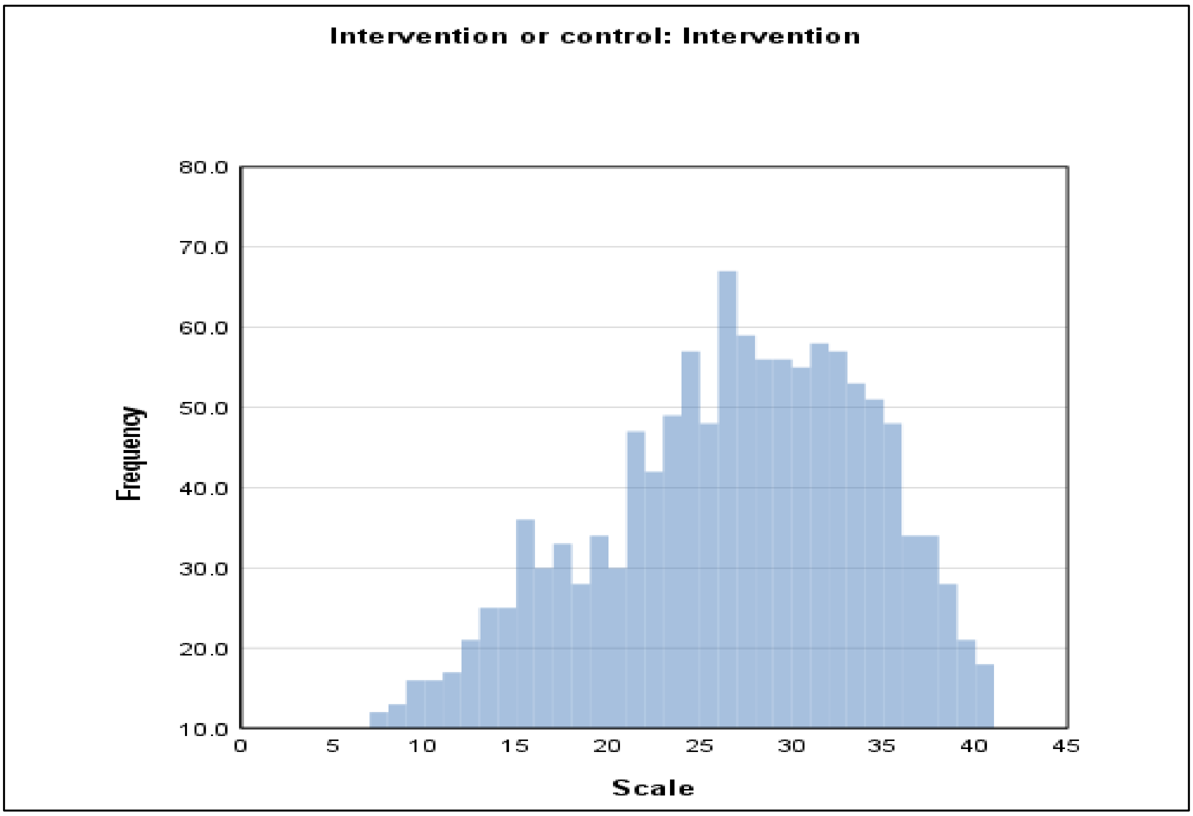

Figure 4: Science attainment test scores (0–45 scale) intervention group histogram

Secondary analysis

The secondary outcomes covered a range of science-related attitudes and beliefs, which were measured via questionnaire responses. Questionnaires were completed by 2,501 pupils (1,261 intervention and 1,240 control), although pupils may not necessarily have answered every question within the questionnaire. The ICC (or ρ) for the secondary outcomes are detailed in Appendix L.

Considering the observed averages (unadjusted means), no differences were apparent across the intervention and control pupils for any of the considered factors (detailed within Appendix M).

When accounting for the pupils' prior Key Stage 1 attainment and the randomisation strata/blocks through the analytical modelling, no differences were predicted across the intervention and control pupils for any of the considered factors. The effect sizes for the secondary outcomes cannot be interpreted as months of additional progress because the outcomes do not relate to attainment.

The secondary outcomes may not necessarily be independent, so the Bonferroni method of adjustment for multiple tests was relevant to interpreting statistical significance (p-values): the 'adjusted' threshold for statistical significance was .050 divided by 7 (the number of outcomes being considered), producing a threshold of .007 (rather than .050). Nevertheless, no observed differences in average responses or in predicted effects were revealed at either the .050 or the .007 threshold.

Table 18: Secondary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n/N (missing)	Mean [95% CI] (SD)	n/N (missing)	Mean [95% CI] (SD)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Interest and enjoyment in science (1–4 scale)	1,257 / 1,480 (223)	3.21 [3.17, 3.24] (.68)	1,233 / 1,402 (169)	3.19 [3.15, 3.23] (.69)	2,409 (1,215; 1,194)	.066 [-.070, .202]	.336
Confidence in science (1–4 scale)	1,255 / 1,480 (225)	2.86 [2.82, 2.89] (.66)	1,229 / 1,402 (173)	2.84 [2.80, 2.88] (.67)	2,403 (1,213; 1,190)	.024 [-.080, .128]	.650
Perceptions of science teachers (1–4 scale)	1,255 / 1,480 (225)	3.28 [3.25, 3.31] (.55)	1,224 / 1,402 (178)	3.30 [3.27, 3.33] (.53)	2,398 (1,213; 1,185)	-.028 [-.173, .116]	.696
Self-regulation of learning in science (1–4 scale)	1,252 / 1,480 (228)	2.88 [2.85, 2.92] (.67)	1,227 / 1,402 (175)	2.88 [2.84, 2.91] (.68)	2,398 (1,210; 1,188)	.012 [-.102, .127]	.832
Self-efficacy for Working Scientifically (1–4 scale)	1,245 / 1,480 (235)	3.10 [3.06, 3.13] (.64)	1,207 / 1,402 (195)	3.06 [3.02, 3.09] (.65)	2,371 (1,203; 1,168)	.057 [-.071, .185]	.376
Working Scientifically beliefs (1–4 scale)	1,244 / 1,480 (236)	3.45 [3.42, 3.47] (.48)	1,203 / 1,402 (199)	3.42 [3.39, 3.45] (.51)	2,366 (1,202; 1,164)	.044 [-.077, .165]	.471
Wider benefits of	1,249 / 1,480 (231)	3.47 [3.43, 3.50] (.59)	1,207 / 1,402 (195)	3.48 [3.45, 3.52] (.57)	2,375 (1,207; 1,168)	-.025 [-.133, .083]	.652

science (1–4 scale)	

CI, confidence interval; SD, standard deviation.

The distribution of the pupils' factor scores appeared to approximate normal distribution but were slightly negatively skewed for confidence in science (intervention skewness = -.313; control skewness = -.328). The distribution of the pupils' factor scores appeared to be more clearly negatively skewed for: interest and enjoyment in science (intervention skewness = -.742; control skewness = -.717); teachers and practices (intervention skewness = -1.002; control skewness = -.882); self-regulation of learning (intervention skewness = -.605; control skewness = -.479); self-efficacy for Working Scientifically (intervention skewness = -.664; control skewness = -.658); wider views about Working Scientifically (intervention skewness = -1.070; control skewness = -1.281); and wider benefits of science (intervention skewness = -1.258; control skewness = -1.338). Visually considering the graphical distributions, most responses appeared to be positive for pupils' perceptions of teachers and practices, wider views about Working Scientifically, and wider benefits of science.

The histograms for the secondary outcomes are provided in Appendix M.

Analysis in the presence of non-compliance

<u>Observed levels of compliance</u>

All of the intervention schools met at least one of the three aspects of compliance; over half of the intervention schools met the binary school-level 'compliance' standard (i.e. 36 of 61 schools met all three aspects of compliance). in detail:

- a total 0 of 61 schools met none of the three aspects;
- a total 12 of 61 schools (19.7%) met one of the three aspects;
- a total 13 of 61 schools (21.3%) met two of the three aspects; and
- a total 36 of 61 schools (59.0%) met all three aspects.

Considered from a cumulative perspective:

- a total 61 of 61 schools (100.0%) met one or more of the three aspects;
- a total 49 of 61 schools (80.0%) met two or more of the three aspects; and
- a total 36 of 61 schools (59.0%) met all three aspects.

Considered from the perspective of each specific aspect of compliance:

- A total 58 of 61 schools (95.1%) met the aspect of at least one relevant staff member attending the introductory meeting in 2019 (which involved science leaders, management staff such as headteachers, and/or equivalent staff attending). Specifically, 58 schools had attendance and 3 schools had no relevant staff able to attend the meeting (and were therefore classified as not having met this aspect of compliance; the delivery team provided a telephone introduction for those 3 schools).
- A total 49 of 61 schools (80.3%) met the aspect of the main Year 5 teacher attending at least three of six live training sessions (or online videos/materials for missed sessions) during 2020/2021. Specifically, 49 schools had Year 5 teachers complete a questionnaire and provide responses that met this compliance aspect, 7 schools had Year 5 teachers complete a questionnaire but their responses did not meet this compliance aspect, and 5 schools did not have Year 5 teachers who completed a questionnaire (and therefore were also classified as not meeting this compliance aspect).
- A total 39 of 61 schools (63.9%) met the aspect of at least four assessment class lessons using Focus4TAPS lesson plans and work samples being undertaken. Specifically, 39 schools had Year 5 teachers complete a questionnaire and provide responses that met this compliance aspect, 17 schools had Year 5 teachers who completed a questionnaire but their responses did not meet this compliance aspect, and 5 schools did not have Year 5 teachers who completed a questionnaire (and therefore were also classified as not meeting this compliance aspect).

Across these aspects of compliance, there were five schools who met the first aspect of compliance but did not have their Year 5 teacher complete a questionnaire, so were classified as not having met the second and third aspects of compliance. With this profile of available data, these schools had one of the three aspects met (and therefore were unable to be classified as having met all three aspects).

On the pupil-level, across all 2,882 pupils, the intervention/control indicator positively correlated with the binary school-level 'compliance' indicator ($R = .637$, $p < .001$) and with the numeric school-level 'compliance' indicator ($R = .901$, $p < .001$).

Instrumental variables compliance modelling: binary compliance via SPSS MIXED multi-level modelling

An 'instrumental variables' approach aims to reveal the 'effect' of the intervention in schools with different extents of compliance. An instrumental variables approach involves two stages: first, an observed compliance indicator is predicted by the intervention/control indicator (together with other predictors), in order to produce a predicted compliance indicator; second, the predicted compliance indicator is then used instead of the intervention/control indicator within the main outcome modelling (together with other predictors).

The 'instrumental variables' approach was undertaken with the binary school-level 'compliance' indicator within SPSS MIXED multi-level modelling. The full details, including the first and second stages of modelling and relevant parameters, are conveyed in Appendix N. The intervention/control indicator positively correlated with the binary 'predicted compliance' indicator from the first stage of the 'instrumental variables' approach ($R = .957$, $p < .001$, across 2,781 pupils with predicted values).

The final effect sizes are summarised in Table 19. The results affirm an effect on the primary outcome of science attainment test scores.

Table 19: Calculated effect sizes via instrumental variables compliance modelling using the binary compliance indicator (SPSS MIXED)

	Instrumental variables modelling via SPSS MIXED modelling using the binary compliance (0–1) indicator		
Outcome	Total modelled n (intervention; control)	Hedges g (95% CI)	p-value
Science attainment test scores (0–45 scale)	2,513 (1,269; 1,244)	.304 ([.047, .561]	.021
Interest and enjoyment in science (1–4 scale)	2,409 (1,215; 1,194)	.116 [-.121, .353]	.336
Confidence in science (1–4 scale)	2,403 (1,213; 1,190)	.042 [-.140, .223]	.650
Perceptions of science teachers (1–4 scale)	2,398 (1,213; 1,185)	-.050 [-.301, .202]	.696
Self-regulation of learning in science (1–4 scale)	2,398 (1,210; 1,188)	.021 [-.178, .221]	.832
Self-efficacy for Working Scientifically (1–4 scale)	2,371 (1,203; 1,168)	.100 [-.123, .323]	.376

Working Scientifically beliefs (1–4 scale)	2,366 (1,202; 1,164)	.077 [-.134, .288]	.471
Wider benefits of science (1–4 scale)	2,375 (1,207; 1,168)	-.043 [-.232, .146]	.652

CI, confidence interval.

Instrumental variables compliance modelling: binary compliance via STATA IVREGRESS modelling

The 'instrumental variables' approach was undertaken with the binary school-level 'compliance' indicator through STATA IVREGRESS modelling. This was additional unplanned exploratory analysis, undertaken as a confirmatory/robustness check and to also provide direct comparability with other evaluations that use this particular software (i.e. offering methodological comparability with other evaluations). The full details, including the first and second stages of modelling and relevant parameters, are conveyed in Appendix N.

The final effect sizes are summarised in Table 20. The results affirm an effect on the primary outcome of science attainment test scores across different statistical software.

Table 20: Calculated effect sizes via instrumental variables compliance modelling using the binary compliance indicator (STATA IVREGRESS)

	Instrumental variables modelling via STATA IVREGRESS modelling using the binary compliance (0–1) indicator		
Outcome	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (0-–45 scale)	2,513 (1,269; 1,244)	.282 [.027, .536]	.030
Interest and enjoyment in science (1–4 scale)	2,409 (1,215; 1,194)	.041 [-.193, .274]	.733
Confidence in science (1–4 scale)	2,403 (1,213; 1,190)	.020 [-.154, .194]	.821
Perceptions of science teachers (1–4 scale)	2,398 (1,213; 1,185)	-.061 [-.305, .183]	.624
Self-regulation of learning in science (1–4 scale)	2,398 (1,210; 1,188)	-.002 [-.196, .192]	.981
Self-efficacy for Working Scientifically (1–4 scale)	2,371 (1,203; 1,168)	.068 [-.142, .278]	.526
Working Scientifically beliefs (1–4 scale)	2,366 (1,202; 1,164)	.072 [-.122, .266]	.470
Wider benefits of science (1–4 scale)	2,375 (1,207; 1,168)	-.058 [-.230, .115]	.512

CI, confidence interval.

Instrumental variables compliance modelling: numeric compliance via SPSS MIXED multi-level modelling

The 'instrumental variables' approach was also undertaken with the linear school-level 'compliance' indicator within SPSS MIXED multi-level modelling. This was additional unplanned exploratory analysis, undertaken in order to maximise insight and utilisation of the available information, and to also offer methodological comparability with other evaluations.

The full details, including the first and second stages of modelling and relevant parameters, are conveyed in Appendix N. The intervention/control indicator positively correlated with the numeric 'predicted compliance' indicator from the first stage of the 'instrumental variables' approach ($R = .996$, $p < .001$, across 2,781 pupils with predicted values), calculated through using the numeric school-level 'compliance' indicator.

The observed and predicted compliance values were numeric (not binary) so the reported effect sizes were calculated as follows (Tymms, 2004): the unstandardised predictive coefficient was multiplied by the SD of that predictor and then divided by the SD of the outcome in order to produce the standardised effect size. These effect sizes are not directly comparable with those from the core analysis or from other forms of compliance analysis, because of the different calculation approach. These effect sizes essentially reflect an effect inferred from increasing compliance rather than an effect inferred from a difference across two groups (while accounting for the other predictors such as Key Stage 1 categories).

The final effect sizes are summarised in Table 21. The results affirm an effect on the primary outcome of science attainment test scores regardless of the handling of the compliance indicator (whether as binary or as numeric).

Table 21: Calculated effect sizes via instrumental variables compliance modelling using the numeric compliance indicator (SPSS MIXED)

	Instrumental variables modelling via SPSS MIXED modelling using the numeric compliance (0–3) indicator		
Outcome	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (0–45 scale)	2,513 (1,269; 1,244)	.088 [.013, .162]	.021
Interest and enjoyment in science (1–4 scale)	2,409 (1,215; 1,194)	.033 [-.035, .102]	.336
Confidence in science (1–4 scale)	2,403 (1,213; 1,190)	.012 [-.040, .064]	.650
Perceptions of science teachers (1–4 scale)	2,398 (1,213; 1,185)	-.014 [-.087, .058]	.696
Self-regulation of learning in science (1–4 scale)	2,398 (1,210; 1,188)	.006 [-.052, .064]	.832
Self-efficacy for Working Scientifically (1–4 scale)	2,371 (1,203; 1,168)	.029 [-.036, .093]	.376
Working Scientifically beliefs (1–4 scale)	2,366 (1,202; 1,164)	.022 [-.039, .083]	.471

Wider benefits of science (1–4 scale)	2,375 (1,207; 1,168)	-.012 [-.067, .042]	.652

CI, confidence interval.

Impact analysis at different levels of compliance

Additional unplanned exploratory analysis was undertaken through an alternate approach in order to potentially gain greater insight and understanding around the implications of different levels of compliance. This exploratory analysis involved applying the primary/secondary impact analysis modelling (using the same model specification as the core analysis) to compare pupils within control schools against pupils within intervention schools with different levels of compliance.

All of the 61 intervention schools met at least one of the three compliance aspects. Comparing the pupils within the 60 control schools against the pupils within the 61 intervention schools (that met at least one of the three compliance aspects) is the main analysis that was undertaken and reported by the impact evaluation. Essentially, the main analysis coincided with this particular level of compliance. The additional analysis then compared the pupils within the 60 control schools against the pupils within the 49 intervention schools that met at least two of the three compliance aspects (pupils with the other intervention schools were not considered within the modelling). The additional analysis then compared the pupils within the 60 control schools against the pupils within the 36 intervention schools that met all three compliance aspects (pupils with the other intervention schools were not considered within the modelling).

The full details from the analysis, covering the primary outcome and secondary outcomes in full (including the observed averages and the results from predictive modelling), is conveyed in Appendix N. Key findings are summarised as follows.

Considering the primary outcome of science test scores, the effect of the intervention appeared slightly higher in the schools that met all three compliance aspects (summarised in Table 22). Nevertheless, these findings are exploratory and indicative; the approach involved undertaking separate analysis across the different compliance circumstances and did not involve explicit tests to compare the magnitudes of effect across the different compliance circumstances.

Table 22: Control schools compared against intervention schools with different levels of compliance aspects: test scores

		Effect size	
Outcome (model features)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (All control schools compared against intervention schools that met one or more compliance aspects)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Science attainment test scores (All control schools compared against intervention schools that met two or more compliance aspects)	2,280 (1,036; 1,244)	.175 [.021, .329]	.026
Science attainment test scores All control schools compared against intervention schools that met all three compliance aspects)	2,006 (762; 1,244)	.208 [.056, .359]	.008

CI, confidence interval.

Considering the various secondary outcomes (summarised in full within Appendix N), the effects of the intervention remained similar across the different compliance circumstances with one partial exception: the effect on interest and enjoyment appeared to be higher in schools with higher levels of compliance although no findings were statistically significant (summarised in Table 23). Considering the observed averages (unadjusted means), pupils in intervention schools that met all three compliance aspects had higher interest and enjoyment in science, higher self-efficacy for Working Scientifically, and more positive beliefs around Working Scientifically, than control pupils; however, when controlling for prior attainment in the predictive modelling, the effect of the intervention was not statistically significant on any of these secondary outcomes (summarised in full within Appendix N). Overall, the results do not clearly show an association between interest/enjoyment in science and intervention compliance through this particular analytical approach (given the absence of statistical significance and the exploratory nature of the analysis); nevertheless, future evaluations considering similar areas may benefit from exploring associations between compliance and pupils' interest/enjoyment and potentially other attitudes and beliefs.

Table 23: Control schools compared against intervention schools with different levels of compliance aspects: interest and enjoyment in science

		Effect size	
Outcome (model features)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Interest and enjoyment in science (All control schools compared against intervention schools that met one or more compliance aspects)	2,409 (1,215; 1,194)	.066 [-.070, .202]	.336
Interest and enjoyment in science (All control schools compared against intervention schools that met two or more compliance aspects)	2,207 (1,013; 1,194)	.104 [-.041, .248]	.158
Interest and enjoyment in science (All control schools compared against intervention schools that met all three compliance aspects)	1,934 (740; 1,194)	.150 [-.011, .312]	.068

CI, confidence interval.

Missing data analysis

Patterns of available and unavailable data

Test data were available for 2,600 of 2,882 pupils (90.2% coverage). Questionnaire data were available for 2,501 of 2,882 pupils (86.8% coverage).

NPD attainment information is unavailable in some circumstances, such as when there was insufficient information for teachers to make judgements at Key Stage 1, or where information is blank/missing. Key Stage 1 reading information was unavailable for 101 of 2,882 pupils; Key Stage 1 mathematics information was unavailable for 100 of 2,882 pupils. There were no average differences in the proportion of Key Stage 1 reading or mathematics information being unavailable across the unavailable/available status of test or questionnaire data. NPD census information (providing indicators of age, gender, and ever-eligible for FSM status) was unavailable for 12 of 2,882 pupils, where these 12 pupils also did not have test data available or questionnaire data available.

The full details of the observed averages of pupil characteristics across data being available or unavailable are shown in Appendix O.

Separate logistic regression models (also provided in Appendix O) revealed that test data and questionnaire data were less likely to be available for pupils from intervention schools, for pupils with ever-eligible FSM status, and for some pupils with lower Key Stage 1 categories (test data were less likely to be available for pupils with the Key Stage 1 reading category of WTS compared to GDS; questionnaire data were less likely to be available for pupils with the Key Stage 1 reading categories of PKF, WTS, and EXS compared to GDS); age and gender were not associated with the availability of test or questionnaire data. These logistic regression models revealed independent associations, for example where the likelihood associated with intervention/control status was independent of the likelihood associated with ever-eligible FSM status. The trial circumstances were such that testing and surveying could not be undertaken within one of the 61 intervention schools, which may have contributed to some (although not necessarily all) of these results.

Analysis with missing categories covering Key Stage 1 outcomes

The main analysis was unable to consider all pupils because some Key Stage 1 information was blank and therefore missing/unavailable. The main analysis also considered Key Stage 1 categories of ‘A’ (Absent for long periods, recently arrived, and insufficient information for teachers to make judgements) and ‘D’ (Disapplied from the National Curriculum) as missing/unavailable data, given that these categories essentially reflect prior attainment information being unavailable.

Further analysis included additional categories to encompass missing/blank cases: a category for subject-specific ‘A / D / blank / missing information’ was added alongside the existing subject-specific categories of BLW, PKF, WTS, EXS, and GDS. This essentially allowed pupils with missing Key Stage 1 information to be encompassed within the analysis.

Overall, the same pattern of results was observed across the main analysis (as reported earlier) and the further analysis (as reported below), so it was reasonable to infer that the findings did not appear to be dependent on the handling of missing/available information. Encompassing more pupils within the modelling appeared to increase the statistical power of the models to reveal the predictive effects on science test scores (shown through improved p-values). The effect size on the primary outcome of science test scores still reflected 2 months of additional progress according to the thresholds from the EEF.

