Education
Endowment
Foundation

Deeper Thinking

Pilot Report

February 2020

Andrew Boyle, Dr Hayley Limmer, Dr Kathy Seymour, Ben Smith (AlphaPlus Consultancy Ltd.) Professor Stephen Morris (Manchester Metropolitan University)

AlphaPlus

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus (formerly Impetus Trust) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

The pilot was funded by the Education Endowment Foundation (EEF) as part of a joint initiative with the Wellcome Trust which aims to generate new evidence about science teaching, with the particular aim of closing the science attainment and progression gap that exists between disadvantaged pupils and their more affluent peers.

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

AlphaPlus Consultancy Ltd. is one of the UK's leading educational consultancies drawing on the professional experience of a team whose work spans the public and private sectors, covering education, assessment and evaluation. We can draw upon a team of over 400 associates, each of whom brings many years' experience in education, covering both schools and post compulsory education, as well as roles in many of the supporting agencies and government departments. We are actively engaged with the latest educational developments, on both a strategic level, through our research and evaluation projects for UK government, and at a practical level, through the experience of our wider team in all aspects of educational development and change.

We have an international reputation for robust and meaningful evaluations, which support our clients to understand the impact on their stakeholders of policy change and/or interventions. We work in collaboration with our clients to ensure that evaluation findings are evidence based and useful. We ensure that our data collection, analysis and interpretation is undertaken with a solid understanding of the stakeholder context and within ethical guidelines.

Our team has expertise in assessment, evaluation and statistics, fieldwork, and evidence gathering – but we are also excellent communicators, speaking and writing clearly and accurately and giving messages that are to the point and free of jargon. We provide realistic and evidence-based reports for our research and evaluation clients.

Contact details

Andrew Boyle
Director of research
AlphaPlus Consultancy Ltd.

Andrew.boyle@alphaplus.co.uk

Executive summary

The project

The Deeper Thinking programme aims to improve outcomes in GCSE science by encouraging pupils to use a variety of metacognitive strategies including the Structure of Observed Learning Outcomes (SOLO) Taxonomy and concept-mapping and revision strategies. SOLO Taxonomy provides pupils with a five-stage, visual model that aims to change how they think about scientific understanding. It encourages learners to link scientific knowledge together and apply it to different contexts. Concept-mapping is also deployed and provides pupils with blank laminated hexagons to populate with scientific concepts before discussing how they link together. Alongside these strategies, Deeper Thinking also encourages the use of revision techniques, including practice testing and distributed practice.

Carmel Education Trust developed this approach, which specifically targets two types of GCSE science question: those that require an extended answer and those relating to the 'required' practicals. Schools involved receive four CPD sessions comprising one day of leadership training for senior leaders followed by three twilight training sessions for science teachers. Teachers are also provided with resource packs and an online portal to support delivery.

Twelve schools in the North East and North Yorkshire took part in this pilot, which ran from January to June 2019. They implemented the strategies with their Year 10 pupils in a condensed version of the programme (which would ordinarily be delivered for two years across Year 10 and 11).

The pilot was funded by the Education Endowment Foundation (EEF) as part of a joint initiative with the Wellcome Trust which aims to generate new evidence about science teaching, with the particular aim of closing the science attainment and progression gap that exists between disadvantaged pupils and their more affluent peers.

How was the pilot conducted?

The pilot aimed to examine whether Deeper Thinking could be summarised in a plausible, doable, testable, and meaningful Theory of Change, before exploring whether the programme may improve teacher practice and perceived pupil behaviour. The evaluation also aimed to discover whether the intervention is sufficiently different to current practice, alongside exploring how feasible and ready for trial the programme is.

A mixed-methods approach was taken to data collection combining evidence from interviews, observations, surveys, and administrative data, in addition to an analysis of selected GCSE science exam paper questions.

Overview of findings

Summary of pilot findings

Question	Finding	Comment
Is there evidence to support the Theory of Change?	Mixed	There is some evidence of positive changes to teaching practice and perceptions of an impact on pupils, but some teachers were already using similar techniques.
Was the approach feasible?	Mixed	The programme has low costs, is easy to set up, and does not require access to IT. Fidelity to the programme implementation and delivery may have been a cause for concern as a substantial proportion of teachers reported that they were not consistently using the SOLO Taxonomy placement and quick concept-mapping hexagons together. Some teachers reported that strategies may support lower-ability learners, but others reported that it was a challenge to apply the Deeper Thinking intervention to lower-attaining pupils and those lacking the skills/capacity to work independently.
Is the approach ready to be evaluated in a trial?	No	The delivery partner would need to adapt the Deeper Thinking classroom packs and training materials to better prepare teachers of lower-attaining and disruptive classes.