Table 24: Outcome analysis with missing categories covering Key Stage 1 outcomes

Analysis with categories for missing/unavailable Key Stage 1 information		Effect size	
Outcome	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (0–45 scale)	2,600 (1,316; 1,284)	.187 [.039, .336]	.014
Interest and enjoyment in science (1–4 scale)	2,490 (1,257; 1,233)	.068 [-.067, .202]	.321
Confidence in science (1–4 scale)	2,484 (1,255; 1,229)	.034 [-.070, .138]	.524
Perceptions of science teachers (1–4 scale)	2,479 (1,255; 1,224)	-.025 [-.166, .117]	.732
Self-regulation of learning in science (1–4 scale)	2,479 (1,252; 1,227)	.020 [-.093, .134]	.722
Self-efficacy for Working Scientifically (1–4 scale)	2,452 (1,245; 1,207)	.067 [-.059, .194]	.294
Working Scientifically beliefs (1–4 scale)	2,447 (1,244; 1,203)	.044 [-.077, .166]	.473

Wider benefits of science (1–4 scale)	2,456 (1,249; 1,207)	-.023 [-.134, .087]	.678

CI, confidence interval.

Analysis with missing categories covering Key Stage 1 outcomes and with estimates/inferences via multiple-imputation

Further analysis was undertaken using categories for missing/unavailable subject-specific Key Stage 1 information, and also including inferences/estimates of missing information for the primary outcome and secondary outcomes via multiple-imputation. This essentially allowed pupils with missing Key Stage 1 information to be encompassed within the analysis, and also encompassed some (but not all) pupils where missing outcome information had been replaced by estimates/inferences.

As before, the same pattern of results was observed across the main analysis (as reported earlier) and the further analysis (as reported below). The effect size on the primary outcome of science test scores still reflected 2 months of additional progress according to the thresholds from the EEF.

Table 25: Outcome analysis with missing categories covering Key Stage 1 outcomes and with estimates/inferences via multiple-imputation

Analysis with categories for missing/unavailable Key Stage 1 information, and with estimates/inferences of missing outcome information via multiple-imputation		Effect size	
Outcome	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (0–45 scale)	2,651 (1,340; 1,311)	.189 [.047, .332]	.009
Interest and enjoyment in science (1–4 scale)	2,499 (1,261, 1,238)	.068 [-.065, .200]	.318
Confidence in science (1–4 scale)	2,499 (1,261; 1,238)	.035 [-.068, .138]	.511
Perceptions of science teachers (1–4 scale)	2,499 (1,261; 1,238)	-.014 [-.152, .125]	.845
Self-regulation of learning in science (1–4 scale)	2,499 (1,261; 1,238)	.020 [-.092, .132]	.727
Self-efficacy for Working Scientifically (1–4 scale)	2,497 (1,260; 1,237)	.069 [-.054, .192]	.269
Working Scientifically beliefs (1–4 scale)	2,497 (1,260; 1,237)	.045 [-.074, .163]	.462
Wider benefits of science (1–4 scale)	2,498 (1,261; 1,237)	-.025 [-.134, .085]	.657

CI, confidence interval.

Subgroup analyses

Subgroup analysis considered whether the intervention had different effects on different groups of pupils, undertaken by interaction models: science test scores were predicted via the main analytical model with an additional 'subgroup' indicator and a 'subgroup × intervention' interaction term. A significant interaction term would reflect a differential effect across the subgroups.

The analysis considered the following groups:

- disadvantaged/advantaged status via the pupils being ever-eligible for FSM;
- gender as reported within the NPD, as reported by pupils through the questionnaire, and as reported within the NPD supplemented/updated by questionnaire reports;
- pupils having parents who attended university or not (as reported by pupils via the questionnaire);
- pupils having a family member who works within a science-related job or not (as reported by pupils via the questionnaire);
- pupils thinking that their parents or guardians are interested in science or not (as reported by pupils via the questionnaire); and
- pupils listed as remaining in school throughout the year (those who did not undertake any remote learning at home), as reported by teachers, compared to all other pupils.

Overall, there were no indications through this analytical approach that the impact of the intervention on science test scores varied across any of the considered subgroups. The full details of the modelling are conveyed in Appendix P.

Subgroup analysis was not undertaken for the secondary outcomes, because no main effects of the intervention were revealed for these outcomes.

Following the established processes and standards from the EEF (2018b), pupils ever-eligible for FSM were also considered as a separate subsample. Science attainment test scores were available for 601 pupils ever-eligible for FSM (296 intervention and 305 control pupils). Considering the observed averages (unadjusted means), the intervention pupils ever-eligible for FSM gained higher scores than the equivalent control pupils (intervention: 22.72; control: 20.27; D = .289, p <.001). However, when accounting for the pupils' prior Key Stage 1 attainment through the primary analysis model, no difference in test scores was predicted across intervention and control pupils (effect size: .149, p = .156). In this context, it is possible that the smaller number of modelled pupils may have limited the potential to reveal significant differences. According to thresholds from the EEF, the effect size on the primary outcome (science test scores) via the modelling of the pupils ever-eligible for FSM can be interpreted as reflecting 2 months of additional progress.

Table 26: Primary analysis for the subsample of pupils ever-eligible for FSM

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n/N (missing)	Mean [95% CI] (SD)	n/N (missing)	Mean [95% CI] (SD)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (0–45 scale) (FSM subsample)	296 / 366 (70)	22.72 [21.69, 23.74] (8.96)	305 / 342 (37)	20.27 [19.37, 21.17] (7.97)	582 (288; 294)	.149 [-.058, .356]	.156

CI, confidence interval; FSM, Free School Meals; SD, standard deviation.

Additional analyses and robustness checks

Analysis with additional pupil-level information

Further analysis supplemented the main analytical model (predicting science test scores) with additional pupil-level information, in order to encompass other potential characteristics and aspects of life that might influence and otherwise associate with science attainment. Overall, this analysis affirmed that the main findings do not appear to be dependent on the inclusion or exclusion of the considered pupil-level information (i.e. the effect sizes were almost unchanged in magnitude and significance). The effect size on the primary outcome of science test scores still reflected 2 months of additional progress according to the thresholds from the EEF. Nevertheless, it remains possible that any number of unmeasured aspects of life may be relevant to science performance/attainment.

Table 27: Primary analysis with additional pupil-level information

		Effect size	
Outcome (model features)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (Main analysis)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Science attainment test scores (Main analysis model + age in months at the start of the academic year)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Science attainment test scores (Main analysis model + gender as recorded within the NPD)	2,513 (1,269; 1,244)	.174 [.026, .321]	.021
Science attainment test scores (Main analysis model + ever-eligible for FSM status)	2,513 (1,269; 1,244)	.175 [.029, .322]	.020
Science attainment test scores (Main analysis model + age, gender [as recorded within the NPD], ever-eligible for FSM status)	2,513 (1,269; 1,244)	.175 [.028, .321]	.020
Science attainment test scores (Main analysis model + age, gender [as recorded within the NPD and supplemented/updated by questionnaire responses], ever-eligible for FSM status)	2,513 (1,269; 1,244)	.175 [.028, .321]	.020

CI, confidence interval; FSM, Free School Meals; NPD, National Pupil Database.

Analysis with Key Stage 1 detailed scores

Further analysis adapted the main analytical model (predicting science test scores), where the Key Stage 1 categories (from the NPD) were replaced by Key Stage 1 detailed test scores (provided by schools). According to thresholds from the EEF, the effect size on the primary outcome (science test scores) via the modelling that considered Key Stage 1 detailed test scores can be interpreted as reflecting 3 months of additional progress; however, this finding is not statistically significant.

This alternate analysis affirmed the plausibility of a positive direction of effect, regardless of the measurement of Key Stage 1 prior attainment, although it remains unclear whether the loss of significance relates to the intervention or relates to the lower sample size reducing the statistical power of the model to determine an effect as being significantly different to zero. Relatively few Key Stage 1 test scores were provided by schools (many schools may not have retained and have access to this information), so this finding can only be indicative. English scores were provided for 855 pupils (29.67% of 2,882) and mathematics scores were provided for 820 pupils (28.45% of 2,882). The Statistical Analysis Plan originally intended to produce estimates/inferences for any missing Key Stage 1 test scores, but this would have involved considering more estimates/inferences than original data.

Table 28: Primary analysis with Key Stage 1 detailed scores

Outcome (model features)	Total modelled n (intervention; control)	Effect size Hedges g [95% CI]	p-value
Science attainment test scores (Main analysis using Key Stage 1 categories)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Science attainment test scores (Alternate analysis using Key Stage 1 scores)	719 (445; 274)	.271 [-.031, .573]	.077

CI, confidence interval.

Science attainment test subscales

Further analysis considered the subscales from the science test; the primary outcome analysis model was reproduced with each test subscale score in place of the overall test score.

Considering the observed averages (unadjusted means), intervention pupils gained higher scores than control pupils for Working Scientifically test items (intervention: 8.87; control: 8.16; D = .185, p <.001), biology test items (intervention: 9.27; control: 8.70; D = .173, p <.001), chemistry test items (intervention: 7.24; control: 6.63; D = .188, p <.001), and physics test items (intervention: 8.80; control: 8.26; D = .179, p <.001). Analysis then applied the main analytical model to predict the subscales; when accounting for the pupils' prior Key Stage 1 attainment through the analysis, intervention pupils were predicted to gain higher scores than comparison pupils for all of these subscales from the science test. Overall, this analysis affirmed that the main findings do not appear to be specific to any particular focus and domain within the test. Nevertheless, findings from the test subscales can only be indicative; future research would need to apply specific and more extensive tests in order to specifically consider effects across topics and domains. According to thresholds from the EEF, the various effect sizes can be interpreted as reflecting 2 months of additional progress.

Table 29: Primary analysis for science attainment test subscales

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n/N (missing)	Mean [95% CI] (SD)	n/N (missing)	Mean [95% CI] (SD)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (0–45 scale)	1,316 / 1,480 (164)	25.31 [24.84, 25.78] (8.69)	1,284 / 1,402 (118)	23.59 [23.13, 24.05] (8.38)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Working Scientifically subscale (0–17 scale)	1,316 / 1,480 (164)	8.87 [8.66, 9.08] (3.88)	1,284 / 1,402 (118)	8.16 [7.96, 8.37] (3.75)	2,513 (1,269; 1,244)	.147 [.000, .294]	.049
Biology domain items subscale (0–15 scale)	1,316 / 1,480 (164)	9.27 [9.08, 9.45] (3.36)	1,284 / 1,402 (118)	8.70 [8.52, 8.87] (3.23)	2,513 (1,269; 1,244)	.152 [.010, .293]	.036
Chemistry domain items	1,316 / 1,480 (164)	7.24 [7.06, 7.41] (3.24)	1,284 / 1,402 (118)	6.63 [6.46, 6.81] (3.19)	2,513 (1,269; 1,244)	.167 [.019, .315]	.027

subscale (0–15 scale)							
Physics domain items subscale (0–15 scale)	1,316 / 1,480 (164)	8.80 [8.64, 8.97] (3.10)	1,284 / 1,402 (118)	8.26 [8.10, 8.42] (2.97)	2,513 (1,269; 1,244)	.145 [.011, .279]	.034

CI, confidence interval; SD, standard deviation.

Additional unplanned analysis: software

All analysis was originally planned to be consistently undertaken with SPSS; however, circumstances unavoidably entailed that analysis involving multiple-imputation needed to be undertaken in STATA rather than SPSS. In order to consider whether different software (potentially implementing estimation approaches in slightly different ways) might potentially influence findings, the main analysis of the primary outcome and secondary outcomes was replicated across SPSS and STATA. Overall, the findings were equivalent (only the main primary outcome analysis is reproduced below for brevity). Nevertheless, it is possible that slightly varying values from different software could become contextually or circumstantially relevant (e.g. a difference between significance values of .049 and .051 becomes relevant when the standard threshold for significance involves .050). The effect size on the primary outcome of science test scores still reflected 2 months of additional progress according to the thresholds from the EEF, regardless of the software being used.

Table 30: Primary analysis with different software

		Effect size	
Outcome (model features)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (Main analysis using IBM® SPSS® Statistics 24)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Science attainment test scores (Alternate analysis using STATA SE 17)	2,513 (1,269; 1,244)	.174 [.028, .320]	.019

CI, confidence interval.

Additional unplanned analysis: Key Stage 1 categories/scales

In order to consider whether the nature/format of the Key Stage 1 information might potentially influence findings, the main analysis of the primary outcomes was replicated where the Key Stage 1 categories were converted into scales (1–5 scales: 1 = BLW; 2 = PKF; 3 = WTS; 4 = EXS; and 5 = GDS). This analysis affirmed that the main findings do not appear to depend on the nature/format of the Key Stage 1 information. The effect size on the primary outcome of science test scores still reflected 2 months of additional progress according to the thresholds from the EEF, regardless of how Key Stage 1 information was handled.

Table 31: Primary analysis with Key Stage 1 scales

Outcome (model features)	Total modelled n (intervention; control)	Effect size Hedges g [95% CI]	p-value
Science attainment test scores (Main analysis using Key Stage 1 categories)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Science attainment test scores (Alternate analysis using Key Stage 1 information as 1–5 scales)	2,513 (1,269; 1,244)	.180 [.032, .328]	.018

CI, confidence interval.

<u>Additional unplanned analysis: analysis with pupil-level reports of frequencies of science teaching and teacher-level reports of science teaching/learning time</u>

Emergent findings from the IPE revealed that pupils and teachers within intervention schools, on average, reported that they encountered or delivered greater extents of teaching/learning around science. Pupils and teachers were surveyed towards the end of the academic year. As conveyed earlier within the trial methods and timeline, pupils completed questionnaires in June 2021 and July 2021 and teachers completed questionnaires in July 2021 through October 2021. Baseline surveying was considered to be an unfeasible burden given the pandemic and could not be undertaken; therefore, it remains unclear whether these observed differences in encountering or delivering extents of teaching/learning around science followed from an initial difference and/or from the Focus4TAPS programme.

The pupil questionnaire included questions of ‘How often do you do science at school?’ and ‘How often do you do investigations (practical work) in science?’, both measured through response categories of: (1) 'Never or almost never'; (2) 'Once or twice a month'; (3) 'Once or twice a week'; or (4) ‘Every day or almost every day’. The modelling considered these indicators as 1–4 scales (and results were unchanged when alternately considering these indicators as categories). Year 5 teachers were asked (via the teacher questionnaire) ‘In a typical week, how much time do Year 5 students spend learning science?’ and ‘In a typical week, how much time do Year 5 students spend learning science?’, where responses were converted into minutes if/as necessary. Pupils’ responses reflect subjective experiences that may or may not match other indicators; for example, responses might be influenced by more recent experiences in the summer term rather than earlier terms. Even if observed differences might also reflect an unobserved characteristic or orientation (such as being inclined to notice or report doing science, being keen or positive, etc.) the observed indicators are still relevant to consider as an avenue towards accounting for any unobserved characteristic or orientation (either the observed aspect of life or hypothetical unobserved aspects of life might link with attainment). Pupils’ experiences and views provide a valuable perspective and are meaningful and important to consider in themselves.

Table 32: Pupil questionnaire responses for reported frequency of science teaching/learning (observed averages)

	Intervention group			Control group			Observed difference	
Questionnaire item (1–4 scales)	Mean	SD	N	Mean	SD	N	D	p-value
How often do you do science at school?	2.81	.49	1,252	2.74	.56	1,227	.138	.001
How often do you do investigations (practical work) in science?	2.27	.66	1,226	2.17	.65	1,199	.151	<.001

SD, standard deviation.

Table 33: Pupil questionnaire responses for reported frequency of science teaching/learning (observed frequencies)

		Responses				Difference	
Questionnaire item	Group	Never or almost never	Once or twice a month	Once or twice a week	Every day or almost every day	Cramer's V	p-value
How often do you do science at school?	Intervention	1.6% (20 / 1,252)	18.7% (234 / 1,252)	76.9% (963 / 1,252)	2.8% (35 / 1,252)	.078	.002
	Control	3.7% (46 / 1,227)	20.8% (255 / 1,227)	73.5% (902 / 1,227)	2.0% (24 / 1,227)		
How often do you do investigations (practical work) in science?	Intervention	9.6% (118 / 1,226)	56.0% (687 / 1,226)	32.0% (392 / 1,226)	2.4% (29 / 1,226)	.076	.003
	Control	12.3% (148 / 1,199)	59.5% (714 / 1,199)	26.7% (320 / 1,199)	1.4% (17 / 1,199)		

Table 34: Teacher questionnaire responses (Year 5 teachers) for science teaching/learning time (observed averages)

	Intervention group			Control group			Observed difference	
Questionnaire item (scale)	Mean	SD	N	Mean	SD	N	D	p-value
In a typical week, how much time do Year 5 students spend learning science? (minutes, coded from free-text response)	103.09	33.05	55	89.06	37.18	48	.400	.047
In a typical week, how much time do you spend teaching science to Year 5 students? (minutes, coded from free-text response)	100.75	34.35	53	87.10	41.49	50	.359	.073

SD, standard deviation.

Given these differences, the impact evaluation undertook unplanned (exploratory) analysis to consider the potential impact or influence of the extent of science teaching/learning on the primary outcome: the pupils' reported frequencies of undertaking science work and/or the teachers' reports of science teaching/learning time were included as additional predictors within the main outcome modelling (i.e. these were additional covariates so that the analysis then 'controlled for' any differences in the extents of teaching/learning around science).

The additional analysis revealed that the Focus4TAPS intervention was positively associated with science test scores, regardless of the inclusion of these various indicators of the extent of science teaching/learning.

The additional analysis revealed different magnitudes of effect sizes associated with the Focus4TAPS intervention, depending on which particular additional indicators were considered with the modelling. Some of this variation may follow from different models considering different numbers of pupils; for example, when including teacher reports of science teaching/learning time, the analysis only considered those pupils where their Year 5 teacher completed the relevant questions within the questionnaire (while other pupils without responses from their Year 5 teacher are not considered).

Additional analysis (not summarised for brevity) also included Year 5 teachers' reports of 'If your major or main area of study was education, did you have a specialisation in science?' and 'During your post-secondary education, what was your major or main area(s) of study?', which were considered separately and also together with the reports of teaching/learning time and pupils' characteristics, and did not entail substantive changes to the intervention effect magnitude/significance. The results were also the same regardless of considering pupil gender as recorded within the NPD, or as recorded within the NPD and supplemented/updated by questionnaire responses. Further analysis affirmed that almost identical results emerged regardless of the operational handling of pupils' reported frequency of science teaching/learning (whether as linear or as categorical).

The Focus4TAPS programme was assumed to refine existing teaching/learning (and was not assumed to inherently require or involve more teaching/learning of science to be undertaken), although the programme may have had the unplanned/unexpected impact of inspiring or facilitating more science teaching/learning within intervention schools. Alternately, it is possible that baseline differences were present; it is also possible that baseline differences were present and the Focus4TAPS programme may have also inspired or facilitated more teaching/learning. Nevertheless, the effect of the intervention remained when controlling for various indicators of the extent of science teaching/learning; essentially, the revealed effect on science test scores did not appear to follow from different extents of teaching/learning.

In summary, and expressed another way, this additional modelling 'controlled for' differences in the extents of teaching/learning science, so that the effect of the intervention on science test scores can be inferred to occur independently from any other modelled aspect including the extent of teaching/learning science. Within educational research, these circumstances are sometimes described as the relevant predictor (the Focus4TAPS intervention programme) having an effect 'over and above' the other predictors (the extent of teaching/learning science, the pupils' prior Key Stage 1 attainment, and the trial randomisation strata).

The effect sizes on the primary outcome of science test scores reflect 2 or 3 months of additional progress according to the thresholds from the EEF, depending on which specific model is being considered.

Table 35: Primary analysis with additional pupil-level/teacher-level information

		Effect size	
Outcome (model features)	Total modelled n (intervention; control)	Hedges g [95% CI]	p-value
Science attainment test scores (Main analysis)	2,513 (1,269; 1,244)	.174 [.027, .321]	.021
Science attainment test scores (Main analysis + pupil reports of 'How often do you do science at school?')	2,349 (1,186; 1,163)	.169 [.016, .322]	.031
Science attainment test scores (Main analysis + pupil reports of 'How often do you do science at school?' [as a categorical predictor])	2,349 (1,186; 1,163)	.169 [.017, .322]	.030
Science attainment test scores (Main analysis + pupil reports of 'How often do you do investigations (practical work) in science?')	2,300 (1,162; 1,138)	.153 [.001, .305]	.049
Science attainment test scores (Main analysis + pupil reports of 'How often do you do investigations (practical work) in science?' [as a categorical predictor])	2,300 (1,162; 1,138)	.153 [.001, .304]	.049
Science attainment test scores (Main analysis + pupil reports of 'How often do you do science at school?', 'How often do you do investigations (practical work) in science?')	2,297 (1,161; 1,136)	.156 [.004, .309]	.045
Science attainment test scores (Main analysis + pupil reports of 'How often do you do science at school?', 'How often do you do investigations (practical work) in science?' [as categorical predictors])	2,297 (1,161; 1,136)	.157 [.006, .309]	.042

Science attainment test scores (Main analysis + pupil reports of 'How often do you do science at school?', 'How often do you do investigations (practical work) in science?', pupil age, gender [as recorded within the NPD], ever-eligible for FSM status)	2,297 (1,161; 1,136)	.155 [.004, .307]	.045
Science attainment test scores (Main analysis + Year 5 teacher reports of 'In a typical week, how much time do Year 5 students spend learning science?')	2,170 (1,156; 1,014)	.247 [.086, .408]	.003
Science attainment test scores (Main analysis + Year 5 teacher reports of 'In a typical week, how much time do you spend teaching science to Year 5 students?')	2,149 (1,102; 1,047)	.267 [.108, .426]	.001
Science attainment test scores (Main analysis + Year 5 teacher reports of 'In a typical week, how much time do Year 5 students spend learning science?', pupil age, gender [as recorded within the NPD], ever-eligible for FSM status)	2,170 (1,156; 1,014)	.245 [.084. .405]	.003
Science attainment test scores (Main analysis + Year 5 teacher reports of 'In a typical week, how much time do you spend teaching science to Year 5 students?', pupil age, gender [as recorded within the NPD], ever-eligible for FSM status)	2,149 (1,102; 1,047)	.265 [.107, .423]	.001

CI, confidence interval; FSM, Free School Meals; NPD, National Pupil Database.

Estimation of effect sizes

The calculation of effect sizes followed the Statistical Analysis Plan, which followed the wider guidance for evaluations by the EEF (Education Endowment Foundation, 2018b). This applied the 'intention-to-treat' principle interpreted as an aim to consider the maximum possible numbers of pupils (ideally the maximum numbers as originally assigned to the intervention and control groups) rather than only considering the smaller number of pupils considered within analytical models. This entailed that different elements within the effect size calculations could consider different numbers of pupils, although in many cases the considered numbers were still unavoidably limited by missing information/data.

As an illustrated example, considering the primary outcome (science test scores), an 'unconditional' multi-level model (with no predictors) considered 2,600 pupils and revealed the relevant school-level variance (τ^2 = 15.087) and the residual variance (which is commonly interpreted as reflecting pupil-level variance although technically reflects anything other than the school-level; σ^2 = 58.771), and therefore revealed the total variance (15.087 + 58.771 = 73.858), which was then used to calculate the SD of the outcome (SD = square root [SQRT] of 73.858 = 8.594).

The primary analysis multi-level model (with the intervention/control indicator, Key Stage 1 categories, and randomisation strata categories as predictors) considered 2,513 pupils and revealed the unstandardised predictive coefficient related to the intervention/control indicator (unstandardised coefficient = 1.496) with associated CIs (.230 and 2.763) and statistical significance (p = .021). The effect size was then calculated by dividing the unstandardised predictive coefficient by the SD of the outcome (effect size = 1.496 / 8.594 = .174), and the same process was followed for the CIs.

Outside of 'intention-to-treat' evaluations, effect sizes are often considered as standardised coefficients calculated from parameters consistently estimated across the number of pupils considered within analytical models. Further analysis affirmed that any differences following from different effect size calculation approaches appeared to be inconsequential for these particular pupils within this trial. For example, the analytical model for the primary outcome (science test scores) considered 2,513 pupils; considering an 'unconditional' multi-level model (with no predictors) for those 2,513 pupils (rather than all 2,600 pupils) produced an SD of 8.545, which produced an almost identical effect size (effect size = 1.496 / 8.545 = .175).

The various parameters are fully detailed in Appendix L.

Estimation of ICC

The calculation of ICCs (or ρ) followed the Statistical Analysis Plan, which followed the wider guidance for evaluations by the EEF (Education Endowment Foundation, 2018b). As an illustrated example, considering the primary outcome (science test scores), an 'unconditional' multi-level model (with no predictors) considered 2,600 pupils and revealed the relevant school-level variance (τ^2 = 15.087) and the residual/pupil-level variance (σ^2 = 58.771), and therefore revealed the total variance (15.087 + 58.771 = 73.858). The ICC was calculated by dividing the school-level variance by the total variance (ICC = 15.087 / 73.858 = .204).

Outside of 'intention-to-treat' evaluations, ICC may be calculated across the number of pupils that are considered within analytical models. For example, the analytical model for the primary outcome (science test scores) considered 2,513 pupils; considering an 'unconditional' multi-level model (with no predictors) for those 2,513 pupils (rather than all 2,600 pupils) revealed the relevant school-level variance (τ^2 = 14.485) and the residual/pupil-level variance (σ^2 = 58.527), and therefore revealed the total variance (14.485 + 58.527 = 73.012). The ICC is then calculated by dividing the school-level variance by the total variance (ICC = 14.485 / 73.012 = .198).

Implementation and process evaluation

Executive summary of this section

The IPE considered teachers' and students' views using quantitative and qualitative data.

COVID-19 circumstances

- During the 2020/2021 academic year, there was a national lockdown between January 2021 and March 2021, during which time most pupils were taught remotely.
- Interviews with pupils indicated that many undertook some science learning while learning at home, although many conveyed that they did not enjoy doing science at home, that their learning was impacted by the pandemic, and that they did not have the right resources to undertake science learning at home.
- The IPE found that the pandemic was a barrier for some schools in adhering to the intervention tasks. There was an impact on the implementation of the Focus4TAPS lessons, including less overall time and opportunity for practical and collaborative work. According to teachers' self-reports, teaching science inside of school helped them to adhere to gap tasks while implementing Focus4TAPS lessons. During school lockdowns, some teachers conveyed challenges around implementing the TAPS lessons and the gap tasks, some of this was due to lack of parental support particularly if they were families who faced other challenges.

Compliance

- As covered earlier within the impact evaluation, 49 of 61 schools (80.33%) met the aspect of the main Year 5 teacher attending at least three of six live training sessions (or online videos/materials for missed sessions) during 2020/2021.
- Training attendance records from the intervention delivery team showed that overall attendance was comprehensive: every intervention school had at least one person (any Year 5 teacher or science lead) attending a live training session during 2020/2021; and 54 of 61 schools (88.52%) had people (any Year 5 teacher or science lead) attending at least three of six live training sessions.
- Future research or trials may benefit from expanding compliance measures to accommodate the potential for involvement from other teachers (in addition to any expected Year 5 teachers), and their extent of teaching/learning with pupils.

Fidelity

- Fidelity was considered through teachers' reported understanding of the aims of the intervention and teachers' reported practices, and also considered teachers' reported extent of adapting or changing aspects of the intervention. Additionally, the evaluation considered whether teachers believed that the intervention was accessible to all pupils, given that the intervention was intended to be applied to reach all Year 5 pupils. All reported quotes are taken from the post-intervention interviews.
- Year 5 teachers within intervention schools who completed the teacher questionnaire conveyed that they understood the Focus4TAPS programme as aiming to encourage improvements around practical work and Working Scientifically including: more frequent/greater emphasis on practical work, developing children's skills, understanding, and confidence; improvements around formative assessment (including elicitation of children's knowledge, ongoing assessment by teachers to develop understanding lesson by lesson); and through wider TAPS principles such as Focused Assessment meaning assessing one element of Working Scientifically at a time, and with less focus around 'writing for the sake of writing'.
- Year 5 teachers within the intervention schools who completed the teacher questionnaire also conveyed that they applied the programme through applying more practical work around the TAPS ideals (including citing use of TAPS lesson plans, children being involved in planning inquiries, work being focused on different aspects of investigations at different times, and including opportunities for child-led learning), through a refined/adapted approach to planning, delivery, and assessment around the TAPS ideals (including citing the TAPS principle of focusing on one skill at a time, and with less emphasis on recording and writing everything), and through some wider applications across their schools (including using and also sharing the TAPS lesson plans, meetings with colleagues, and applying or planning to apply wider implementation across year groups).

- Some teachers also reported that they understood the Focus4TAPS programme as aiming to encourage children's autonomy and independence, and encourage teachers' confidence.
- Overall, Year 5 teachers appeared to understand and apply the programme with few adaptations, and they often conveyed that the programme was accessible to most pupils.
- An interview with the trainer suggested that she felt that the conversations she had with the teachers indicated that the adaptations envisaged in the logic model were being met.

Usual practice

- Year 5 teachers from control schools and from intervention schools gave similar responses around their schools being involved or not involved in other science programmes and initiatives, and for resources available or not available for Year 5 pupils.
- Year 5 teachers from control and intervention schools conveyed a relatively similar range of activities and approaches that can infer to be elements of 'usual practices' within teaching/learning science. These included: pupils being able to plan their own scientific investigations in order to answer their own questions (including making predictions), with attention towards identifying independent and dependent variables and how to control them as necessary (with teachers often noting the concept of fair testing); pupils taking measurements with a range of equipment while working safely; pupils recording and conveying results through various tables and charts (and as a scientific report with introduction, prediction, method, results, and conclusion sections); and pupils drawing conclusions with explanations and using results to make further predictions.
- Teachers within intervention schools also conveyed specific practices around their use of TAPS resources, such as a Working Scientifically wheel, and their use of Focus4TAPS principles, such as specifically focusing on one aspect of Working Scientifically at a time.

Dosage

- As covered earlier within the impact evaluation, 39 of 61 schools (63.93%) met the compliance aspect of at least four assessment class lessons using Focus4TAPS lesson plans and work samples being undertaken.
- Some disruption may have followed from the COVID-19 pandemic context during 2020/2021. Some interviewed pupils conveyed that learning at home had an effect on science learning and that almost half of the interviewed pupils reported that they did not have the right resources at home to take part in science learning. Teacher interviews indicated that there was some way around the pandemic by teaching the non-practical elements of the programme during online learning.
- As covered earlier within the impact evaluation, the intervention did not appear to have different impact on different groups/categories of pupils. The IPE teacher and pupil interviews did not reveal difference in experiences of the intervention across different groups/categories of pupils. One teacher conveyed that pupils with English as an Additional Language (EAL) were considered to find TAPS less accessible; conversely, one teacher via the questionnaire conveyed that pupils with EAL were perceived to develop their language through the practical experiences. Year 5 intervention teachers who completed the questionnaire tended to convey that the programme was considered to be accessible to pupils and especially for those who might have found routine teaching/learning less accessible (including conveying that they believed the reduced focus on written work helped make science more accessible across those with differences in skills/characteristics). There was also some (qualitative) data to suggest that higher ability children did better than lower ability pupils, but these findings were not uniform across the sample.

Compliance

Compliance measures the extent to which teachers adhered to the three elements of the intervention. Compliance analysis was undertaken through the impact evaluation. As conveyed earlier within the impact evaluation, all of the intervention schools met at least one of the three aspects of compliance; over half of the intervention schools met the binary school-level 'compliance' standard (i.e. 36 of 61 schools met all three aspects of compliance).

As conveyed earlier within the impact evaluation, considered from the perspective of each specific aspect of compliance: 58 of 61 schools (95.08%) met the aspect of at least one relevant staff member (science leader or management staff) attending the introductory meeting in 2019. Around 49 of 61 schools (80.33%) met the aspect of the main Year 5 teacher attending at least three of six live training sessions (or online videos/materials for missed sessions) during 2020/2021. Around 39 of 61 schools (63.93%) met the aspect of at least four assessment class lessons using Focus4TAPS lesson plans and work samples being undertaken.

On the school-level, average attendance by the main Year 5 teachers was four of six CPD sessions (live attendance) via training attendance from intervention delivery team and five of six sessions (live or video/recording) via teacher questionnaire reports. The attendance records from the intervention delivery team considered live attendance only, while the teacher questionnaire considered live attendance and engaging with the recordings (if a live session might have been missed). The compliance indicators focused on the main Year 5 teacher, although science subject leaders and additional teachers could also attend the training. From the perspective of the training attendance records from the intervention delivery team, overall attendance was comprehensive: every intervention school had at least one person attending a live training session during 2020/2021; 54 of 61 schools (88.52%) had any one person (any Year 5 teacher and science lead) attending at least three of six live training sessions. Some teachers may not have answered the particular compliance questions within the teacher questionnaire and information could have varied across sources for any number of reasons, so it is possible that the compliance indicators may not completely reflect actual practices.

The Focus4TAPS programme included access to a range of new teaching resources, which were not available for control schools, and science subject leadership guidance on how to embed the programme within the school, all of which equated to support for the Focus4TAPS classes. The structure of the training, via gap tasks, was intended to help teachers apply and adapt Focus4TAPS plans and activities within lessons. On the school-level, on average five lesson plans were undertaken by the main Year 5 teachers via the teacher questionnaire reports. As before, if teachers may not have answered the particular compliance questions, then it is possible that indicators may not completely reflect actual practices; additionally, the compliance indicators do not consider any lesson plans delivered by additional teachers and science leaders (unless the science leader was listed by the school as also being the main Year 5 teacher for the focused class).

Future research or trials may benefit from expanding compliance measures to accommodate the potential for involvement from other teachers (in addition to any expected Year 5 teachers), and their extent of teaching/learning with pupils.

To what extent do teachers, schools (and trainers) adhere to the intervention tasks and requirements and what factors affect compliance?

Triangulating from various data sources indicated that the pandemic was one common factor, which was a barrier to adhering to the intervention tasks. For example, Year 5 teachers within the intervention schools who completed the teacher questionnaire highlighted that the COVID-19 pandemic entailed some wider issues or difficulties, which may have influenced their implementation of the Focus4TAPS lessons, including:

- less overall time and opportunity for practical and collaborative work;
- complicated circumstances within schools (which could, e.g. limit group work and/or require more resource in order to have many smaller groups undertaking the same activities compared to larger groups); and
- general needs to streamline teaching/learning, prioritise key learning and skills, and be concise.

Some teachers highlighted that this limited or reduced coverage, and that lockdown limited practical (or any) work.

Cross-checking teachers' answers to that of the trainer 'What do you see as the teacher practices that the Focus4TAPS programme is trying to encourage?' indicated that schools and trainers were adhering to the intervention programme. Additionally, interviews with teachers in intervention schools indicated that teachers were better able to adhere to tasks while implementing Focus4TAPS lessons and gap tasks while teaching took place in school. During times of school lockdowns some teachers struggled with implementing the TAPS lessons and the gap tasks, much of this was due to lack of parental support particularly if they were families who were faced with other challenges.

During periods of school lockdowns or bubbles of pupils being sent home, learning was often reliant on parents being able to support their children. Teachers, in particularly challenging contexts, had to deal with issues around mental health, which invariably would impact the delivery of all kinds of teaching including TAPS. For example, a Year 5 teacher within an intervention school conveyed:

> *I think the mental health of a lot of our children as a result of COVID has really suffered, and in the catchment we're in there's a lot of mental health issues with parents as well, and with the children, and it's been probably the hardest year in teaching I've ever had … the families are under more pressure, they really struggled to support their children, they were very stressed, a lot of them have lost jobs and various different things, not had enough food and we're having to provide food, education, but a lot of support that the families used to have when I first started teaching just isn't there anymore, and there's a lot more being put on to schools and specifically on to teachers, and you're having to go that extra, well it's not just an extra mile, it's an extra 50 miles to kind of compensate for that, and it's all been like, it's always been hard at our school, we're in a challenging catchment, but the pandemic has magnified it massively.*

How does the degree to which teachers and schools respond to the intervention (and CPD) vary and what factors affect this?

As covered in more detail in a later section and Table 42, the intervention teachers who completed the questionnaire tended to be positive in their agreement to questionnaire items including 'The training prepared me for using Focus4TAPS in my classroom': 65.4% responded with 'Agree a lot', 30.8% 'Agree a little', 1.9% 'Disagree a little', and 1.9% 'Disagree a lot'.

Many of the interviewed teachers responded positively around the CPD and the support that they received. Some teachers reported that it was positive to have other teachers to speak to about what works and what does not. For example, a Year 5 teacher within an intervention school conveyed:

> *I learned so much and it was lovely to see the resource being used ... [she would say to herself] I'm gonna do that in class next week and I knew I would have known exactly what I needed to get and how to do it. So I would have obviously like to have more in-house sessions [training done online], but I think it's given me such confidence to teach it because there's such simple ideas, but teach science in a way that I wouldn't have thought up with myself. I like how it's got the key vocabulary on there and it's that the plans are really useful for people who are not confident in teaching science because they're absolutely foolproof. They're so easy to follow and they tell you exactly the assessment focus, what, the big questions they put it into real life context ... And yeah, it tells you how to assess what to look for in terms of assessment as well … whilst it still gives you lots of scope to be creative with it, it does really support you to use all the right terminology and know exactly what the children are taking from it because it gives you all of those indicators up to look for and to see if the children have understood it.*

Teachers also reported that the training impacted their attitudes to science and this came out more with those whose specialism was not science. For example, a Year 5 teacher within an intervention school conveyed:

> *Definitely [training impacted attitude to science]. I was always someone who enjoyed science anyway, but there was always a sense of I guess anxiety and teaching science in a classroom of 30 Year 5s. I enjoyed science myself, but I never ask. Or science was always quite a stressful thing. So I really enjoyed teaching science now [because of the training]. I love it, and especially seeing how, I guess enthusiastic […] [the programme lead] is about it and all of the other schools [on the training programme] that we worked with it kind of gave us [all teachers] a buzz about science and gave us a little bit more energy to kind of go at it in a completely fresh mindset. So, definitely [the training] affected my attitude towards teaching it. I look forward to it, whereas before I might have looked forward to it, been quite a lot more anxious about teaching it [historically before the training].*

Fidelity

Fidelity considers the extent to which teachers applied the intended intervention within the context of the aims and practices that were intended by the Focus4TAPS programme. Fidelity was considered through teachers' reported understanding of the aims of the intervention and teachers' reported practices and also considered teachers' reported extent of adapting or changing aspects of the intervention. Additionally, the evaluation considered whether teachers believed that the intervention was accessible to all pupils, given that the intervention was intended to be applied to reach all Year 5 pupils. Teaching can be applied to all pupils but may be more or less accessible to different pupils, so any potential inaccessibility may not necessarily reflect a lack of fidelity but remains relevant within an evaluation.

Year 5 teachers within the intervention schools who completed the teacher questionnaire conveyed (through their free-text responses to 'What do you see as the teaching practices that the Focus4TAPS programme is trying to encourage?') that they understood the Focus4TAPS programme as aiming to produce:

- improvements around practical work and Working Scientifically (including more frequent/greater emphasis on practical work, developing children's skills, understanding, confidence, and with less focus around writing for the sake of writing);
- improvements around formative assessment (including elicitation of children's knowledge, ongoing assessment by teachers to develop understanding lesson by lesson, and through the TAPS principle of Focused Assessment meaning assessing one element of Working Scientifically at a time); and
- some teachers also reported that they understood the Focus4TAPS programme as aiming to encourage children's autonomy and independence, and encourage teachers' confidence.

The trainer interview also uncovered she had come across evidence (quote reported in a later section) that basic teacher practices that the programme was trying to implement were being done at least at the minimum level and some went beyond that. The pupil interviews found that 67 of 93 interviewed pupils reported that their teacher let them plan and do their own practicals. When pupils were asked 'Does your teacher tell you how to do better in science?' a large proportion (85 of 93) indicated that they did. Many pupils (78 of 93 interviewed pupils) indicated that their teacher told them how well they were doing in science.

Year 5 teachers within the intervention schools who completed the teacher questionnaire also conveyed (through responses to 'In what ways have you applied the programme in your teaching and across your school?') that they applied the programme through:

- applying more practical work around the TAPS ideals (including citing use of TAPS lesson plans, children being involved in planning inquiries, work being focused on different aspects of investigations at different times, and including opportunities for child-led learning);
- through a refined/adapted approach to planning, delivery, and assessment around the TAPS ideals (including citing the TAPS principle of focusing on one skill at a time, and with less emphasis on recording and writing everything); and
- through some wider applications across their schools (including using and also sharing the TAPS lesson plans, meetings with colleagues, and applying or planning to apply wider implementation across year groups).

An interview with the trainer suggested that she felt that the conversations she had with the teachers indicated that the adaptations envisaged in the logic model were being met.

> *I definitely know that they were doing the lesson plans, so taking those lesson plans off the website and trying some in class. And some of them did loads of them, I think some of them did the minimum, one per term or four across the year or whatever, but they were definitely being done. And I am very sure about that because teachers came back and talked about it and gave examples … The people who had the capacity were doing innovative things, so they were taking the activities and doing them in a different way online because of school closures or changing them to home learning activities. So there was some really innovative stuff going on.*

Year 5 teachers within the intervention schools who completed the teacher questionnaire often conveyed (through responses to 'To what extent have you needed to adapt or change the TAPS activities to better fit your teaching, the students, and other circumstances?') that they needed or applied few if any adaptations, although some teachers variously highlighted:

- the need for attention around pace and topics;

- application of scaffolding/support depending on existing knowledge/circumstances such as Special Educational Needs (SEN);
- adjustments to apply the content/activities with wider ages of pupils; and
- adjustments given the COVID-19 pandemic (such as science work happening within smaller groups than might otherwise be used, and adaptations to use materials available at home).

The teacher interviews affirmed that teachers did indeed adapt resources, although this was not possible within all school contexts particularly where there was little support from parents or those from EAL backgrounds. The TAPS programme developer believed that while school lockdowns happened, the TAPS resources enabled teachers to adapt the resources and apply them within their own contexts and deliver TAPS-based lessons. The initial Focus4TAPS intervention conceptualisation included an expectation that teachers might begin by following lesson plans closely and then adapt and broaden their teaching and assessment of science as they become confident with the approach. Within this context, the teacher responses broadly support fidelity considered through the application of lesson plans, and fidelity with the underlying expectation or ideal that also recognises and applies adaptation when/where necessary.

Year 5 teachers within the intervention schools who completed the teacher questionnaire also tended to convey that they believed that the programme was accessible to all pupils and especially for those who might have found routine teaching/learning less accessible (including citing that they believed the reduced focus on written work helped make science more accessible across those with differences in skills/characteristics). For example, a Year 5 teacher within the intervention schools conveyed (in response to considering if the programme worked particularly well for all groups or certain pupil groups):

> *I think it did. It stretched the more able, made middle abilities more confident and the practical aspect was a revelation with the less able where many were able to shine with their organisational and practical skills.*

Nevertheless, some Year 5 teachers within the intervention schools highlighted that some pupils still required support, for example for those with lower skills, abilities, and confidence. There was an example with one of the teacher interviews where he felt pupils who were not fully fluent in English required some support and scaffolding to be able to benefit from the lessons and he also was not sure if this was because of the extremely disadvantaged context he taught in.

During the 2020/2021 academic year, schools had to manage teaching most of their pupils using online methods during a national lockdown between January 2021 and March 2021. Interviews with pupils and teachers indicated that it is difficult to determine how the programme was applied during lockdown and the extent of any disruption. It is possible that schools faced challenges and were less able/unable to cover science. The pupil interviews indicated that 70 of 98 interviewed intervention pupils reported that they took part in science lessons at home when schools were in lockdown. Just over half of the interviewed pupils reported that the science lessons learnt at during lockdown were interesting. Teacher interviews indicated that there was an issue with pupil disadvantage and accessibility to online learning environments. Some teachers reported that pupils from disadvantaged backgrounds were less likely to participate in science lessons or have resources at home to participate. There was also some variability about the type of teaching that went on and the pressures schools were under. Some teachers reported having to spend some time on non-teaching elements during lockdown, so for example putting time into ensuring pupils welfare issues were dealt with.

The wider evaluation process involved/affirmed that one class per school was identified as being the key class taking part in the programme, although some schools conveyed that further classes also received the programme. One key teacher was identified as the one delivering the TAPS lessons. As indicated in the logic model, pupils were provided with TAPS-based lessons. It was reasonable to conclude that the programme was broadly delivered with fidelity. Overall, it was reasonable to infer that Year 5 teachers understood and applied the programme with few adaptations, and they often conveyed that the programme was accessible to most pupils.

(a) In what ways do teachers and schools adapt the intervention and how do these adaptations affect implementation? (b) What adaptations have teachers had to make to implement Focus4TAPS during school

or class closures during the pandemic? Have the pandemic-related adaptations impacted implementation? How practical is it for schools to implement the Focus4TAPS programme designed by Bath Spa?

The various sources of data indicated that schools were able to adapt the intervention to fit in with their particular school contexts; the adaptations had no impact on implementation. Given that there is some degree of autonomy at primary level (particularly with senior leadership support) to move around lessons, the programme was able to fit in with the overall teaching plans of intervention teachers. For example, a Year 5 teacher within an intervention school conveyed:

> *We moved elements around in the timetable, so when I wanted to do certain key investigations, I would move blocks of learning and adapt it for home learning if it kind of lent itself, so that as soon as the children came back we prioritised all of the practical learning that I knew they wouldn't be accessing at home. So, it was very much moving around the timetable, and science has taken priority over some of the other … well, over the foundation subjects, so I haven't done as much history and geography as I normally maybe would have, because I've been trying to prioritise the science, although science is a core subject, so I always try to prioritise it a bit more than those subjects anyway to be honest.*

Other adaptations were about access to materials whether it was in lockdown or in school being able to access materials that were to hand. The resources from TAPS offered a range of activities for potential use within lessons, with a further degree of autonomy where changes could be made in order to run an experiment. For example, a Year 5 teacher within an intervention school conveyed:

> *I would say in school there was some experiments which if we were doing kind of week-on-week weren't feasible we've just kind of changed some of the resources needed. For example, we would just use salt and sugar as opposed to using a range of resources.*

Other adaptations were about enabling all pupils to access the learning goals, something which teachers already undertake within their day-to-day work. For example, a Year 5 teacher within an intervention school conveyed:

> *I wouldn't say in year five we had to make huge adaptations and it was more about differentiating it to allow all our children within our class to access that learning. But those differentiations take place kind of in every single lesson, regardless of what that kind of learning intention is. To make sure that all the children can access it. So, we would just make adaptations to the scaffolding of it and how we would scaffold each child to be able to reach those learning intentions.*

There was also support from the programme trainer to help schools switch back from online learning to school-based learning as the trainer took into account that there may be gaps in knowledge created by the lockdown. For example, a Year 5 teacher within an intervention school conveyed:

> *I think TAPS gave us a gap spotting document and when we returned from online learning we did use that sort of to help us with a couple of elicitation tasks to kind of assess where our children were and if there were any gaps that needed filling in from that time. So that kind of helped us with adapting lessons a little bit.*

The recommenced programme involved some online delivery to help accommodate potential disruption following from the COVID-19 pandemic, which was accessible and helpful for some teachers. For example, a Year 5 teacher within an intervention school conveyed:

> *The online courses I found really, really, really useful and it was nice to be able to speak to other kind of teachers which aren't in your school and see what they were doing, see how their children like it's gonna go, see how their children were getting on with the learning kind of discussed. Everyone is at different points at the start of the course, so we could kind of discuss 'I've done this investigation, this went really well, this didn't go as well'.*

However, some teachers conveyed that it was not always easy to deliver the programme during the pandemic despite the contingency plans put in place by the programme developer and schools. Teachers had to deal with a myriad of issues including some parents potentially not being able to support children in their learning, some children's lack of engagement, and wider issues around disadvantaged circumstances, which may have been impacting many aspects of children's lives including their school attendance (whether in-person or remote). This was particularly an issue in schools where there were high proportions of pupils from disadvantaged backgrounds. For example, a Year 5 teacher within an intervention school conveyed:

> *The fact that we were not only having to teach, we were trying to manage and control the children online and with many distractions … I found it really difficult. We didn't have many supportive role models at home for the children, which encouraged them to learn. And so that was really challenging, because you couldn't be there to support children. We have so many kind of individual needs and requirements, our school and which we weren't able to kind of implement. Some children which kind of had access kind of difficulties we couldn't differentiate as much as we could in school. We can offer interventions and then I think just the kind of the well being of staff as well online and made a massive impact on kind of quality of teaching and I felt it was OK at first but after a while teaching online to get quite draining staring at a screen all day, you've got technical issues. I found it by towards the end it was definitely difficult [teaching online] and I know a lot of teachers felt really similar. So, it was challenging. It was really challenging.*

Could teachers who attended training and delivered the Focus4TAPS programme continue to implement it without any further support or training? What do teachers think will be enablers and barriers to impact?

Data from teacher interviews and questionnaires indicated that many teachers felt that they were about to continue to implement TAPS without further training or support. Responses through the teacher questionnaire indicated that teachers, on average, agreed with being able to continue to implement the programme with no further training or support via 'I will be able to continue to implement Focus4TAPS without any further support or training' (an average of 3.43 on the 1–4 agreement scale from across 54 intervention teachers): 51.9% responded with 'Agree a lot', 42.6% 'Agree a little', 1.9% 'Disagree a little', and 3.7% 'Disagree a lot'.

On average, teachers also tended to agree (to a slightly lesser extent) with 'I will be able to continue to implement Focus4TAPS with further support or training' (an average of 3.19 on the 1–4 agreement scale from across 53 intervention teachers): 43.4% responded with 'Agree a lot', 37.7% 'Agree a little', 13.2% 'Disagree a little', and 5.7% 'Disagree a lot'.

The interviews revealed further elaboration and clarification from teachers. While some teachers talked about needing training for colleagues across the school, others talked about what they personally would need.

> *Because it's working for our school. The teachers are quite clear on it. We have our Working Scientifically walls on the classrooms of our displays. They understand what they need to be hitting. That's something that then gets passed on to the next year group teacher and the following year so they know where the children hit them, where they need to focus on the different skills in that particular year.*
>
> *I think any further kind of training would definitely be beneficial to keep and maintain it. I think it's quite disrupted last year [the training], so it would still be useful if the four TAPS [teachers] continue it. I think for it to kind of be [run] across the school. Other staff would therefore need training as well… for them to kind of access it and have the same knowledge and understanding of it. It would be really useful for them to go through a similar or if not the same training that we did. But yeah, I think I would be able to maintain it, but I think to develop it and kind of keep improving and promoting it. I would need further training*.
>
> *I think I can do it, I think I can manage it. I think it probably wouldn't be a bad idea to have a refresher course now and then. Just to [sic] you know, it's like anything to just kind of. Make sure you're, you know, bring you back to the right track and so on. But I feel pretty confident at the moment.*

How acceptable do teachers and school leaders feel it is for pupils to engage in the programme for the purpose of boosting science attainment?

Data from teacher interviews and teacher surveys indicated that there were varied responses about joining the TAPS programme for the purposes of enhancing pupils' science attainment. Some joined it to boost attainment, others for engagement.

> *Oh, absolutely [joined the programme for boosting attainment]. I was offered it, it came through on an email and I'd been trying to research ways to improve science assessment for years, and it finally came up with a very logical solution, so yeah, that's exactly why we went for it.*

Not all Focus4TAPS teachers made the decision to join the programme for boosting attainment, in some schools this was decided by school leaders.

> *Yes, the science lead chose, chose to join the program for science attainment.*

> *Yeah [joined the programme to boost science attainment], I think in general our school is really science based and really tries to implement science and attainment really well.*

Looking at the data across the teacher surveys and interviews it is clear that another key benefit of joining the TAPS programme is boosting teachers own confidence to teach science alongside boosting attainment in science.

> *Yeah yeah, definitely. Definitely boosting science attainment, boosting science teaching, and the confidence or for teachers to teach it. I think it would be wonderful if TAPS were able to provide in-school training—because I think you know, predominantly my other colleagues were quite sort of upset that here I was doing it and being able to access all of this, all of these resources and they weren't able to come—so that definitely would benefit teachers.*

Do teachers think there will be a sustained impact of the intervention on pupils' outcomes (in particular, attainment)? To what extent does this vary by pupil characteristic (e.g. FSM status, gender)?

Year 5 teachers within the intervention schools who completed the teacher questionnaire conveyed (through their free-text responses to 'What difference do you feel the Focus4TAPS programme has made for students?') that they believed that the programme benefited pupils through:

- the provision of more practical work (and less writing);
- practical/enquiry skills;
- confidence in general and confidence in undertaking investigations;
- engagement, excitement, and interest;
- independence, autonomy, and ownership;
- more discussions about science and opportunities to learn new science vocabulary; and
- more experiences including generally fostering being/thinking like scientists.

For example, Year 5 teachers within intervention schools conveyed:

> *The children have been more focused each lesson, and they clearly understand what they need to do to be successful each lesson.*

> *They definitely are more engaged and enjoy science more because they can have more ownership and the freedom to explore and investigate for themselves.*

> *They are more confident in practical aspects. Their enthusiasm and intrigue for science is amazing. So many positive emails and letters from parents about how their children have loved the science work.*

> *They have a greater understanding of how scientists work and behave and the different skills involved in doing this. More practical work = greater engagement and enjoyment.*

> *Children feel that they are able to ask a question and seek out the answer independently. They have learnt to accept that some plans will fail and have developed a resilience in seeking a new method.*

> *The children have really enjoyed engaging in the TAPS lessons and have been really enthusiastic towards the investigations. They seem to enjoy the lessons more and look forward to science each week. They have enjoyed using a wider range of scientific equipment and can use some of these with greater accuracy.*

The teacher interviews also suggested that teachers were either positive about a current change in pupils' attainment, or a change in their skills, which will enable pupils to engage with and understand science better.

For example, Year 5 teachers within intervention schools conveyed:

> *In terms of their inquiry skills, I'd say that across the school has gotten better.*

> *I think it would definitely have a positive impact on attainment within science.*

> *I think it will definitely if they… if we continue to use the scheme of work and the investigation, ideas and the focused teaching and recording. I think it will definitely have a positive impact. It's definitely become clearer what how each child is doing in science. I think it's developing their Working Scientifically skills. So it's just more obvious where each child is in their attainment and how they can move on.*

While most teachers reported that there was no impact by pupil characteristics (on attainment), one school reported that the TAPS programme was less accessible to their pupils with EAL. Some of the Year 5 teachers within the intervention schools who completed the teacher questionnaire conveyed that they believed that the Focus4TAPS programme was accessible for most pupils and especially those who might have found routine teaching/learning less accessible. For example, one teacher conveyed that they believed the programme '*worked well for children with literacy difficulties as [the] focus on speaking and listening and allowed group recording therefore less writing',* and the teacher was then '*able to assess their science skills and knowledge as opposed to their writing skill'.* One teacher conveyed that programme worked especially well for *'SEN, because of the reduced amount of written work and the practical element made it a level playing field Low floor high ceiling activity'*; another teacher similarly conveyed that *'The SEN seemed to really engage though'*; and another teacher conveyed that *'It worked well for all students but especially LA [lower ability] and SEN children who find it difficult to access text heavy lessons'.* Nevertheless, another teacher conveyed that the programme *'worked well for all groups. Perhaps less so for SEN children as they are a lot less independent than the rest of the class'*.

How (if at all) do teachers think the intervention will impact on pupils in other ways (e.g. interest in other subjects)?

Year 5 teachers within the intervention schools who completed the teacher questionnaire often conveyed (through their free-text responses) that they believed that the programme benefited pupils' interest and enjoyment, recognising that pupils often appreciate and enjoy practical work. For example, Year 5 teachers within intervention schools conveyed perceived benefits to pupils including:

> *More practical work which they love!*

> *More excitement in the learning.*

> *Children are more engaged and willing to partake in science lessons with more enjoyment. They still find written science challenging.*

> *They have enjoyed the practical focus with a lesser emphasis on writing.*

> *They have really enjoyed the activities and not having to write out whole investigations.*

The pupil and teacher interviews revealed that while there was no sustained impact on other subjects the TAPS programme helped to increase pupils' engagement in science, pupils' intrinsic motivation, and for some attainment.

> *I think they are definitely, when … if you write, 'Science', on the board you can hear them, you can see the excitement, they're definitely more motivated and engaged in science than they used to be, and I know one of them was saying to me the other day, 'Oh, my previous teachers, they didn't like science, they liked RE [Religious Education]. We love it now'. I don't know whether that's because I'm enthusiastic about it or whether it's because I use the TAPS as well, but I'll never know, because I do.*

How scalable do teachers think the Focus4TAPS programme is and what are their suggestions for change if the programme is to be implemented more widely?

Responses through the teacher questionnaire and interviews indicated that many teachers believed that the Focus4TAPS programme was scalable. In some of the schools they already began rolling out TAPS to other parts of the school.

> *We've run it, we put it across the whole school apart from reception, we've taken aspects of advice from the TAPS sessions and given it to reception to use but not fully implemented because they've got their own curriculum going on.*

Some teachers felt that scaling the programme across the country would be possible.

> *It's good, just it. It's clear it's straightforward […] Because one of the main reasons why we decided to go with TAPS at our schools that we were struggling with a kind of focus—we notice that our while our knowledge was good, the actual record scientifically end was not very good at all. So when we had the opportunity to be part of the TAPS project. We just jumped on that because it just seemed to be enough right now.*

Some teachers felt changes were needed to make TAPS accessible to a wide range of children. One school had a higher-than-average number of EAL children and in this particular school, teachers reported that EAL children were less able to access the TAPS lessons than children who were of high or average ability. To make TAPS scalable the teachers suggested making resources more applicable to EAL children.

> *I think maybe looking at some of the resources you have on offer and how you can differentiate and scaffold those resources for a much wider scale of the children. I know a lot of the people on the course said it was really accessible for all their children … very different demographic [his school with high proportion of EAL students compared to other schools]. So it's looking at schools similar to ours and how the TAPS program can really relate to schools like ours. Maybe that be what they do differently to teach for TAPS or allow children to access for TAPS learning. So maybe look at how you can differentiate those even further.*

In terms of making the programme more scalable (within the school), the impact on teachers own time was important particularly in the light of the COVID-19 pandemic. To make this programme work well, what is required is an internal support network where teachers can support each other to learn and implement TAPS, although of course, this brings across certain burdens for individual teachers. It certainly was the case that the science lead in a number of schools had to hold together the TAPS programme particularly when there was a change in staff because of maternity or sickness.

> *I'd be hoping now that you know things are kind of beginning to become more normal in schools because we have had a lot of competing demands on our time as teachers. But I would like to see a trickle-down effect of what we're doing in years five and six to kind of get that approach going in years three and four and possibly years one and two. I do accept that those teachers haven't been on the training and so they would be relying on me …*

There were a number of suggestions to help make the TAPS programme more scalable across the country. Some teachers reported that online training helped them to access the professional development training they needed.

> *I think actually doing it online this way has been great. It's made it a lot easier for lots of teachers. You have made it a lot easier.*

What impact does Focus4TAPS have on pupils' attitudes towards science?

The impact evaluation indicated via the pupil survey that there was no difference in pupil attitudes to science in the Focus4TAPS schools compared to the control schools. However, responses through teacher interviews and pupil interviews indicated that pupils in Focus4TAPS schools were felt to be more positive about science after having participated in Focus4TAPS for an academic year. Teachers believed that pupils' engagement had increased.

> *I'd say maybe an increased engagement in science—like children look forward to it more in my class.*

Teachers also reported that pupils who may not have ordinarily engaged with science were more able to do so. One teacher gave an example of girls, where the Focus4TAPS method of teaching had helped girls to engage with discussions and using oral communication skills.

> *The girls or some a lot of the time I've railroaded by the boys in terms of being able to speak up and have their own say. And this is especially the collaborative nature of it and working in groups and things like that. I think the girls are now forced into speaking them. And when I say forced, it's good because I think a lot of them are not used to not speaking, so forcing them to speak has allowed them to actually find their voice a little bit more than previous, I would say.*

On the whole, teachers reported that pupils already had intrinsic motivation in science, which was further increased as a result of the Focus4TAPS way of teaching.

> *I think in science it [Focus4TAP teaching] definitely motivated them and they looked forward to science a lot more and their engagement with science lesson definitely had a massive increase.*

Another teacher highlighted that engagement and benefits did not necessarily link with simple notions of ability.

> *And there was no sort of hard and fast rule [who was more or less engaged with TAPS] … What we found is we had actually some children who would have been classed as less able in maybe English or maths, but actually when it came to actually carrying out something practically and setting up an investigation or setting up equipment and actually came into their own. It was surprising to see which children actually understood what was going on, and sometimes the more academic ones—who were more keen on, you know, doing what they're told and not really thinking outside the box—they were the ones that sometimes just needed that extra support … it wasn't always the children that you expected [that needed the support].*

What are the views of the intervention pupils about the teaching and learning methods applied by Focus4TAPS, and do such views vary by pupil characteristics (e.g. FSM status, gender)?

The pupil interviews themselves did not indicate that there was variation in experience received of the TAPS programme by pupil background although some data from teachers indicated that the TAPS programme may have been less accessible to pupils from backgrounds with EAL. One teacher spoke about having to provide scaffolding around the resources to make them accessible although the same teacher also indicated that the pupils' level of English understanding was particularly below average.

Pupil interviews in intervention schools demonstrated that pupils noticed that there were more investigations in Year 5 than in Year 4. The reports from pupils about the teaching and learning methods were quite positive, where pupils spoke about Working Scientifically, having more time for discussion and participating in investigations. For example, a pupil conveyed: *'I think work in Year 5 has been so much more better (compared to Year 4).'*

Some pupils were able to talk about science being important to them and being happier with the lesson in Year 5 because of the more inquiry-based science. For example, a pupil conveyed:

> *Yes [science Year 5 better than Year 4] because in Year 4 we didn't really do a lot because we were in lockdown and it's just really, I feel like I've really improved since Year 4. Yes [science is important], because I really want to be either a vet or a doctor because, and those [Year 5 lessons] involve science and I really like science.*

While there was lockdown learning in both Years 4 and 5, pupils were more positive about the learning in Year 5. For example, a pupil conveyed:

> *I remember in Year 4 we did an egg experiment and that was very good, we got to watch how they coped under different circumstances in vinegar and water but that was really the only, I would say the only fun part that I really enjoyed because we had to go in to lockdown pretty early and we didn't even do online learning, we just sat at home and did like homework for a few weeks—but in Year 5 it's always been constant and there's always been excitement and every week I always look forward to science.*

Does Focus4TAPS have an impact on teachers' perceptions of their job satisfaction and workload?

A large proportion of teachers in the surveys and interviews reported feeling positive about the impact of Focus4TAPS on their workload and job satisfaction. Positive impact on job satisfaction was because of a number of things. So, for example one teacher spoke about their view that being involved in Focus4TAPS felt like being involved in research:

> *I think I feel positive about it, I like being involved in research, I've done research projects before for TAPS, and that's an element that I'm really interested in. I think research is a really positive thing to be involved in, and I think it kind of builds your confidence as a teacher and it means that yeah, I can go in and help other teachers, and I can give you some resources, and I feel confident in what I'm telling them because I've seen the impact of it in my own class*.

Asking teachers about job satisfaction appeared to be closely linked to learning about and using the TAPS investigations. For example, one teacher within an intervention school conveyed:

> *I found some of the teaching more fun … was probably based on the fact that we would do more investigations and that's something I enjoyed.*

However, some teachers spoke about the challenges they faced because of school closures impacting learning. For example, one teacher within an intervention school conveyed:

> *When the children were in school we had to pick them up from, well, literally. You know they when they came back they were all really at different places in their learning and then they would have been if they had been in school and we felt that some of the learning that they've done in previous years before the lock downs that had been lost to a certain extent. Children's approach to learning changed and they find it much harder to settle and to do things in a in a in a [sic] set time so things seem to take longer for children to do some things I'm beginning, I think to move back up, but it's been hard work. You feel as though you're, you're just kind of constantly trying to retrieve to get those children back up to where they should be … it has felt challenging.*

Some teachers were able to report via the teacher survey and interviews that despite working on the Focus4TAPS programme in the pandemic there were positives that in turn impacted teachers own feelings of job satisfaction. For example, one teacher within an intervention school conveyed:

> *I mean it's come at a very difficult year where I suppose job satisfaction is difficult due to how I guess how stressed all teachers are, however, I can honestly say that my science sessions and some of the brilliant lessons that we've done have been the lessons where I've smiled the most and laughed the most. It's some of those lessons where I've really enjoyed spending time with my children […] Enjoyed being their teacher. So, it definitely, I guess made me realize how nice the job can be and gave us a lift when there's so much other rubbish going on. So, definitely made me appreciate how lovely the job can be. Thank you very much.*

The TAPS programme also had some impact on teachers' confidence and intrinsic value of teaching. For example, one teacher conveyed that: *'It definitely has given me more confidence as a teacher and my enjoyment of teaching science has increased*.' Nevertheless, some variation was present and another teacher noted: *'Probably not, because the job is bigger than science.'*

Teachers understanding of learning objectives and criteria for success had an impact on teachers feeling more positive about teaching science and thus having an effect on feelings of job satisfaction. For example, one teacher within an intervention school conveyed:

> *I think because it's helped … towards science lessons. That's more positive. I think it's had a positive impact. It's probably not been an enormous impact overall throughout the whole job … but it's definitely helped a bit, yes. OK, so it's helped a bit with job satisfaction.*

The teacher survey indicates that the TAPS programme did have an impact on pedagogical planning to elicit pupils' science knowledge and skills. The interviews suggested that achieving this helped teachers feel more professionally competent, which in turn had an impact on job satisfaction. For example, teachers within intervention schools conveyed:

> *I feel much stronger [as] a science coordinator. And a better professional teaching science myself. So I've felt more secure.[…] And it's been really sweet—like through the PSQM [Primary Science Quality Mark] I had to do teacher feedback and I had tremendous positive feedback from my colleagues who have appreciated the insight of the training that I provided—as I said, copying and pinching all the good ideas from TAPS and bringing it to my colleagues and they're really open to all. Like I said, we are doing now whole science weeks or science days on top of our lessons where it's all investigations and it's all using mostly the TAPS material and from reception all the way to year six TAPS has been used. Teachers said they are more confident in teaching science, and when I did a review, you know I can see the science has changed at school, as has become a lot more hands on practical and the recording of it.*

Usual practice

The evaluation considered teachers' reports of their school context and their reported teaching/learning practices, in order to determine what could be inferred to be 'usual practice' or 'business as usual' within control schools, in order to help determine whether and how the intervention programme differed.

On average, Year 5 teachers within the control schools and within the intervention schools gave similar responses through the teacher questionnaire regarding their schools being involved or not involved in science programmes, initiatives, or other interventions during the past two academic years (with questions considering PSQM, PSTT, STEM Learning/Science Learning Centre courses, Association of Science Education [ASE] courses, and with an option to note other initiatives). On average, the teachers also gave similar responses for their schools having resources available or not available for Year 5 pupils (with questions considering computers/tablets, a library, a science laboratory, assistance during practical work, and specialised science equipment for practical work).

Table 36: Year 5 teacher reported contexts

	Intervention group			Control group			Observed difference	
Questionnaire item (scale)	Mean	SD	N	Mean	SD	N	D	p-value
Has your school been involved in any specific science programmes, initiatives, or other interventions during the past two academic years? Primary Science Quality Mark (0=No, 1=Yes)	.05	.23	56	.02	.13	55	.189	.320
Has your school been involved in any specific science programmes, initiatives, or	.02	.13	56	.02	.13	55	.002	.990

other interventions during the past two academic years? Primary Science Teaching Trust (0=No, 1=Yes)								
Has your school been involved in any specific science programmes, initiatives, or other interventions during the past two academic years? STEM Learning / Science Learning Centre courses (0=No, 1=Yes)	.23	.43	56	.25	.44	55	.052	.786
Has your school been involved in any specific science programmes, initiatives, or other interventions during the past two academic years? Association of Science Education (ASE) courses (0=No, 1=Yes)	.02	.13	56	.00	.00	55	.188	.322
Has your school been involved in any specific science programmes, initiatives, or other interventions during the past two academic years? Other (0=No, 1=Yes)	.07	.26	56	.16	.37	55	.287	.135
Does your school have computers and/or tablets that can be used by Year 5 students? (0=No, 1=Yes)	.96	.19	56	1.00	.00	53	.266	.159
Does your school have a library that can be used by Year 5 students? (0=No, 1=Yes)	.91	.29	55	.88	.32	52	.080	.681
Does your school have a science laboratory that can be used by Year 5 students? (0=No, 1=Yes)	.02	.13	56	.06	.23	53	.205	.294
Do teachers usually have assistance available when Year 5 students are undertaking practical work? (0=No, 1=Yes)	.54	.50	56	.43	.50	53	.203	.292
Does your school have specialised science equipment for practical work? (0=No, 1=Yes)	.74	.44	54	.60	.50	52	.308	.117

SD, standard deviation; STEM, science, technology, engineering, and mathematics.

Essentially, it was possible to infer that the intervention and control schools appeared to have broadly similar contexts/circumstances, excepting the Focus4TAPS programme itself (and the slightly higher number of science specialists reported by those teachers in the intervention group who completed survey returns, detailed in the earlier 'Pupil and school characteristics' section describing the samples; Table 12), although it remains possible that unmeasured differences were present.

There were similarities across control schools and intervention schools in terms of some reported teaching practices. Year 5 teachers within the control schools (and also those within the intervention schools) who completed the teacher questionnaire conveyed (through their free-text responses to 'Please briefly describe how you teach Working Scientifically to your Year 5 students') that they applied a range of approaches and activities to teach Working Scientifically with Year 5 pupils. These responses conveyed a range of practices including:

- pupils being able to plan their own scientific investigations in order to answer their own questions (including making predictions), with attention towards identifying independent and dependent variables and how to control them as necessary (with teachers often noting the concept of fair testing);
- pupils taking measurements with a range of equipment while working safely; and pupils recording and conveying results through various tables and charts (and as a scientific report with introduction, prediction, methods, results, and conclusion sections); and
- pupils drawing conclusions with explanations and using results to make further predictions. Some teachers noted that this broadly formed a 'science enquiry cycle'.

Teachers conveyed that investigations might be undertaken in pairs, groups, and as a class. For example, Year 5 teachers within the control group wrote via the questionnaire:

> *I use an investigation planner template for children to independently plan scientific enquiries focusing on identifying independent/dependent variables, controlled variables and making sound predictions. I then allow children to carry out their investigations using measuring equipment to record results and present their findings using tables, graphs, bar charts and so on.*

> *Children plan their own experiments in each unit of work. We begin with an enquiry question and the children plan what equipment they will need to plan an effective experiment. They make predictions, set up their tests, observe and measure their own data and evaluate it using the information they have gathered from their assessment. Depending on resourcing, this may be as a whole class together, or in pairs.*

Year 5 teachers within the control schools (and also those within the intervention schools) also noted that teaching Working Scientifically involved discussions between pupils and teachers, including thinking about prior strategies that were applied and learning that was undertaken, and also making connections with the wider world. The teachers often conveyed that a varied range of practical work was used to cover Working Scientifically, but where exploring the science behind investigation could also be covered and supplemented through other approaches such as class discussions, videos, examples, and teacher demonstrations. Some teachers noted that each session was intended to have a practical aspect and a Working Scientifically focus; others noted that investigations covered a series of lessons to allow planning, undertaking the work, and analysing the results. For example, Year 5 teachers within the control group wrote:

> *We use the strands from the curriculum and build specific areas into our teaching so that each session has a 'Working Scientifically' focus. Children also carry out in-depth investigations. We use 'Big Questions' and apply scientific skills such as observing and pattern spotting in other subject areas.*

> *By following a scheme, every lesson has the opportunity for the children to practice their Working Scientifically skills.*

> *I try to teach this strand through as many different avenues/skills during the school year to ensure that children have access to and the opportunity to develop all the strands.*

Year 5 teachers within the control schools (and also those within the intervention schools) who completed the teacher questionnaire also conveyed (through their responses to 'Please briefly describe how you undertake formative assessment in science for your Year 5 students') that they applied a range of approaches and activities within assessment for learning. Responses included:

- initial consideration or assessments of understanding;
- observations and questions/discussions with pupils and classes during learning;
- pupil self-assessments, paired work, and peer feedback during learning; and
- subsequent assessments and notes on assessment grids, and with formal systems to record notes and judgements; and planning and delivering subsequent follow-up lessons to address misconceptions.

Responses from Year 5 teachers within the control schools and the intervention schools broadly conveyed a relatively similar range of activities and approaches, including initially and subsequently using 'what I know, what I want to know, what I've learnt' grids or other forms of initial elicitation and subsequent review that then informs follow-up lessons. For example, Year 5 teachers within the control group conveyed via the questionnaire:

> *This is done in the classroom during the lesson through paired work, higher-order questioning, peer feedback and then after the lesson through marking. This is often followed up in the following lesson with any misconceptions addressed.*

> *Children take part in a pre-assessment test for each topic which informs group and planning going forward. At the end of the unit, children take a unit test. During the unit, children have the expectations at the back of the book which are ticked when lessons are completed to show the children's understanding and then ticked again if they show understanding on the end of unit test.*
>
> *We use retrieval quizzes at the start of every lesson. Ask questions which link to prior learning throughout the lesson. The children complete a pre and post assessment for every topic.*

Overall, responses from Year 5 teachers from the control schools and from the intervention schools conveyed a relatively similar range of activities and approaches that can be inferred to elements of 'usual practices' within teaching/learning science. Nevertheless, a broad and diverse range of approaches were conveyed, and it remains difficult to determine whether every school would necessarily apply every approach or apply an approach in the same way. Applied practices may also differ from reported practices. From a wider perspective, the range and detail within the responses from teachers potentially coheres with one of the wider ideas within the Focus4TAPS intervention conceptualisation: the conceptualisation/specification of the Focus4TAPS programme essentially involves adapting and refining existing teaching/learning practices; the programme was not conceptualised around the assumption of a 'deficit model' (where teachers are assumed to not undertake group work or formative assessment or suchlike). While these are teacher self-report of practice, the quantitative findings from the teacher survey do add weight to the assumptions in the logic model that intervention teachers were more likely to use the programmes' intended teaching and learning practices (see teacher survey findings Table 40).

Additionally, the COVID-19 pandemic may have impacted general and science-specific teaching/learning approaches and circumstances within schools although as Table 37, Table 39, and Table 40 indicate pupils and teachers in intervention schools were more likely than control school pupils and teachers to say that the quantity and quality of science teaching was different. It is unavoidably difficult to characterise 'usual practices' within potentially changed and changing circumstances. Teacher responses through the questionnaire item that considered involvement in any other initiatives did not clearly involve any related to COVID-19 recovery. The interview with the Focus4TAPS trainer indicated that advice and support was given to teachers to deliver the teaching remotely.

Further responses from the teachers who completed questionnaires highlighted similar points around disruption during lockdown/home learning from those within control and within intervention schools (via responses to 'This academic year (2020/2021), how has the COVID-19 pandemic impacted teaching and learning for Year 5 students?'). Teachers highlighted impacts including:

- limited coverage of science content and undertaking practical work (introducing or risking gaps in knowledge on return, and additionally complicated with recovery/catch-up foci towards English reading/writing and mathematics);
- children may not have completed (or been able to complete) assigned remote work;
- limited scope of or delays to assessment and feedback; and
- limited group working, discussion, and potential for support.

Some teachers from intervention schools and control schools cited use of Oak Academy resources for home/remote learning and some additional science-related activities from various organisations such as the ASE. Some teachers conveyed that Year 5 pupils ultimately had greater knowledge gaps and misconceptions than would usually be observed.

To what extent can the intervention (and its different elements) be distinguished from existing practice and what factors affect the degree of differentiation from normal practice? To what extent is Focus4TAPS distinguishable from control schools' model of learning?

Table 37, Table 39, and Table 40 indicate there were some significant findings from the teacher and pupil surveys to demonstrate that there were elements of the intervention, which were different from existing practice namely:

- intervention pupils were doing more science;
- intervention pupils were doing more practical work;
- intervention pupils reported that in a typical week, they spent more time learning science;

- intervention teachers were more likely to inspire students to learn science;
- intervention teachers were more likely to explain science concepts or principles by doing science experiments;
- intervention teachers were more likely to help students appreciate the value of learning science;
- intervention teachers were more likely to assess student comprehension of science;
- intervention teachers were more likely to teach science using inquiry methods;
- intervention teachers were more likely to provide challenging tasks for the highest achieving students; and
- intervention teachers were more likely to ask students to explain their answers.

Year 5 teachers within the intervention schools who completed the teacher questionnaire conveyed that some particular practices were also applied within Working Scientifically and assessment for learning (in addition to the range of practices as above), which were not present within responses from teachers within the control group. First, some teachers within intervention schools explicitly cited their use of TAPS resources, such as a Working Scientifically wheel; second, some teachers within intervention schools explicitly cited their use of Focus4TAPS principles, such as specifically focusing on one aspect of Working Scientifically at a time. When teachers were asked within the teacher survey (‘In what ways have you applied the programme in your teaching and across your school?’), Year 5 teachers within the intervention schools who completed the teacher questionnaire similarly affirmed/conveyed that they applied the Focus4TAPS programme through various ways including: applying more practical work around the TAPS ideals; and through a refined/adapted approach to planning, delivery, and assessment around the TAPS ideals such as focusing on one aspect of Working Scientifically at a time. Nevertheless, some teaching/learning approaches such as focusing on one aspect of Working Scientifically at a time are not necessarily unique to Focus4TAPS and could be discovered and adopted in schools outside of Focus4TAPS (although the responses from teachers from control schools did not cite these particular examples). Future research may benefit from isolating practices or resources that are entirely unique to an intervention programme (impossible or unfeasible for control schools to apply), and practices that schools could conceivably adopt or apply outside of an intervention programme.

Pupil reports about usual practice

Pupils within intervention schools, on average, reported that they experienced more frequent science work and investigations (practical work) in science compared to pupils within control schools, on questionnaire scales of: (1) ‘Never or almost never’; (2) ‘Once or twice a month’; (3) ‘Once or twice a week’; and (4) ‘Every day or almost every day’. The survey data was collected in summer 2021 when pupils had almost a full academic year of teaching. This suggests that the Focus4TAPS programme may have increased the provision of science including practical work; alternately, this may reflect greater focus towards science or greater visibility of science from the perspective of pupils, which may or may not follow from changes in the actual frequencies of provision. The pupils completed the questionnaires in June 2021 and July 2021, and it cannot be determined whether their responses reflect the (potentially varying) frequency of science teaching/learning across the academic year.

Table 37: Pupil questionnaire responses for science teaching/learning frequency

	Intervention group			Control group			Observed difference	
Questionnaire item (1–4 frequency scales)	Mean	SD	N	Mean	SD	N	D	p-value
How often do you do science at school?	2.81	.49	1,252	2.74	.56	1,227	.138	.001
How often do you do investigations (practical work) in science?	2.27	.66	1,226	2.17	.65	1,199	.151	<.001

SD, standard deviation.

The pupils’ responses to other questionnaire items that were intended to consider aspects of the wider TAPS approach (such as ‘My teacher plans and discusses science lessons with us’ and ‘I check my own work to find out what I have learned in science’) involved no clear differences across the intervention and control groups. It is possible that these aspects may not necessarily have been central to the delivery of the Focus4TAPS programme compared to other aspects. Other elements of the IPE (as above) suggest that teachers within the intervention schools and the control school broadly apply a range of teaching/learning approaches, including discussion and elicitation (i.e. could be inferred to be aspects of ‘usual practice’ within science teaching/learning).

Table 38: Pupil questionnaire responses for IPE items

	Intervention group			Control group			Observed difference	
Questionnaire item (1–4 agreement scales)	Mean	SD	N	Mean	SD	N	D	p-value
My teacher plans and discusses science lessons with us	3.28	.92	1,240	3.25	.91	1,204	.033	.421
I check my own work to find out what I have learned in science	2.88	.97	1,241	2.90	.96	1,197	.015	.702
Students check each other’s work to find out what we have learned in science	2.27	1.06	1,236	2.34	1.03	1,203	.063	.117
My teacher explains how science is relevant to everyday life	3.02	.94	1,238	3.07	.95	1,199	.048	.233

IPE, implementation and process evaluation; SD, standard deviation.

Teacher reports about usual practice

Year 5 teachers within the intervention schools reported that Year 5 pupils spent more time learning science, on average, compared to those in control schools.

Table 39: Teacher questionnaire responses (Year 5 teachers) for science teaching/learning time

	Intervention group			Control group			Observed difference	
Questionnaire item (scale)	Mean	SD	N	Mean	SD	N	D	p-value
In a typical week, how much time do Year 5 students spend learning science? (minutes, coded from free-text response)	103.09	33.05	55	89.06	37.18	48	.400	.047
In a typical week, how much time do you spend teaching science to Year 5 students? (minutes, coded from free-text response)	100.75	34.35	53	87.10	41.49	50	.359	.073

SD, standard deviation.

The Focus4TAPS programme and logic model intended to (essentially) refine science teaching/learning rather than increase the extent of teaching/learning. It is possible that the Focus4TAPS programme may have indirectly inspired more science teaching/learning and directly facilitated more science teaching/learning through aspects of the training (such as guidance on how to undertake science activities during the pandemic). However, given that baseline data was unable to be collected, it remains unclear whether these findings around teaching time reflect a baseline difference and/or an increase (there may have been a baseline difference that remained constant, or an increase over time, or a baseline difference and also an increase over time).

The degree of differentiation between intervention and control is impacted by a number of things that brought about changes via the TAPS programme. So, for example (as detailed below), Focus4TAPS teachers reported that they were more likely to inspire pupils to learn science, they were more likely to explain science concepts or principles by doing science experiments. In addition, Focus4TAPS teachers provided challenging tasks for the highest achieving pupils. Teachers in intervention schools were also more likely to report that they helped pupils appreciate the value of learning science. In terms of inquiry the intervention teachers were more likely to use this to teach science and assess pupils understanding of science than those from control schools.

Teacher interviews in intervention schools spoke about how engaging in Focus4TAPS has helped improve and change their practice, which were affirmed by some observed differences between control and intervention schools in responses to the teacher questionnaire. The teacher questionnaire gathered information around the teachers' reported frequencies of applying various activities and aspects of teaching/learning including: relating lessons to daily lives; asking students to explain answers; bringing interesting materials to class; asking students to complete challenging tasks; and taking part in discussions. These questionnaire items were informed by earlier research studies in order to enhance potential contextualisation and comparability (Martin, *et al.*, 2016b), and so encompassed practices that were intended to be changed by Focus4TAPS and also practices that might not have been intended to be changed by Focus4TAPS.

Year 5 teachers within the intervention schools (on average compared to those in control schools) reported that they more frequently asked pupils to explain their answers in class (which cohered with principles of Focus4TAPS) and were more likely to bring interesting materials to class (which was not necessarily an aim of Focus4TAPS). Otherwise, the teachers reported similarly for a range of diverse aspects of teaching/learning. Some of these aspects (marked with an *) were not necessarily target foci within the Focus4TAPS programme (such as 'Relate the lesson to students' daily lives'), which indirectly enhances the plausibility and validity of the various findings from the teacher questionnaire.

Table 40: Teacher questionnaire responses (Year 5 teachers) for frequencies of teaching

How often do you do the following in teaching science to Year 5 students? (1=Never, 2=Some lessons, 3=About half the lessons, 4=Every or almost ever lesson)	Intervention group			Control group			Observed difference	
Questionnaire item (1–4)	Mean	SD	N	Mean	SD	N	D	p-value
Relate the lesson to students' daily lives*	3.17	.86	54	3.28	.82	53	.138	.476
Ask students to explain their answers	3.95	.23	55	3.74	.52	53	.521	.009
Bring interesting materials to class*	2.95	.83	55	2.60	.77	53	.428	.028
Ask students to complete challenging exercises that require them to go beyond the instruction*	2.82	.75	55	2.68	.67	53	.195	.312
Encourage classroom discussions among students	3.93	.33	55	3.79	.57	53	.293	.135
Link new content to students' prior knowledge	3.81	.39	54	3.81	.48	53	.008	.967
Ask students to decide their own problem-solving procedures	2.89	.69	55	2.83	.75	53	.084	.662
Encourage students to express their ideas in class	3.91	.35	55	3.84	.42	51	.172	.382

* Item not necessarily target foci within the Focus4TAPS programme.
Focus4TPS, Focus for Teacher Assessment of Primary Science; SD, standard deviation.

Changes in practice around working and assessing scientifically

Summary of findings to support better teaching informed learning

Findings indicated that as hypothesised in the logic model the intervention led to teachers using assessment for learning informed by Focus4TAPS, which then had an impact on students' learning. So, for example, the statistical findings in the impact evaluation demonstrated there was an impact on student learning as pupils in intervention schools had higher attainment results. Table 37 indicates that intervention pupils reported in the student survey that they were doing more science than control schools and that they were doing more practical work than control schools. Teachers in interventions school were more likely to report that they were inspiring students to learn science, explain science concepts or principles by doing science experiments, helping students appreciate the value of learning science, assessing student comprehension of science, teaching science using inquiry methods, and providing challenging tasks for the highest achieving students.

Further significant findings indicated that intervention pupils reported that in a typical week, they spent more time learning science (Table 37) and that Table 40 reports from the teacher survey that teachers in intervention schools were more likely to ask students to explain their answers.

Given that the trial was restarted in the middle of the pandemic, it was jointly decided between the EEF and the evaluators that baseline IPE data would not be collected in order to minimise burden to schools. Given the lack of data the evaluation was unable to ascertain how practice may have changed between the end of the treatment and baseline for both the control and intervention schools.

Teacher reflections on the year

Year 5 teachers within the intervention schools who completed the teacher questionnaire tended to agree that the Focus4TAPS programme entailed various positive benefits, which help affirm the other findings around changes/differences/refinements in practices around Working Scientifically and assessment and (from a wider perspective) therefore also help affirm aspects of the logic model.

Table 41: Teacher questionnaire responses (Year 5 teachers) for reflections on the year

Reflecting on your experience of teaching science to Year 5 students this year, how much do you agree or disagree with the following statements?	Intervention group			Control group			Observed difference	
Questionnaire item (1–4 scale; 1=Disagree a lot, 2=Disagree a little, 3=Agree a little, 4=Agree a lot)	Mean	SD	N	Mean	SD	N	D	p-value
My students have enjoyed their science lessons	3.70	.46	53	3.57	.54	53	.263	.178
My students have made good progress in science	3.23	.58	53	3.04	.63	52	.313	.112
My students are confident in science	3.23	.54	53	2.96	.59	53	.467	.018
My students can work independently in science	3.02	.64	53	2.77	.67	52	.381	.054
My students come up with their own scientific ideas	3.09	.56	53	2.67	.79	52	.617	.002
My students do a lot of writing in science	2.23	.75	53	2.56	.85	52	.414	.037
I have changed the way I teach science	3.43	.57	53	2.74	.76	53	1.035	<.001
I enjoy teaching science	3.70	.46	53	3.42	.72	53	.468	.018

I have a good knowledge of science	3.52	.50	52	3.26	.59	53	.463	.019
My students have been engaged with science	3.62	.49	53	3.40	.60	53	.414	.036
I have explained to my students how science is relevant to everyday life	3.64	.59	53	3.45	.50	53	.344	.080
I have talked to my students about jobs that involve using science	3.43	.67	53	3.28	.77	53	.210	.282
I have conveyed to my students that anyone can do science and be a scientist	3.67	.51	52	3.56	.54	52	.219	.266
I have conveyed to my students that it is important to do well in science	3.38	.60	53	3.35	.59	52	.053	.788
I have conveyed to my students that it is important to learn about science to get ahead in the world	3.36	.68	53	3.26	.62	53	.144	.459

SD, standard deviation.

The average responses from the questionnaire (Table 42 below) affirm the logic model, where teachers conveyed that their experiences of the programme changed/improved their practices and entailed other benefits. The data also demonstrated that just as the logic model suggested, improved teachers' practices fed into teachers' experiences with Focus4TAPS and that there was a circular relationship between the two. So, for example, intervention teachers were very positive about the training preparing them for using TAPS in the classroom (with an average extent of agreement of 3.60 on the 1–4 agreement scale) and excited about using TAPS strategies in their teaching (with an average extent of agreement of 3.59 on the 1–4 agreement scale).

Teachers tended to report positively around continuing to implement Focus4TAPS, with slightly more positive response to 'I will be able to continue to implement Focus4TAPS without any further support or training' than to 'I will be able to continue to implement Focus4TAPS with further support or training'. Further evaluations may benefit from applying explicit questions around the extent and nature of any further support or training that may be beneficial for sustained implementation.

Table 42: Teacher questionnaire responses (Year 5 teachers) for the Focus4TAPS programme

Reflecting on your experience of applying the Focus4TAPS programme this year in 2020/2021, how much do you agree or disagree with the following statements?	Intervention group		
Questionnaire item (1–4 scale; 1=Disagree a lot, 2=Disagree a little, 3=Agree a little, 4=Agree a lot)	Mean	SD	N
The training prepared me for using Focus4TAPS in my classroom	3.60	.63	52
The resources increased my understanding of the Focus4TAPS ideas around Working Scientifically	3.69	.61	54
The resources increased my understanding of the Focus4TAPS ideas around formative assessment	3.59	.60	54
Taking part in Focus4TAPS has increased my job satisfaction	3.09	.73	54
Taking part in Focus4TAPS has increased my workload	2.35	.78	54
The Focus4TAPS strategies have been useful to me for teaching science	3.65	.59	54

I have been excited to use Focus4TAPS strategies in my teaching	3.59	.66	54
I used the Focus4TAPS strategies regularly in my teaching	3.54	.72	54
I was able to implement the Focus4TAPS strategies effectively in my teaching	3.33	.73	54
Using the Focus4TAPS strategies increased students' engagement	3.44	.63	54
My students seemed to understand the science content better after I used the Focus4TAPS strategies in science lessons	3.28	.66	54
My students enjoyed the Focus4TAPS activities I used in my classroom	3.66	.59	53
The Focus4TAPS programme has made me more effective in teaching science	3.49	.64	**53**
The Focus4TAPS programme was inspiring	3.50	.69	54
I will be able to continue to implement Focus4TAPS without any further support or training	3.43	.72	54
I will be able to continue to implement Focus4TAPS with further support or training	3.19	.88	53

Focus4TAPS, Focus for Teacher Assessment of Primary Science; SD, standard deviation.

What impact does Focus4TAPS have on teachers' attitudes towards science and confidence to teach/assess Working Scientifically in science?

Responses through the teacher questionnaire and teacher interviews revealed that the Focus4TAPS programme can be inferred to have been a positive influence on teachers' confidence to teach Working Scientifically in science alongside being able to assess students' understanding of science. Specifically, Year 5 teachers within the intervention schools who completed the teacher questionnaire (on average compared to those in control schools) reported that they were more confident in teaching many aspects of science education (including inquiry methods) compared to control schools.

Table 43: Teacher questionnaire responses (Year 5 teachers) for confidence in science teaching

In teaching science to Year 5 students, how would you characterise your confidence in doing the following? (1=Low, 2=Medium, 3=High, 4=Very high)	Intervention group			Control group			Observed difference	
Questionnaire item (1–4)	Mean	SD	N	Mean	SD	N	D	p-value
Inspiring students to learn science	3.11	.60	55	2.85	.72	53	.394	.044
Explaining science concepts or principles by doing science experiments	3.15	.63	54	2.81	.69	52	.518	.009
Providing challenging tasks for the highest achieving students	2.73	.65	55	2.47	.70	53	.379	.052
Adapting my teaching to engage students' interest	3.02	.57	54	2.87	.68	53	.241	.217
Helping students appreciate the value of learning science	3.20	.62	55	2.85	.86	53	.468	.018

Assessing student comprehension of science	2.84	.66	55	2.57	.72	53	.391	.045
Improving the understanding of struggling students	2.69	.84	55	2.70	.64	53	.010	.960
Making science relevant to students	3.02	.62	55	2.91	.79	53	.158	.415
Developing students higher-order thinking skills	2.72	.66	53	2.57	.82	53	.203	.300
Teaching science using inquiry methods	3.02	.59	55	2.58	.66	53	.690	.001

SD, standard deviation.

Majority of Year 5 teachers within the intervention schools who completed the teacher questionnaire further conveyed that the Focus4TAPS programme had specifically changed how they taught Working Scientifically (via responses to 'Has Focus4TAPS changed how you teach Working Scientifically?'; 47 free-text responses to that question conveyed positive changes while the other 3 responses conveyed no changes occurring). Teachers' responses tended to cite a greater emphasis towards Working Scientifically in general, and a focused/clearer approach (exemplified through the TAPS approach of 'one focus at a time') that then led to greater coverage/improvement over the year.

Further narratives within the interviews showed that teachers were able to gather evidence of their pupils' learning through observation and use that to provide responsive teaching. For example, a Year 5 teacher within the intervention group conveyed:

> *I suppose it made me a little bit more [focused]. I'm focused and how I was approaching things in particular, so I guess certain terminology was being a lot more focused … in terms of their own independent investigation, so come and letting them go free. You could actually see how they were working things out for themselves a lot better. So yeah, definitely it was helpful … Yeah much, much better [at being confident in teaching science and assessing children Working Scientifically] because I know what I'm looking for now, it just it just [sic] makes it makes sense.*

Taking part in the training has also enabled teachers to appreciate a different style of teaching scientific knowledge across to pupils. For example, a Year 5 teacher within the intervention group conveyed:

> *Yes, I think it's [TAPS training] made me feel less overwhelmed about science lessons. Actually I think before I would think oh it's a whole investigation. We've got to get all of this equipment. We've got to get all of this written in their books. But now it's less overwhelming. I look forward to it a bit more than I did before. I've always liked to teach in science, but it's now—it's less overwhelming, I think would be the right way to put it. I think [confidence] it's definitely grown. I think by feeling like there's less to fit in a lesson then after I automatically feel more confident and there's more time for discussions and listening to different ideas. So that that [sic] makes you feel a lot more comfortable.*

Focus4TAPS contributed to enhanced confidence in being able to teach pupils to work scientifically and assess them Working Scientifically and this was also apparent in teachers who had a science subject specialism. While some teachers may have had a prescriptive mindset because of their scientific subject specialism there was still some gains to be had by taking part in the TAPS training. For example, a Year 5 teacher within the intervention group conveyed:

> *I was a chemistry teacher way back. That was what I started my teaching career doing and then when I came back into teaching primary, I definitely had that same sort of mindset which was actually quite prescriptive and quite knowledge based … But certainly when I started attending the training days … I thought this is great because it is so much more applicable to primary and the focus was much less on learning these facts, writing down, recording what you did in a certain order and doing everything you know from start to finish. In investigation the focus on certain areas. Of the reporting and whether it was in designing the investigation, using equipment, taking accurate measurements, being able to*

> *use graphs to show your results or how to draw conclusions, whatever it is you were focusing on. I felt it was much better because I think children then had a much better understanding of each process rather than trying to do all of them for each investigation, which actually you couldn't really do in one session … I feel that children have got much more out of this way of doing things because it focuses on certain areas at a time and I think children can do that much better, so it's definitely changed my approach to teaching science in primary school … to assess where children are [in understanding], which is the most difficult bit really to assess what children really understand about what they're doing, and not just following up a set of instructions. But do they really understand? And can they explain what they've been doing and why they've been doing it? And so yes, I feel very confident to keep going with it, and I will keep going with it.*

The planned impact of the Focus4TAPS programme from the perspective of the trainer was:

> *The main hope is that it will support their confidence in teaching science, which is still a big issue in primary schools because we're generalists, we're not specialists, and so the main aim is to support the teachers to be confident in what they're doing. I'm hoping that that will then support them to be able to make decisions about their teaching which will support the learners rather than stick to a pre-planned scheme necessarily that is generic, if that makes sense. So that confidence is the main thing*.

The survey and the interviews revealed that the TAPS training helped with Working Scientifically more so for teachers who did not have a subject specialism. For example, a Year 5 teacher within the intervention group conveyed:

> *Yeah, absolutely. I mean it did. With my confidence I was able to try more daring things. We made a sound wave. Which was a huge sound wave that crossed our whole room and it was fantastic. So yeah, it gave me confidence that you can do quite daring things within a class and children can be quite in or [in awe] of science. So that's really and that's when they learned the most is when you do something quite extraordinary.*

The findings are not entirely clear what impact science subject specialism had on the delivery of the teaching on pupils (and what effect this had on pupils) and whether having a specialism in science had an influence on how useful the programme was for those with and without specialisms.

Compared to those in the control group, Year 5 teachers in the intervention group had a higher proportion of those who studied science/STEM in post-secondary education (via the questionnaire item 'During your post-secondary education, what was your major or main area(s) of study?') and, on average, also reported higher confidence for many areas within the questionnaire. Given that teachers were only surveyed at one point in time, it remains unclear whether the confidence of teachers in the intervention group was initially high or whether their confidence increased across the year. It also remains unclear whether teachers' confidence follows from their subject specialism and/or from other aspects of life. Further research may benefit from considering this area in more detail.

What impact does Focus4TAPS have on teachers' assessment (and other) practices?

Data from various sources (teacher survey, teacher interviews, pupil interviews, and the trainer interview) indicated that the Focus4TAPS programme can be inferred to have been a positive influence on teachers' assessment practices.

Specifically (as below), Year 5 teachers within the intervention schools who completed the teacher questionnaire (on average compared to those in control schools) reported that they were more confident in many aspects of science assessment and Working Scientifically.

Table 44: Teacher questionnaire responses (Year 5 teachers) for confidence in assessment and Working Scientifically

Thinking about the assessment of science (including Working Scientifically) for Year 5 students, how would you characterise your confidence for the following? (1=Low, 2=Medium, 3=High, 4=Very high)	Intervention group			Control group			Observed difference	
Questionnaire item (1–4 scale)	Mean	SD	N	Mean	SD	N	D	p-value
Understand the learning objectives and criteria for success	3.21	.62	56	3.08	.68	53	.214	.268
Undertake pedagogical planning to elicit students' science knowledge and skills	2.91	.67	56	2.58	.69	53	.479	.014
Involve students within planning	2.38	.89	56	1.92	.79	52	.538	.006
Gather evidence of students learning through questioning, discussion, and observation	3.04	.66	56	2.75	.70	53	.412	.034
Gather evidence of students learning through studying the products of activities and tasks	2.93	.63	56	2.79	.72	52	.207	.287
Deliver responsive teaching such as adapting the pace, challenge, and content of activities	3.05	.62	55	2.79	.64	52	.423	.031
Allocate time for students to reflect on and assess their own work	2.64	.80	56	2.42	.80	53	.286	.138
Use moderation and discussion with other teachers to align judgements about students	2.50	.85	56	2.06	.84	53	.523	.007
Base summative judgements of students learning on a range of types of activity	2.82	.64	56	2.35	.86	52	.632	.002
Draw on a range of information to summarise students' progress across the cohort	2.68	.61	56	2.43	.89	53	.323	.098
Summarise achievements in terms of what students can do, not only in terms of levels, grades, or percentages	2.89	.65	56	2.42	.91	53	.607	.002
Provide parents/carers with oral and written reports that identify the next steps for their children	2.71	.88	55	2.43	.95	53	.301	.121
Provide feedback to students about how to improve	3.05	.67	56	2.66	.87	50	.509	.011

SD, standard deviation.

Majority of Year 5 teachers within the intervention schools who completed the teacher questionnaire also conveyed that the Focus4TAPS programme had specifically changed how they applied formative assessment (via responses to 'Has Focus4TAPS changed how you apply formative assessment?'; 41 free-text responses conveyed positive changes while the other seven responses conveyed no changes occurring). Responses included that the programme had provided an emphasis on formative assessment, provided an avenue for approaching assessment (through the principle of assessing one aspect of science at a time), and through various approaches including assessment feeding into planning, listening to and observing pupils, and using more frequent elicitation of understanding/knowledge.

Pupil interviews indicated that they felt their learning of science in Year 5 has been better than their learning of science in Year 4 (79 of 88 interviewed pupils reported this). Also, the pupil interviews indicated that majority of pupils felt that their teachers tell them how to do better in science (85 of 93 interviewed pupils), supports them in learning science (91 of 94 interviewed pupils), and their teacher tells them how well they are doing in science (78 of 93 interviewed pupils). In addition, 91 out of 94 interviewed pupils reported that they had a chance to take part in class discussions.

Triangulating findings from various data sources indicates that while the programme boosted confidence in science teaching, it also skilled teachers about Working Scientifically, and being confident about Working Scientifically and built a shared understanding with peers about what progress can be expected with the programme (e.g. see Table 41, Table 43, and

Table 44). The teacher interviews indicated that the Focus4TAPS programme helped build their skills in teaching science. Teacher interviews reported that their skills at assessing children's knowledge of Working Scientifically had improved. For example, a Year 5 teacher within the intervention group conveyed:

> *I think that that's probably the thing that I need to focus on the most, I would say. I think we're quite good at assessing children's knowledge of science, the actual factual side of it, but perhaps assessing specifically where we need to develop their Working Scientifically skills is something that TAPS has really made me think about. Particularly, I think I mentioned this before, you know, just the parts of the process that often get left out. So, when we were trying to teach an entire experiment in one go, you would definitely do the planning stage of the experiment then talk about the testing, and then collect the data and perhaps the end of the process was the part that got missed out with it, where you really interpret the data and think about the conclusion. So yeah, I think it's just made me think about that, and that's probably something that I really need to think about next year, going forwards, I've got year five again, just really using those assessments to make sure that we're using our time in the best possible way. So time is a constraint for science, but if we assess them really carefully, we can really see where we need to focus our teaching.*

Dosage

How does dosage vary across classrooms, what factors affect dosage and to what extent does dosage impact on outcomes?

The indicators of compliance (including the number of TAPS activities/lesson plans reported as delivered by the main Year 5 teacher), together with the teacher and pupil interviews, suggested that dosage could vary, potentially linked with and influenced by the impact of the pandemic on schools.

Interviews with pupils from intervention schools indicated that a large number of pupils (70 of 88 interviewed pupils) reported that they took part in science lessons at home when schools were in lockdown. In addition, just over a half of pupils interviewed in intervention schools reported that science lesson at home were interesting. Pupil interviews asked whether the pandemic had an effect on learning science in general across Year 5 (not just at home); just over half of the interviewed pupils (52 of 99 pupils) reported that the pandemic had an effect on science learning at Year 5. Equally around half of the interviewed pupils reported that they didn't have the right resources at home to take part in science learning (48 of 89 pupils). Teacher interviews indicated that there was some way around the pandemic by teaching the non-practical elements of the programme during online learning.

Nevertheless, the reported number of TAPS lesson plans/activities delivered by the main Year 5 teacher had no associations with their other responses on the teacher questionnaire; in this context, the relatively small numbers of teachers may limit the potential for associations to be revealed (responses were available from 56 main Year 5 teachers within intervention schools alongside responses from 55 main Year 5 teachers within control schools). Similarly, it is difficult to ascertain via the pupil and teacher interviews whether variation had impacts on the primary outcome or secondary outcomes. For example, the wider context also involved the COVID-19 pandemic impacting teaching/learning in general, as a Year 5 teacher within a control school conveyed:

> *COVID has impacted negatively. Due to the lockdown children experienced in the previous academic year, when the curriculum was not being covered in the same way, it meant that pupils entered Year 5 with huge gaps in learning which have had to be addressed. Those children who did not engage in online learning during Year 5 have added to these gaps further—some of them missing entire units of work that their classmates completed.*

The observed correlations between the pupil-level indicators and aspects of the school-level compliance indicators suggested that increased dosage (inferred as being shown through increased compliance and provision of lesson plans/activities) associated with some although not necessarily all pupil outcomes.

Table 45: Observed correlations between outcomes and compliance indicators for intervention group pupils

	Observed correlation with school-level numeric (0–3) 'compliance' indicator			Observed correlation with school-level number of delivered TAPS lesson plans/activities (as reported by the main Year 5 teacher)		
Pupil-level outcome or indicator	R	p-value	N	R	p-value	N
Science attainment test scores	.091	.001	1,316	.010	.725	1,225
Interest and enjoyment in science	.089	.002	1,257	.100	.001	1,193
Confidence in science	.046	.101	1,255	.053	.065	1,192
Perceptions of science teachers	.079	.005	1,255	.085	.003	1,191
Self-regulation of learning in science	.087	.002	1,252	.074	.011	1,188
Self-efficacy for Working Scientifically	.042	.140	1,245	.014	.623	1,185
Working Scientifically beliefs	.058	.040	1,244	.014	.635	1,184
Wider benefits of science	.047	.096	1249	-.007	.803	1,187
'How often do you do science at school?'	.016	.568	1,252	.012	.668	1,189
'How often do you do investigations (practical work) in science?'	.154	<.001	1,226	.073	.013	1,164

TAPS, Teacher Assessment in Primary Science.

Further correlations between the pupils' responses suggest that pupils' reports of undertaking more frequent practical work positively associated with many of their attitudes and orientations towards science, including their interest and enjoyment in science. Nevertheless, there were no clear correlations between the pupil-level response to the question 'How often do you do investigations (practical work) in science?' and their science attainment test scores. This may suggest that the provision of practical work alone may not be sufficient for higher attainment, and that any associations may be indirect and complex (such as practical work being perceived as interesting and enjoyable, which may then motivate or inspire pupils to study and engage further, which may then foster attainment). Nevertheless, more extensive research would be required in order to consider associations, given that science attainment and attitudes may be influenced by many aspects of education and life.

Table 46: Observed correlations between outcomes and reported frequency of science teaching/learning

	Observed correlation with the pupil's response to 'How often do you do investigations (practical work) in science?' for intervention group pupils			Observed correlation with the pupil's response to 'How often do you do investigations (practical work) in science?' for control group pupils		
Pupil-level outcome or indicator	R	p-value	N	R	p-value	N
Science attainment test scores	.004	.901	1203	-.037	.209	1173
Interest and enjoyment in science	.154	<.001	1226	.144	<.001	1199
Confidence in science	.092	.001	1225	.090	.002	1198
Perceptions of science teachers	.237	<.001	1224	.221	<.001	1191
Self-regulation of learning in science	.178	<.001	1224	.192	<.001	1196
Self-efficacy for Working Scientifically	.137	<.001	1215	.170	<.001	1176
Working Scientifically beliefs	.081	.005	1214	.108	<.001	1173
Wider benefits of science	.088	.002	1219	.071	.016	1176

How well are different elements of the intervention delivered and what factors affect this quality of implementation?

Triangulation from various sources of data indicated that a key issue that impacted the quality of implementation was school closures via lockdown rather than the programme itself. Pupil interviews highlighted that their perceptions about the quality of learning at home was not the same as what they had at school. Teachers in some schools continued to implement TAPS but the engagement from the pupils' perspective was not the same. For example, a pupil within an intervention school conveyed:

> *We didn't do any practical experiments, we saw people doing them but actually one of our teachers showed us PowerPoints but it was quite a big disappointment not being able to do it but then again I do understand why we couldn't do it and maybe people didn't have the right equipment and it could be dangerous if you had other siblings in your house so I do understand it but we didn't get to do practical experiments and that was a real shame to me.*

From the perspective of teachers some were able to juggle around the learning requirements during school closures and try and balance out the theory with the investigations. Some schools were able to continue with practical experiments in lockdown and achieve a balance across the school year about the amount of theory and practical that was needed. For example, a Year 5 teacher within an intervention school conveyed:

> *So we used one of their investigations [in lockdown], we used the Crater investigation, and the children had been in lockdown, and when they came back we used lots of the investigations to make sure that*

> *we had that practical element, because they'd done a lot of theoretical learning via lockdown, and we used the Crater experiment.*

Perceived areas for improvement

Year 5 teachers within the intervention schools who completed the teacher questionnaire conveyed (through their free-text responses to 'Are there any aspects of the Focus4TAPS programme that could be improved?') that the Focus4TAPS programme could have been improved through:

- increasing and refining the scope/range of topics;
- refining/clarifying the plan/materials in general; and
- providing refined or wider guidance (including how to integrate lessons within sequences/the National Curriculum framework and regarding assessment in general).

Nevertheless, the range of other responses suggested that teachers were often receptive and positive, so it remains difficult to determine the extent of any potential limiting factors.

To what extent does the intervention reach all (intended) pupils and what factors affect this?

It was difficult to directly ask pupils through interviews whether they felt the intervention had an impact on them, as many pupils would not have known they had different teaching than what they ordinarily would have received. So, questions were asked in order to consider the intended outcomes from the intervention logic model. As indicated earlier, majority of pupils reported positively about teachers' support in learning science (91 of 94 pupils), on receiving information on how well they were doing (78 of 93 pupils), feedback on doing better, taking part in class discussion (85 of 93 pupils said that they had opportunities), and planning their own practicals (67 of 93 pupils).

The quantitative analysis indicated that the intervention did not appear to have a different impact on different categories of pupils. The triangulation between the qualitative data (teacher and pupil interviews) indicates that there was no apparent difference between pupils in their experiences of the intervention. One teacher via interview conveyed that pupils with EAL were considered to find TAPS less accessible; conversely, one teacher via the questionnaire conveyed that pupils with EAL were perceived to develop their language through the practical experiences, and other teachers conveyed that less focus on written work was considered to increase accessibility for pupils in general. There was also some (qualitative) data to suggest that higher ability children did better than lower ability pupils, but these findings were not uniform across the sample. These conflicting findings will be discussed in the implications section.

Nevertheless, some of Year 5 teachers within the intervention schools who completed the teacher questionnaire conveyed that they believed that the Focus4TAPS programme was accessible for most pupils and especially those who might have found routine teaching/learning less accessible. For example, one teacher conveyed: '*It worked well for all students but especially LA [lower ability] and SEN children who find it difficult to access text heavy lessons'.* Other teachers conveyed that they believed the programme worked particularly well for particular groups including:

> *The children that struggle to record their learning in written form. The practical aspects have allowed me to assess understanding using a whole range of methods e.g. recording voice notes, photographs, diagrams/pictures.*

Another teacher conveyed that they believed the programme *'worked well for children with literacy difficulties'* because it allowed a focus on speaking and listening and therefore less writing, and the teacher was then '*able to assess their science skills and knowledge as opposed to their writing skill'.*

To what extent does practice in control schools impact pupil outcomes? Is there a relationship between practice and how pupils can work scientifically?

Pupils within intervention schools, on average, reported that they experienced more frequent science work and investigations (practical work) in science compared to pupils within control schools. This suggests that the

Focus4TAPS programme may have increased the provision of science including practical work (or that the provision was already high within intervention schools); alternately, this may reflect greater focus on science or greater visibility of science from the perspective of pupils (which may be different to the frequency or type of provision). The responses through the teacher questionnaires (as conveyed elsewhere) nevertheless suggest that intervention pupils may have indeed received more science teaching/learning compared to control pupils.

The correlations between pupils' responses (reported earlier in

Table 46) show correlations between some (although not all) measured outcomes and the pupils' reported frequency of experiencing science teaching and learning. For pupils in control schools and also for pupils in intervention schools, pupil's reported frequency of involvement in practical investigations had positive correlations with their reported:

- interest and enjoyment in science;
- confidence in science;
- perceptions of science teachers;
- self-regulation of learning in science;
- self-efficacy for Working Scientifically;
- Working Scientifically beliefs; and
- wider benefits of science.

The pupils' responses to other questionnaire items (Table 38; covered in an earlier section) that were intended to consider aspects of the wider TAPS approach, such as 'My teacher plans and discusses science lessons with us' and 'I check my own work to find out what I have learned in science', involved no clear differences across intervention and control groups. It is possible that these aspects may not necessarily have been central to the delivery of the Focus4TAPS programme compared to other aspects, such as following and adapting various lesson plans.

Pupils in control schools compared to the TAPS schools were equally likely to report that they would like to do more investigations and that they liked to do science investigations (Table 47). Pupils from control and intervention schools had similar perceptions about how important science is and that it is important to do well in science.

Table 47: Pupil questionnaire responses for further IPE items

	Intervention group			Control group			Observed difference	
Questionnaire item (1–4 scales)	Mean	SD	N	Mean	SD	N	D	p-value
I like to do science investigations	3.33	.92	1,240	3.32	.95	1,201	.015	.708
I would like to do more science investigations	3.25	.97	1,240	3.24	1.01	1,200	.010	.806
I would like a job that involves using science	2.15	1.12	1,244	2.18	1.10	1,207	.023	.570
It is important to learn about science to get ahead in the world	3.18	.87	1,241	3.27	.83	1,199	.100	.013
My parents think that it is important that I do well in science	2.94	.93	1,226	2.91	.94	1,182	.033	.416
It is important to do well in science	3.22	.83	1,240	3.28	.82	1,201	.071	.078

You have to be clever to do science	1.93	1.06	1,241	2.02	1.10	1,204	.080	.049
Anyone can do science and be a scientist	3.35	.94	1,244	3.37	.91	1,202	.015	.710

IPE, implementation and process evaluation; SD, standard deviation.

Costs

Information about the time and costs associated with the Focus4TAPS programme were gathered through the teacher questionnaire and directly from the delivery team. Teacher questionnaires were completed by Year 5 teacher and science leaders (some teachers may have held both roles).

The delivery team conveyed the following:

- in-person training involved 3 days (4.5 hours each day for a total of 13.5 hours); online training involved 9 hours (six sessions of 1 hour, with 30 minutes expected for preparation/follow-up per session);
- preparation for lessons was expected to require at least 2 hours (30 minutes per lesson plan) and more likely to require 4–5 hours;
- delivery of lessons was expected to require at least 4 hours (1 hour per lesson) and more likely to require 8–10 hours; and
- applying wider aspects was expected to vary depending on teachers/schools, potentially requiring at least 2 hours and more likely to require 8–10 hours.

Questions for teachers covered the time (in hours) spent: attending Focus4TAPS training; preparing Focus4TAPS lesson plans, tasks, and other aspects for delivery; delivering Focus4TAPS within science lessons; and applying other Focus4TAPS aspects within the school such as approaches to assessment and other aspects of the TAPS pyramid. From intervention schools, questionnaire responses were received from Year 5 teachers and science leaders (where some teachers had Year 5 and science leader roles). Responses were summarised per question per role: from the perspective of those with Year 5 teacher roles (56 teachers) and from the perspective of those with science leader roles (58 teachers), with some unavoidable overlap (18 teachers had Year 5 and science leader roles). Some teachers (7–12 of the Year 5 teachers, varying per question) provided narrative/generalised responses that could not be quantified (and were treated as missing for the quantitative summaries), mainly conveying that they were unsure with some general responses (such as a ‘few’ hours). Accordingly, it is possible that the findings may potentially underestimate or overestimate times; for example, asking people to recollect time spent may unavoidably introduce some uncertainty.

Table 48: Teacher questionnaire responses for reported time spent onFocus4TAPS

	Teachers	Reported time in hours
Focus4TAPS aspect (teacher role)	n/N	Mean [95% CI] (SD)
Attending Focus4TAPS training (All teachers with Year 5 teaching roles)	45 / 56	14.38 [11.12, 17.64] (10.86)
Attending Focus4TAPS training (All teachers with science leader roles)	47 / 58	13.83 [11.09, 16.57] (9.32)
Preparing Focus4TAPS lesson plans, tasks, and other aspects for delivery (All teachers with Year 5 teaching roles)	38 / 56	9.57 [7.25, 11.88] (7.05)
Delivering Focus4TAPS within science lessons (All teachers with Year 5 teaching roles)	38 / 56	14.03 [11.00, 17.05] (9.21)
Applying other Focus4TAPS aspects within the school, such as approaches to assessment and other aspects of the TAPS pyramid (All teachers with Year 5 teaching roles)	35 / 56	4.60 [2.82, 6.38] (5.17)

Applying other Focus4TAPS aspects within the school, such as approaches to assessment and other aspects of the TAPS pyramid (All teachers with science leader roles)	37 / 58	8.00 [4.88, 11.12] (9.36)

CI, confidence interval; Focus4TAPS, Focus for Teacher Assessment of Primary Science; SD, standard deviation.

Given the average responses from the teachers (and some variability), the cost calculations then assumed the following (rounding up to the nearest 5, in order to provide more conservative and easily interpretable values that were still informed by the data):

- training required 15 hours from Year 5 teachers and 15 hours from science leaders;
- preparation for lessons required 10 hours from Year 5 teachers;
- delivery of lessons required 15 hours from Year 5 teachers; and
- applying wider TAPS aspects required 5 hours from Year 5 teachers and 10 hours from science leaders.

A further question considered what existing equipment, materials, and resources teachers used (or were expected to be needed by the delivery team) in order to deliver the Focus4TAPS programme (and how much these costed).

The delivery team believed that teachers would apply their existing equipment and resources (essentially, whatever teachers would be using in 'business as usual' lessons). The questionnaire responses from Year 5 teachers/science leaders conveyed that a variety of existing resources were used, including common science-specific equipment (such as beakers, thermometers, magnifying glasses, sieves, magnets, etc.), science-related consumables (flour, sugar, salt, food dyes, etc.), and general/craft-related consumable materials (such as paper and card, tape, bubble wrap, straws, string, bottles, etc.). As these covered existing resources, many teachers/leaders conveyed that they were not easily able to quantify the costs; values from £4 to £300 were cited. One exception involved computer tablet devices, which were cited as used within a research-focused lesson (cited cost of £7,000); however, it is reasonable to infer that this may be an exceptional case, given no other mention of computer equipment and given the intended scope of the Focus4TAPS activities.

A further question then considered what new equipment, materials, and resources (which schools did not already have) were used by teachers (or were expected to be needed by the delivery team) to deliver the Focus4TAPS programme (and how much these costed).

The delivery team believed that the activities would not require additional equipment and resources (essentially, the activities would involve adapting existing teaching rather than undertaking something entirely new). Summarised on the school-level (considered across Year 5 teacher and science leader responses): 15 schools conveyed no need of additional resources; 30 schools had a teacher/leader citing new/additional resources; 8 schools did not know; while the other intervention schools had no information (the question may not have been answered or teachers may not have completed the questionnaire). The responses tended to convey that new/additional resources involved the same or similar consumables from the array of earlier science/craft materials; values from £1 to £300 were cited. Two schools listed particular scientific equipment (sets of beakers, thermometers, microscopes, stopwatches, minibeasts sets, and heart rate monitors); values of £50 to £250 were cited for each piece/set of equipment giving one response a total of £250 and another response a total of £560.

Overall, it was possible to infer that Focus4TAPS involves applying existing equipment and resources that may be already available within many schools, and that any additional requirements typically involve 'more of the same' resources such as consumables. It is possible that additional but common scientific equipment might also be required, although it was reasonable to infer that this may involve 'case by case' instances depending on particular schools and their circumstances.

The final cost calculations assume 24 pupils per class (for per pupil cost per year) and assume £40.17 per hour for teaching cover. The calculations show a default position where Focus4TAPS is assumed to require no further resources, materials, and equipment and is assumed to adapt existing teaching/learning work. Essentially, the process assumes that teaching/learning in schools would involve time spent preparing lesson plans, delivering lessons, and applying wider assessment for learning principles and practices in some way.

Table 49: Cost of delivering Focus4TAPS

Item	Type of cost	Cost per school per year	Total cost per school over 3 years	Total cost per pupil per year over 3 years (assuming 24 pupils per class per year)
Training teacher cover: Year 5 teacher for 15 hours	Start-up cost (Year 1 only)	15 hours × £40.17 = £602.55	£602.55	£602.55 / (3 × 24) = £8.37
Training teacher cover: science leader for 15 hours	Start-up cost (Year 1 only)	15 hours × £40.17 = £602.55	£602.55	£602.55 / (3 × 24) = £8.37
Training costs: online course (estimated costs by delivery team)	Start-up cost (Year 1 only)	£150.00	£150.00	£150.00 / (3 × 24) = £2.08
Programme costs: estimated by delivery team	Recurring cost (Year 1, 2, and 3)	£0.00	£0.00	£0.00
Facilities, equipment, and materials	Not applicable	–	–	–
Preparation for lessons: Year 5 teacher for 10 hours per year	Not applicable	–	–	–
Delivery of lessons: Year 5 teacher for 15 hours per year	Not applicable	–	–	–
Supporting/applying other TAPS aspects: Year 5 teacher for 5 hours per year	Not applicable	–	–	–
Supporting/applying other TAPS aspects: science leader for 10 hours per year	Not applicable	–	–	–
Total			£1,355.10	£1,355.10 / (3 × 24) = £18.82

TAPS, Teacher Assessment in Primary Science.

Table 50: Cumulative costs of Focus4TAPS (assuming delivery over 3 years)

	Year 1	Year 2	Year 3
Focus4TAPS	£1,355.10	£0.00	£0.00

Focus4TAPS, Focus for Teacher Assessment of Primary Science.

Conclusions

Table 51: Key conclusions

Key Conclusions
Pupils in Focus for Teacher Assessment of Primary Science (Focus4TAPS) schools made the equivalent of two additional months' progress in science, on average, compared to pupils in other schools. This result has a high security rating.
Pupils eligible for Free School Meals (FSM) in Focus4TAPS schools made the equivalent of two additional months' progress in science, on average, compared to pupils eligible for FSM in other schools. This result is based on smaller numbers of pupils and may need to be interpreted with caution.
Pupils in Focus4TAPS schools had similar, and positive, attitudes and orientations towards science to pupils in other schools.
The programme ran as expected, with 80% of teachers attending the minimum number of training sessions and 63% of teachers conducting at least four 'Focused Assessment' lessons using Focus4TAPS materials. Greater compliance with the programme (e.g. attending more training sessions and conducting more lessons) was associated with higher science test scores.
Teachers in Focus4TAPS schools reported higher confidence for some aspects of teaching and assessing science than teachers in other schools, and believed that various benefits followed from the programme, e.g. changing how they taught Working Scientifically and how they applied formative assessment.

Impact evaluation and IPE integration

Evidence to support the logic model

The Focus4TAPS programme aimed to refine teaching approaches and assessment around Working Scientifically in science. The programme was conceptualised through the logic model as having the potential to influence teachers' practices (which are also potentially dependent on and influenced by teachers' background and context); teachers' practices may then influence aspects of pupils' learning (which are also potentially dependent on and influenced by pupils' background and context); pupils' learning is then observable through pupils' attainment in science and pupils' attitudes and orientations towards science (which are likely to associate). The evaluation of the Focus4TAPS programme involved a randomised trial to consider whether any differences were observable across those assigned to an intervention group who received the Focus4TAPS programme and those assigned to a control group who undertook their usual practices within science teaching/learning.

The Focus4TAPS trial commenced with Year 5 pupils during the 2019/2020 academic year. From March 2020, the trial was unable to continue given teaching/learning disruption following from the COVID-19 pandemic. The trial then recommenced with Year 5 pupils during the academic year of 2020/2021. Recommencing the trial had minimal to no observable impact on the profiles of schools and pupils within the trial.

After recommencing, the Focus4TAPS trial considered 2,882 pupils in Year 5 within 121 schools (1,480 pupils within 61 intervention schools; 1,402 pupils within 60 control schools) during the 2020/2021 academic year. The profiles of pupils and schools were broadly balanced across the intervention group and the control group, with the partial exception of the pupils' prior attainment in Key Stage 1 reading (more pupils in the control group had the WTS category in reading than the intervention group); the primary and secondary analysis accounted for the pupils' prior attainment in reading and in mathematics order to mitigate any potential initial imbalance. Year 5 teachers within the intervention group and the control group also, on average, reported similar backgrounds (other than more STEM specialists in intervention schools), characteristics, and contexts including for their average number of years spent teaching and for facilities and resources available for Year 5 pupils at their schools. However, more teachers in the intervention group than the control group reported that their post-secondary education (i.e. university degree or equivalent) involved STEM subjects although it remains unclear how/why such a difference may have arisen.

In total, 99 pupils from intervention schools took part in interviews before the intervention and we invited all 99 to take part in interviews after the intervention to inform the IPE, covering 10 pupils per school in 10 schools (except in one case with nine pupils per school). We invited all teachers to participate in interview pre- and post-intervention. All 10 were interviewed before the intervention; six were interviewed immediately after the intervention. A further three were interviewed around 5 to 6 months after the intervention ended (due to the impact of COVID, or teacher workload, or retention, or personal issues). In one of the schools interviewed 5 to 6 months after the end of the intervention, a second teacher joined in, this was not requested from our end rather the teachers were enthusiastic in participation.

The impact evaluation revealed that, when considering observed averages, pupils assigned to the Focus4TAPS intervention group gained higher science test scores than pupils assigned to the control group. Analytical modelling revealed that even when accounting for the pupils' prior Key Stage 1 attainment, the intervention positively associated with science test scores. Additional modelling affirmed that the positive association with pupils' science test scores was revealed even when also accounting for other characteristics of the pupils (including their age, gender, and ever-eligible for FSM status), and also through alternate analysis that accounted for some (although not necessarily all) missing information. Further modelling also associated higher levels of compliance with the Focus4TAPS programme with higher science test scores. Interview narratives from some of the teachers indicated that an increase in attainment levels were expected, some schools joined the programme for the purpose of boosting attainment levels, which supports the outcomes in the logic model.

The impact evaluation also revealed that, considering observed averages, pupils assigned to the Focus4TAPS intervention group and pupils assigned to the control group conveyed similar (and positive) attitudes/orientations towards science, considered across a range of views including the pupils' interest in science, confidence in science in general, and confidence in undertaking aspects of Working Scientifically. Analytical modelling affirmed that there were no associations between the intervention and these outcomes, when accounting for the pupils' prior Key Stage 1 attainment. The IPE further affirmed that the interviewed pupils within intervention schools were indeed generally positive about science at school towards the start and towards the end of the Focus4TAPS programme. Within these pupil interviews occurring towards the start and towards the end of the Focus4TAPS programme, similar numbers of pupils reported liking practical experiments the best about science and having opportunities to take part in class discussions about science topics. A large majority of the interviewed pupils reported that they preferred the science lessons that took place in Year 5 as compared to Year 4, and 91 of the 94 pupils conveyed positive views about science at school and believed that their teachers supported them in their learning of science at school. The pupil interview narratives also suggested that their views around what they liked best about learning science in school remained relatively unchanged over time; when interviewed towards the start and towards the end of the programme, similar numbers of pupils conveyed that they liked practical work best within science (66 pre-intervention and 70 post-intervention). Future research and trials may gain more insight through surveying pupils at multiple time points in order to gain greater clarity into trajectories of changes or consistency in views, and to consider associations when accounting for pupils' prior attitudes/orientations rather than their prior attainment. The evaluation originally planned to survey pupils at multiple time points, although this became unfeasible during the recommencement of the trial given the disruption following from the COVID-19 pandemic.

The Focus4TAPS programme intended for teachers to apply TAPS lesson plans/activities and select a focus for teaching, learning, and assessment for each of their practical science lessons (so that the full breadth of enquiry skills would be covered across the year) and then adapt the level of support and challenge during subsequent lessons. Many of Year 5 teachers within the intervention group provided narrative responses through the teacher questionnaire that conveyed an understanding of the Focus4TAPS principles including focusing on one aspect of Working Scientifically at a time and conveyed that programme materials such as lesson plans were applied that, strategies were applied to elicit initial views/knowledge from pupils and that, writing was reduced, along with various other applications. The Focus4TAPS programme was initially commissioned pre-pandemic, and schools may have faced many challenges over 2019/2020 and 2020/2021 including when schools were in lockdown. For example, some teachers within intervention schools conveyed through interviews that, at times, they were reliant on parental support or parental/pupil access to resources and this was not always possible for some home learning. The pupil interviews within intervention schools revealed that only half of pupils reported having access to resources to do science learning at home. Nevertheless, the IPE unavoidably focused on intervention schools, and it remains unclear whether control schools faced similar/different challenges.

Year 5 teachers in the intervention group reported, on average, higher confidence then those in the control group for some (although not all) of the considered aspects of science teaching and assessment. These differences support the plausibility of the intervention enhancing particular aspects of practice, as reflected through teachers' reported confidence. These findings can be considered to help characterise the programme as it was experienced and applied by teachers; these findings can also be interpreted through the logic model as helping to potentially explain the findings from the impact evaluation.

The Focus4TAPS programme and logic model intended to refine science teaching/learning rather than increase the extent of teaching/learning. However, the IPE found that pupils and teachers within intervention schools reported experiencing or delivering more science teaching/learning, on average compared to those in control schools. It is possible that the Focus4TAPS programme may have indirectly inspired more science teaching/learning and directly facilitated more science teaching/learning through aspects of the training (perhaps including guidance on how to undertake science activities and mitigate disruption around the pandemic). Alternately, it is possible that baseline differences were present and the Focus4TAPS programme may have also inspired or facilitated more teaching/learning. Baseline surveying was considered to be an unfeasible burden given the pandemic and could not be undertaken; therefore, it remains unclear whether these observed differences in encountering or delivering extents of teaching/learning around science followed from an initial difference and/or from the Focus4TAPS programme. Nevertheless, further analysis revealed that the effect of the intervention remained when controlling for the extent of science teaching/learning, which suggested that the programme effect in science test scores followed from Focus4TAPS teaching/learning in intervention schools rather than following from the extent of teaching/learning.

Overall, the findings suggest that the Focus4TAPS programme positively associated with Year 5 pupils' science attainment but had no clear association with their (on average already positive) attitudes and orientations towards science. The Focus4TAPS conceptualisation and logic model assumed that practical science may be positively received by pupils. The IPE highlighted that the teachers within the intervention group, and also the control group, conveyed that they believed that pupils tended to enjoy practical work and science in general. This suggests that practical work in science may indeed be an avenue for benefits in attitudes/orientations towards science, and helps affirms the underlying logic model. Within wider research, younger children often convey positive attitudes and orientations towards science (Archer, *et al.*, 2010; Hamlyn, *et al.*, 2017; Martin, *et al.*, 2016a; Mullis, *et al.*, 2020). The impact evaluation may not necessarily have been able to reveal small differences and consider the implications for all potential subgroups. Future research and trials may gain more insight through surveying pupils at multiple time points in order to gain greater clarity into trajectories of changes or consistency in views, through considering pupils' particular views about teaching/learning approaches, and through considering differences across various groups of pupils in more detail. Future research and trials may also benefit from reflection on whether particular ages and groups of pupils might be targeted, and how that might be best undertaken.

The circumstances of the trial may complicate any consideration of whether/how the findings may generalise to other situations. Majority of Year 5 teachers in intervention schools were the same in 2019/2020 and 2020/2021, and may have gained experience and familiarity with some elements of the Focus4TAPS programme across 2 years (rather than 1 year). It remains unclear whether similar or different findings would be observed within another trial. For example, experienced teachers may have increased the potential to achieve higher effects within this trial; conversely, some disruption due to the pandemic may have reduced the potential to achieve higher effects within this trial.

Does the logic model suitably address the factors that are related to both the primary outcome and secondary outcomes?

The logic model covers some factors that may be related to the primary outcome and secondary outcomes, providing an abstracted and simplified overview. Nevertheless, further clarification, detail, and insights into particular teaching/learning approaches and contextual/wider factors may be beneficial.

Data from IPE sources indicated that those involved with the programme perceived that there was improved pupil learning. The trainer interview indicated that during school lockdown because of the pandemic there were examples of pupils enthusiastically taking part in TAPS-based learning (with the help of parents) by using the environment around them. Data from teachers also indicated that there was an influence on secondary outcomes. For example, a teacher within an intervention school conveyed:

> *I think they are definitely [students more engaged because of TAPS], when … if you write, 'Science', on the board you can hear them, you can see the excitement, they're definitely more motivated and engaged in science than they used to be*.

The pupil interviews indicated that teachers were providing a supportive learning environment using the TAPS resources and methods of teaching and thus had an impact on children's learning and engagement with science (outcomes of the logic model). For example, a pupil within an intervention school conveyed:

She explains things that we're not quite sure about and if we go wrong somewhere then she'll just calmly just say to us, 'Okay, I think you might have made a mistake, could you see it?' and if you're still not sure where you've gone wrong then she'll get someone normally a partner or someone else who's finished the topic that we're doing and then we just understand it once we've, once it's been explained to you.

And of science the student reported:

I think they're normally quite interesting [science lessons] and we normally just kind of be interested in it … [interested in science because] I think just the way that it's been set out, all the experiments that we can do, that's what I enjoyed, just getting stuck in with like.

Engaging in science in an interactive way helped TAPS pupils to really understand science. For example, a pupil within an intervention school conveyed:

Definitely the experiments because that's probably the most exciting part of the science lesson. Planning [experiments] it is good because you know what you're doing but doing the experiments is always very exciting and with your partner where you just work together and it just kind of before you know it the lesson is over and you wish it would continue.

TAPS investigative lessons also helped pupils link the relationship between theory and experiments while also igniting a special interest about what pupils learnt. For example, a pupil within an intervention school conveyed:

I like the experiments and listening to theories because when I listen to theories it makes me think like about it for the whole day.

Data from teachers also indicated that there was an increase in engagement with pupils in their TAPS lessons as well as having an impact on attainment (outcome of the logic model). For example, a teacher within an intervention school conveyed:

I believe so [TAPS teaching led to a positive effect on attainment] ... I think if it had been a normal year, they would have attained, they would have achieved more. I think their engagement in science has massively increased and I can see that their attainment has definitely increased in terms of their knowledge about the scientific process, they understand what a prediction is, they understand why observations are so important, they understand how to carry out a fair test, they know what a variable is, and I think those skills were really weak beforehand. And they do seem to really love science, I think if they love it then hopefully they'll hang on to it and there'll be more interest in the future.

Interpretation

Pupil attainment

The Focus4TAPS programme provided support to facilitate the delivery and assessment of science, including assessment lesson plans and activities for covering Working Scientifically together with wider guidance. Various sources of data from teachers indicated that there was a consensus that the programme would have an impact on attainment and that some schools may join this programme for the purposes of boosting science attainment at their schools.

The Focus4TAPS trial helps provide new evidence and insights for primary school contexts. Practical work has often been promoted and encouraged within science, although with more focus towards secondary school contexts (Gatsby Charitable Foundation, 2017). From a wider perspective, the EEF had summarised some teaching/learning activities that may help improve science (at secondary schools), which included some ideas common to Focus4TAPS: facilitating pupils to direct their own learning (fostering self-regulation of learning); providing feedback to pupils (which would help pupils self-regulate their learning); and providing practical work (Education Endowment Foundation, 2018a). The EEF summary also included other areas that may not necessarily be applied through Focus4TAPS: considering pupils' preconceptions; using models to support understanding; supporting memory and retrieval (following from psychological/cognitive ideals aiming to assist memory/recollection in general rather than promoting memorisation at the expense of any/all other teaching/learning aspects such as active inquiry/creativity); and developing pupils' scientific language and writing (Education Endowment Foundation, 2018a).

Pupil attitudes and orientations towards science

The IPE highlighted that the teachers within the intervention group, and also the control group, conveyed that they believed that pupils tended to enjoy practical work and science in general. For example, teachers in control schools conveyed (when prompted for any other comments about their experiences of teaching and learning science this year through the questionnaire) that: '*Children love practical lessons and majority are always excited to see it on the timetable!*'; '*Pupils in my class really enjoy science*'; and '*When the children have been able to come into school, they have loved the practical elements of science, even more so than usual*'. Similarly, teachers in intervention schools conveyed various points (when prompted for what differences they feel the Focus4TAPS programme has made for pupils), including around pupils' interest, enjoyment, and engagement. For example, teachers in intervention schools conveyed: '*They enjoy science more*'; '*They look forward to the practical learning and activities*'; '*They are more engaged and many say science is their favourite lesson, especially practical science'*; and '*The children have really enjoyed engaging in the TAPS lessons and have been really enthusiastic towards the investigations. They seem to enjoy the lessons more and look forward to science each week. They have enjoyed using a wider range of scientific equipment and can use some of these with greater accuracy*'.

These findings broadly affirm an assumption of the Focus4TAPS logic model, where practical science may be positively received by pupils. The pupils' attitudes and orientations towards science tended to be positive, for those assigned to the intervention group and, also for those assigned to the control group, although there were no observable differences across the groups. It may have been harder to reveal differences/changes in the pupils' attitudes without undertaking baseline measures within the trial. Additionally, within wider research, younger children often convey positive attitudes and orientations towards science (Archer, *et al.*, 2010; Hamlyn, *et al.*, 2017; Martin, *et al.*, 2016a; Mullis, *et al.*, 2020). Attitudes towards science often become less positive as pupils grow older across secondary school (Mujtaba, *et al.*, 2020), so it is possible that interventions may have greater scope for impact with older pupils; if many pupils already hold positive views, then fostering increases through interventions may be less feasible. It is also possible that impacts/differences may be more nuanced than questionnaire items can consider and reveal. The pupil questionnaire considered various attitudes/orientations towards science and was designed to be comparable with various national and international research, including TIMSS (Martin, *et al.*, 2016b). Nevertheless, future research or evaluations may need to also consider more specific views from pupils regarding particular aspects of teaching/learning.

Emerging guidance around fostering inclusion/accessibility to science at primary school (Nag Chowdhuri, *et al.*, 2021) and secondary school (Godec, *et al.,* 2017) has broadly focused around recognising and supporting diversity around being and doing science (rather than focused around fostering attainment). Future evaluations may benefit from considering whether/how aspects of teaching such as science could be made more inclusive and accessible via interventions, which may involve (for example) considering more dimensions of advantage/disadvantage in addition to ever-eligible for FSM status (and considering how multiple dimensions may intersect).

The Focus4TAPS programme and trial was primarily orientated around refining teaching/learning to primarily foster attainment (with any benefit to pupil attitudes being secondary or indirect), rather than refining teaching/learning to primarily foster engagement or interest (with any benefit to pupil attainment being secondary or indirect). Pupils' attainment and attitudes towards science may associate, but indirect benefits may be delayed and arise through complex associations. Nevertheless, the IPE, through the various data sources and particularly through the lens of the teachers, identified a range of perceived impact on pupils. For example, the interviewed teachers tended to convey that their pupils had benefited from the Focus4TAPS programme, including reporting that it had a positive impact on pupils' engagement and learning in science and some talked about the cognitive benefits; these teachers often reported that the programme was worth doing and that it was worth rolling out.

The Focus4TAPS programme also did not explicitly target particular groups of pupils. The IPE highlighted that some teachers who completed questionnaires within intervention schools believed that some pupils may have found science to be more accessible, facilitated through (for example) the reduced emphasis on writing through the Focus4TAPS programme; other statements from teachers who were interviewed suggested that more scaffolding may be required to help make the programme accessible to pupils with EAL. Future research or evaluations may need to consider (and potentially target) subgroups in more detail.

Limitations and lessons learned

The Focus4TAPS trial was a two-arm (intervention and control) cluster randomised efficacy trial (with schools as the units of clustering/randomisation). Potential limitations are relevant to most if not all research and trials.

For the Focus4TAPS trial, potential pupil-level confounding factors were assumed to be mitigated through the design. Schools (not pupils) decided to be involved with the trial, and schools were then randomly assigned to the intervention group or the control group, such that pupil-level characteristics, motivations, and other aspects of life (whether measured by the trial or not) should (in theory or ideally) be balanced across both groups. Teacher/school-level factors such as motivations towards delivering optimised science teaching/learning could potentially inspire participation in a trial, greater engagement with an intervention programme, and (regardless of an intervention) foster attainment in pupils. Ideally, such issues/risks would also be mitigated through a randomised design. Nevertheless, it is possible that school-level continuation or attrition at the recommencement of the trial involved confounding factors, although any such factors could not be known or measured (given that teachers could not be subsequently surveyed in cases of attrition). For example, a potential confounding variable may be the level of subject specialism in teachers; once schools had been assigned to the intervention group, it is possible that some teachers (perhaps those with science specialisms) may have been more likely to want to become involved (or not become involved) with a programme focused around science. Future research may benefit from exploring whether/how teacher subject specialism associates with pupil performance or gains in performance.

The Focus4TAPS trial was unable to undertake baseline surveying and was unable to explore why schools (and/or teachers) may have initially engaged with the trial; additional data collection at different stages of the trial would have placed an unfeasible burden onto schools and teachers. Within the recommenced schools for 2020/2021, the profile of Year 5 teachers who completed questionnaires was similar across the intervention and control groups for some characteristics such as gender, number of years spent teaching, and having an educational specialisation in science; however, on average, more teachers in the intervention group compared to the control group conveyed that their post-secondary education (i.e. university degree or equivalent) involved STEM subjects. It remains unclear how/why such a difference may have arisen, such as whether initial selection or nomination may have influenced the profile of teachers. The trial recruitment occurred in the spring and summer of 2019 and some schools had not determined teachers for 2019/2020 (or staffing changes were occurring or pending). Schools were notified of their assignment to the intervention to control groups in June 2019; once schools were aware of their assignment, it is possible that some teachers may have had more interest in engaging with the Focus4TAPS programme within intervention schools. From the schools who recommenced, many were unable to specify a Year 5 teacher pre-randomisation (24 of 61 intervention schools and 13 of 60 control schools) although some schools had the same specified Year 5 teachers from pre-randomisation to 2019/2020 (20 of 61 intervention schools and 27 of 60 control schools). Majority of schools that recommenced then had the same specified Year 5 teachers across 2019/2020 and 2020/2021 (58 of 61 intervention schools and 54 of 60 control schools). From a wider perspective, between 2019/2020 and 2020/2021, it is possible that the COVID-19 pandemic may have reduced the extent of staff changes across schools in general.

The responses through the teacher questionnaire were similar across the intervention and control schools regarding receiving/undertaking any other specific science programmes, initiatives, or other interventions during the past two academic years. Nevertheless, it remains possible that unmeasured/unknown involvement in other concurrent initiatives occurred.

The circumstances of the trial, involving some (but curtailed) training and delivery during 2019/2020 and then recommenced training and delivery during 2020/2021, may complicate any consideration of whether/how the findings may generalise to other situations. For example, any teachers who were involved with the trial during 2019/2020 and 2020/2021 may have gained more experience and familiarity with the Focus4TAPS programme, compared to any new teachers who were only involved with the programme during 2020/2021. As conveyed earlier within the sample/school profiles, only 3 of 61 intervention schools had changes in Year 5 teachers from 2019/2020 to 2020/2021. In addition, during the academic year 2020/2021 schools experienced another 3-month lockdown so it is difficult to ascertain the quality of science learning that took place. Both teacher and pupil interviews revealed there was some disparity between the ways in which the schools operationalised the programme during school lockdown. The interview analysis indicated that 70 of 88 interviewed pupils from intervention schools reported that they took part in science lessons at home when schools were in lockdown; just over a half of the interviewed intervention pupils reported that those lessons were interesting. Learning science at home may have had impacted the extent of exposure to Focus4TAPS teaching overall and for particular pupils. Some of the interviewed teachers indicated that pupils from disadvantaged backgrounds were less likely to participate in science lessons or have resources at home to participate. Teacher interviews revealed mixed results about the type of teaching that was undertaken and the pressures schools were under. Given that the IPE focused on pupils within intervention schools, it remains unclear whether pupils within intervention schools and within control schools would have conveyed similar or dissimilar experiences around the COVID-19 pandemic such as the extent of undertaking science at home.

The Focus4TAPS trial was undertaken within schools in particular geographical regions, which may limit the extent of generalisation to wider regions and schools with any different circumstances. Other factors may impact generalisation, such as the profile of teachers/schools/pupils. The teacher interviews suggested (although not confirmed as universal across the sample) that some of those with subject specialism in science already had a confident, enthusiastic approach to teaching science. Future research could consider exactly what teacher gains are to be had for those with or without science as a subject specialism and what impact this may have on pupils.

Teachers selected Focus4TAPS resources from a range of materials (including lesson plans and activities) in order to deliver them within the Focus4TAPS programme. The trial did not assess whether particular lessons and tasks were more or less effective for students with particular characteristics. Future research would be necessary to help explore whether there are some elements of the intervention, which are better able to engage students with particular characteristics. Some of the interviews with teachers within intervention schools suggested that some scaffolding was required to make lessons accessible for those with EAL; it would be important to consider, which resources may be better able to be scaffolded. Additionally, it was not within the remit of the trial to explore what had more of an impact; Was it the actual resources or teachers' own enthusiasm that impacted pupils? One interviewed teacher from an intervention school highlighted that she could not be sure whether her pupils were more engaged because of her own enthusiasm or because of the teaching materials themselves. This would be an important future direction to look into, to help consider whether teachers from particular backgrounds need more support.

The Focus4TAPS trial originally intended to apply baseline questionnaires and follow-up questionnaires. However, undertaking baseline surveying on recommencement at the start of the 2020/2021 academic year was considered to be unfeasibly burdensome for schools and pupils, and was removed from the design. The applied design was only able to undertake surveying towards the end of the 2020/2021 academic year. Future research and trials would ideally measure changes over time through testing and surveying at multiple time points; this would help to measure what gains were made by control and intervention schools in terms of both attainment and attitudes and then test the difference between the control and the intervention schools. More generally, prior attainment in mathematics and reading (as recorded within the NPD) may not necessarily reflect prior attainment in science (which is not considered within Key Stage 1 outcomes within the NPD), and attainment may not necessarily reflect pupils' views such as their interest in science. Additionally, pupils could potentially learn and progress at different rates between the time of the prior attainment measures and the start of Year 5. Nevertheless, some of these general issues are common to any evaluation that follows the guidance to use baseline attainment measures sourced from the NPD (Education Endowment Foundation, 2018b).

The Focus4TAPS trial originally intended for tests and questionnaires to be undertaken under external invigilation, although the COVID-19 pandemic entailed that this was unfeasible; the tests and questionnaires were then administered by teachers following instructions/guidance. The trial was not able to test and survey all pupils: test data were available for 2,600 of 2,882 pupils (90.22% coverage); and questionnaire data were available for 2,501 of 2,882 pupils (86.78% coverage). Tests and questionnaires were provided as separate paper booklets for pupils to complete. Schools were provided with guidance explaining that, in order to reduce pupil fatigue, the test and survey could be administered a few hours apart, such as one first thing in the morning and one late morning or afternoon. If pupils were missing on the day, schools were encouraged to arrange for them to complete the test/survey as soon possible when they returned to school. It remains unclear why numbers/coverage varied across the tests and questionnaires. For example, it is possible that fatigue may have increased if the questionnaire was scheduled after the test. Additionally, some pupils may have chosen not to complete tests/questionnaires. It is also possible that scheduling tests and questionnaires on different days (if this occurred) may have complicated reaching all pupils due to absences on different days.

The Focus4TAPS programme and trial aimed to involve one main Year 5 teacher and their class per school. This involved complexity where schools with multiple classes and interested teachers had to specify one particular teacher/class as the focus for the trial. Future research and trials could instead consider encompassing any/all classes within schools. The randomised design was intended to mitigate against potential issues/risks following from class selection, given that schools were intended to specify a class/teacher prior to randomisation. However, in practice, some schools had circumstances where this was not possible; staff may have been changing from one academic year to the next, new staff may not have been recruited, assignments and workloads for the next academic year may not have been determined at that stage of the current academic year, and other circumstances may have been relevant. Future trials could potentially avoid these issues by applying restrictive recruitment criteria (i.e. schools would have to specify a teacher/class in advance or would not be allowed to participate), which would influence the resulting profile of schools/teachers (i.e. restrictive recruitment criteria would entail that schools with staffing changes would be unable to participate; only schools with the same teachers from academic year to academic year would be able to participate).

The Focus4TAPS trial undertook an array of planned analysis, which restricted exploratory analysis. This prevented any selective reporting of just significant findings. Nevertheless, this entailed that gaining greater understanding into (for example) what factors might influence pupils' science tests scores and views around science (other than any difference across pupils being assigned to an intervention group or a control group) would require further research such as through more elaborate quantitative predictive modelling.

The circumstances around the COVID-19 pandemic may complicate the extent of generalisation from recent research, such as the Focus4TAPS trial. Nevertheless, changed or changing circumstances also complicates the extent of generalisation from historical research; given that socio-cultural aspects of life, including education, continually change and develop. Some historical research may be less generalisable to contemporary circumstances given the COVID-19 pandemic.

Future research and publications

The Focus4TAPS programme was considered through an efficacy trial. Further evaluation could be undertaken through an effectiveness trial to consider impact at wider scales (involving more schools and pupils in more geographical areas), and with wider or different foci (such as involving more or different year groups of pupils). The teacher interviews conducted indicated that some schools had already begun to roll the programme out to younger year groups. Research could also look at the effects of Focus4TAPS teaching over a longer period. Given the actual disruptions in this trial because of school lockdown in January 2021, it may be helpful to ascertain what impact a trial would have over a 2-year period and without a lockdown period. Some teachers reported that they would need further training to continue rolling the Focus4TAPS programme while others appeared confident with the training they had. Future research could examine, for example, how much training teachers would need and whether a bank of professional development materials would suffice to help teachers refresh skills every year.

Further research could be undertaken to gain greater understanding into underlying aspects of science education, teaching/learning approaches, and also wider aspects of education and life that may be relevant (e.g. sense of school belonging, support for learning science, availability and encouragement of extra-curricular activities). Essentially, greater understanding into what associates with pupils' attainment and attitudes may benefit the delivery and refinement of interventions (including logic models and theories of change), as well as understanding wider inequalities such that they could be addressed. Further research may also help inform outcomes and approaches within evaluations; for example, pupils could be asked more specific questions about their confidence and interest in particular teaching approaches, and evaluations could consider particular groups of pupils.

Teaching/learning and wider circumstances may continue to change, such that understanding 'usual practice' and 'typical circumstances' may be complicated for the foreseeable future. For example, it remains unclear whether 'typical circumstances' in education would (or could) involve a return to circumstances as per a few years ago. From a wider perspective, this introduces (potentially positive) scope for reflecting on what any 'new normal' circumstances could or should involve. Essentially, it may be timely to reflect on which aspects of science education could or should be further developed or refined in order to mitigate inequalities.

Further publications will explore and involve reflections for practitioners on how to deliver on an evaluation during unprecedented times, more focused findings for teachers and including the programme developers' views about the successes and challenges of the programme during the pandemic, alongside implications for future programmes. Further publications will also explore associations between pupil attainment and attitudes/orientations towards science in more detail.

References

Archer, L. et al., 2015. “Science capital”: A conceptual, methodological, and empirical argument for extending Bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching,* 52(7), pp. 922-948.

Archer, L. et al., 2010. “Doing” science versus “being” a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education,* 94(4), pp. 617-639.

Archer, L. et al., 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), pp. 88-126.

Black, P. & Wiliam, D., 1998. Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice,* 5(1), pp. 7-74.

Black, P. & Wiliam, D., 2009. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability,* 21(1), pp. 5-31.

Bøe, M. V. & Henriksen, E. K., 2015. Expectancy-Value Perspectives on Choice of Science and Technology Education in Late-Modern Societies. In: E. K. Henriksen, J. Dillon & J. Ryder, eds. *Understanding Student Participation and Choice in Science and Technology Education.* Dordrecht: Springer, pp. 17-29.

Culliney, M. et al., 2021. *REACH Primary evaluation report,* London: Education Endowment Foundation.

Davies, D. et al., 2014. *Approaches to Science Assessment in English Primary Schools: Interim findings from the Teacher Assessment in Primary Science (TAPS) project,* Bristol: Primary Science Teaching Trust.

Davies, D. et al., 2017. Development and exemplification of a model for Teacher Assessment in Primary Science. *International Journal of Science Education,* 39(14), pp. 1869-1890.

Department for Education, 2014. *The national curriculum in England: Framework document,* London: Department for Education.

Department for Education, 2021. *Get Information about Schools,* London: Department for Education.

Earle, S. et al., 2016. *The Teacher Assessment in Primary Science (TAPS) school self-evaluation tool,* Bristol: Primary Science Teaching Trust.

Education Endowment Foundation, 2018a. *Improving Secondary Science: Guidance Report,* London: Education Endowment Foundation.

Education Endowment Foundation, 2018b. *Statistical analysis guidance for EEF evaluations [March 2018],* London: Education Endowment Foundation.

Education Endowment Foundation, 2019a. *Cost evaluation guidance for EEF evaluations,* London: Education Endowment Foundation.

Education Endowment Foundation, 2019b. *Implementation and process evaluation guidance for EEF evaluations, August 2019,* London: Education Endowment Foundation.

EngineeringUK, 2017. *Engineering UK 2017: The state of engineering,* London: EngineeringUK.

Freund, P. A. & Kasten, N., 2012. How Smart Do You Think You Are? A Meta-Analysis on the Validity of Self-Estimates of Cognitive Ability. *Psychological Bulletin,* 138(2), pp. 296-321.

Gatsby Charitable Foundation, 2017. *Good Practical Science,* London: Gatsby Charitable Foundation.

Godec, S., King, H. & Archer, L., 2017. *The Science Capital Teaching Approach: engaging students with science, promoting social justice,* London: University College London.

Goldsworthy, A., Watson, R. & Wood-Robinson, V., 2000. *Developing understanding in scientific enquiry (AKSIS),* Hatfield: Association for Science Education.

Hamlyn, R., Matthews, P. & Shanahan, M., 2017. *Young people's views on science education: Science Education Tracker Research Report February 2017,* London: Wellcome Trust.

Hanley, P., Slavin, R. & Elliott, L., 2015. *Thinking, Doing, Talking Science Evaluation report and executive summary,* London: Education Endowment Foundation.

Hattie, J. & Timperley, H., 2007. The Power of Feedback. *Review of Educational Research,* 77(1), pp. 81-112.

Humphrey, N. et al., 2016. *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook,* London: Education Endowment Foundation.

Institute of Physics, 2012. *The Importance of Physics to the UK Economy,* London: Institute of Physics.

Joshi, K. et al., 2021. *Development of a Year 5 Science Assessment for the EEF,* London: Education Endowment Foundation.

Kitmitto, S., González, R., Mezzanote, J. & Chen, Y., 2018. *Thinking, Doing, Talking Science: Evaluation report and executive summary,* London: Education Endowment Foundation.

Leonardi, S., Lamb, H., Howe, P. & Choudhoury, A., 2017. *'State of the nation' report on UK primary science education: Baseline research for the Wellcome Trust Primary Science Campaign,* London: Wellcome Trust.

Mansell, W. & James, M., 2009. *Assessment in schools: fit for purpose?,* London: Teaching and Learning Research Programme.

Martin, M., Mullis, I., Foy, P. & Hooper, M., 2016a. *TIMSS 2015 International Results in Science,* Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Center / International Association for the Evaluation of Educational Achievement.

Martin, M., Mullis, I. & Hooper, M. eds., 2016b. *Methods and Procedures in TIMSS 2015.* Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Center / International Association for the Evaluation of Educational Achievement.

McMahon, K. & Davies, D., 2003. Assessment for inquiry: supporting teaching and learning in primary science. *Science Education International,* 14(4), pp. 29-39.

Mujtaba, T. & 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), pp. 371-393.

Mujtaba, T., Sheldrake, R. & Reiss, M. J., 2020. *Chemistry for All: Reducing inequalities in chemistry aspirations and attitudes,* London: Royal Society of Chemistry.

Mujtaba, T., Sheldrake, R., Reiss, M. J. & Simon, S., 2018. Students' science attitudes, beliefs, and context: associations with science and chemistry aspirations. *International Journal of Science Education,* 40(6), pp. 644-667.

Mullis, I. et al., 2020. *TIMSS 2019 International Results in Mathematics and Science,* Chestnut Hill, Massachusetts: TIMSS & PIRLS International Study Center / International Association for the Evaluation of Educational Achievement.

Nag Chowdhuri, M., King, H. & Archer, L., 2021. *The Primary Science Capital Teaching Approach: Teacher Handbook,* London: University College London.

National Foundation for Educational Research, 2011. *Exploring young people's views on science education,* London: Wellcome Trust.

Nuffield Foundation, 1998. *SPACE research project,* Available at: https://www.stem.org.uk/resources/collection/3059/nuffield-primary-science [Accessed 10 January 2018].

Nuffield Foundation, 2012. *Developing policy, principles and practice in primary school science assessment,* London: Nuffield Foundation.

Nunes, T. et al., 2017. *Review of SES and Science Learning in Formal Educational Settings: A Report Prepared for the EEF and the Royal Society,* London: Education Endowment Foundation.

OECD, 2017. *PISA 2015 Technical Report,* Paris: OECD Publishing.

O'Mara, A., Marsh, H., Craven, R. & Debus, R., 2006. Do Self-Concept Interventions Make a Difference? A Synergistic Blend of Construct Validation and Meta-Analysis. *Educational Psychologist,* 41(3), pp. 181-206.

Pampaka, M. et al., 2021. *Increasing Competence and Confidence in Algebra and Multiplicative Structures (ICCAMS) evaluation report,* London: Education Endowment Foundation.

Primary Science Teaching Trust, 2016. *Teacher Assessment in Primary Science (TAPS).* [Online] Available at: https://pstt.org.uk/resources/curriculum-materials/assessment [Accessed 10 January 2018].

Regan, E. & DeWitt, J., 2015. Attitudes, Interest and Factors Influencing STEM Enrolment Behaviour: An Overview of Relevant Literature. In: E. K. Henriksen, J. Dillon & J. Ryder, eds. *Understanding Student Participation and Choice in Science and Technology Education.* Dordrecht: Springer, pp. 63-88.

Richardson, M., Abraham, C. & Bond, R., 2012. Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological Bulletin,* 138(2), pp. 353-387.

Rosenzweig, E. & Wigfield, A., 2016. STEM Motivation Interventions for Adolescents: A Promising Start, but Further to Go. *Educational Psychologist,* 51(2), pp. 146-163.

Royal Society, 2008. *Science and mathematics education, 14–19: A 'state of the nation' report on the participation and attainment of 14–19 year olds in science and mathematics in the UK, 1996–2007,* London: The Royal Society.

Royal Society, 2010. *Science and mathematics education, 5–14. A 'state of the nation' report,* 2010: The Royal Society.

Royal Society, 2014. *Vision for science and mathematics education,* London: The Royal Society.

Roy, P. et al., 2021. *Generation STEM Work Experience evaluation report,* London: Education Endowment Foundation.

Rubin, D., 1976. Inference and missing data. *Biometrika,* 63(3), pp. 581-592.

Shaw, B. et al., 2016. *Ethnicity, Gender and Social Mobility,* London: Social Mobility Commission.

Social Mobility Commission, 2016. *State of the Nation 2016: Social Mobility in Great Britain,* London: Her Majesty's Stationery Office.

Speckesser, S. et al., 2018. *Embedding Formative Assessment: Evaluation report and executive summary,* London: Education Endowment Foundation.

Toering, T. et al., 2012. Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). *International Journal of Sport and Exercise Psychology,* 10(1), pp. 24-38.

Tymms, P., 2004. Effect sizes in multilevel models. In: I. Schagen & K. Elliot, eds. *But What Does It Mean? The Use of Effect Sizes in Educational Research.* Slough: National Foundation for Educational Research, pp. 55-66.

Valentine, J., DuBois, D. & Cooper, H., 2004. The Relation Between Self-Beliefs and Academic Achievement: A Meta-Analytic Review. *Educational Psychologist,* 39(2), pp. 111-133.

Wiliam, D., 2011a. *Embedded formative assessment.* Bloomington: Solution Tree Press.

Wiliam, D., 2011b. What is assessment for learning?. *Studies in Educational Evaluation*, *37*(1), pp. 3-14.

Appendix A: EEF cost rating

Figure 5: 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

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%					4
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%			Adjustment for threats to internal validity [0]		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
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	Threat to internal validity?	Comments
Threat 1: Confounding	LOW	Well-designed RCT design, randomisation by evaluation team and comparable arms produced. The analysis controlled for pre-test measure. Greater proportion of science (STEM) specialism among nominated teachers from treatment schools. This was controlled for in additional analysis to produce consistent findings. Minimal impact from recommencement.
Threat 2: Concurrent Interventions	LOW	No evidence of concurrent interventions in either intervention or control groups. Schools implementing similar science programmes (TDTS/PSQM) or previously involved with Focus4TAPSexcluded from the trial.
Threat 3: Experimental effects	LOW	No reported experimental effects.
Threat 4: Implementation fidelity	MODERATE	Some adaptations were made to the intervention and its delivery as a result of COVID, but the IPE otherwise suggests relatively high fidelity to the intervention. Some teachers were trained twice because of the trial stopping and recommencing. Recommencement training was delivered online rather than in-person as originally intended, and some lessons were delivered remotely; this poses a moderate risk to internal validity.
Threat 5: Missing Data	LOW	The percentage of missing data was moderate, but this has already been resulted in the loss of a padlock. There was very marginal differential attrition (14.3% intervention, 11.3% control). Analyses accounting for missing data produced similar results.
Threat 6: Measurement of Outcomes	MODERATE	Primary outcome measurement was appropriate. However, due to unexpected Covid-related restrictions, tests were administered by teachers (who were not blind to allocation, and potentially invested in the outcome of the evaluation) and later marked independently.

Threat 7: Selective reporting	LOW	Analyses were pre-specified within published protocol and SAP. Trial registered on ISRCTN. Evaluation followed protocol closely. Additional analyses are stated as unplanned and exploratory; altogether contribute to thorough exploration of findings

- **Initial padlock score:** 4 padlocks; well designed two-arm cluster randomised trial; MDES of .199 at randomisation and 12.8% attrition experienced.
- **Reason for adjustment for threats to validity**: 0 padlocks lost. Two moderate threats to validity for the trial. First, assessments were administered by teachers and not independently as planned, though tests were marked independently by invigilators. Second, moderate threat to implementation fidelity with unplanned online delivery of training and reported incidences of teachers delivering sessions remotely. Some threats to validity captured in the initial padlocks related to loss to follow-up. Missing data is marginally differential, and whilst direction of potential bias could go either way, additional missing data analysis found consistent results. Overall no adjustment in padlocks.
- **Final padlock score:** 4 padlocks

Further appendices

Further appendices are available on the EEF website.

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