The findings were mixed. The pilot found that the Deeper Thinking intervention could be summarised as a distinct intervention, with a well-grounded, doable, testable, and meaningful Theory of Change. However, while there was some evidence of positive changes to teaching practice, there is only mixed evidence that the programme would improve GCSE science outcomes. Some teachers did report that the programme led to positive changes in their practice, such as identifying scientific misconceptions more quickly, implementing a greater focus on practicals and writing style, and fostering links between different scientific content. Some teachers also appeared to be using the techniques beyond the pupils involved in the pilot, while others reflected that the intervention may have supported pupils to communicate scientific knowledge and make their own links between concepts. Two of the case studies also suggested that Deeper Thinking may reduce teacher planning and marking workload (although this was not explicitly measured in the pilot).

However, the pilot also suggests that many teachers may already be using techniques which are similar to those deployed in Deeper Thinking. Some teachers also reported that it was a challenge to deliver the programme to, and to elicit outcomes from, lower-attaining pupils and those lacking the skills and capacity to work independently. While some teachers reported that certain strategies (such as quick concept-mapping) may support lower-attaining learners, others suggested that lower-ability pupils may find the programme more challenging. Given the mission of the EEF—and the aims of the funding round—to narrow the attainment gap, this may be a concern.

Although the programme has low costs and does not require access to IT, the pilot identified challenges around feasibility. There may have been evidence that a substantial proportion of teachers did not consistently deliver the programme as intended as teachers may have selected different strategies from the programme rather than delivering the suite of techniques together. The delivery team would argue that this still represented delivery as intended as they maintain that teachers could have delivered the strategies flexibly. Development may be required to define more clearly what teachers are expected to deliver. The training was rated highly by participants. Although some noted that it could have been condensed and suggested improvements—such as providing more training materials—more than three quarters of respondents found the training to be ‘helpful’ or ‘very helpful’.

The intervention is not ready to be evaluated in a trial as development would be required to ensure that lower-attaining pupils are able to access the intervention.

The pilot also examined whether it would be possible for an evaluation team to conduct a future trial of the approach. An analysis of AQA exam papers suggested that there are a sufficient number of relevant questions included in the sample papers analysed for a trial to measure the impact of the programme. However, the ability to monitor the use of similar activities in a control group of schools is anticipated to prove challenging as the Deeper Thinking intervention includes some activities that many teachers use already. Whilst the Deeper Thinking intervention is unique in that it combines several approaches, these do exist in isolation, therefore the apparent lack of difference to current practice could also make demonstrating an impact more challenging.

Introduction

Intervention

The description below refers to the intended implementation of this intervention. For practical reasons, the following changes and conditions applied to the current piloting of the intervention:

- The pilot was delivered by the Carmel Education Trust across 12 schools in the North East and North Yorkshire. This number of schools is comparable to other pilot trials and was considered sufficient to allow for some conclusions to be drawn across a variety of schools whilst keeping the resources required to deliver the pilot manageable.

- The pilot was a shortened version of what is intended to be a two-year intervention running from the beginning of Year 10 through to the GCSE exam period in Year 11. The pilot ran from January to June 2019 and the intervention was used with Year 10 pupils only.

1. Brief name: Deeper Thinking.

2. Why (rationale/theory)? Carmel Education Trust's research suggests that pupil performance in science examinations is limited by poor responses to extended answer questions (those worth typically four to six marks) and questions around experiments that took place in class (also known as 'required' practicals). Performance appears to be limited by the pupils' inability to respond correctly to the command words used in examinations. Some pupils, particularly the least able and those with lowest literacy (of which a high proportion are disadvantaged pupils) often do not attempt to answer extended answer questions. Those who attempt the questions often do not connect their knowledge together in a way that answers the question. Many teachers appear not to have suitable techniques, tools, or programmes to prepare the pupils well for these questions. Revision for examinations is often limited and occurs late in the GCSE programme.

Deeper Thinking aims to:

- improve pupil performance at GCSE by intervening over one to two years;
- develop pupils' understanding of the marking and question-setting process for GCSE science;
- develop a deeper understanding in pupils of the 'required' practicals; and
- improve pupil revision strategies.

3. Who (recipients)? Pupils in Years 10 and 11 receive the intervention as part of their GCSE science programme.

4. What (materials)? Teachers receive materials, including templates and guidance documentation, to support the delivery of Deeper Thinking approaches. These approaches aim to work together to improve:

1. performance in extended answer and analysis questions; and
2. reasoning around the 'required' practicals.

The approaches include the Structure of Observed Learning Outcomes (SOLO) Taxonomy, concept-mapping, and revision techniques.

The SOLO Taxonomy is a model that aims to change how pupils frame their scientific understanding. It categorises pupils' responses to a question, in this case a mock extended answer science exam question, according to five levels of sophistication ranging from no understanding to a complex understanding of how several pieces of information can be linked together and applied to different contexts. These are formally labelled as:

- prestructural—the question is not understood;
- unistructural—one piece of information is identified;
- multistructural—several pieces of information are identified but they are treated independently;
- relational—the different pieces of information are linked together; and

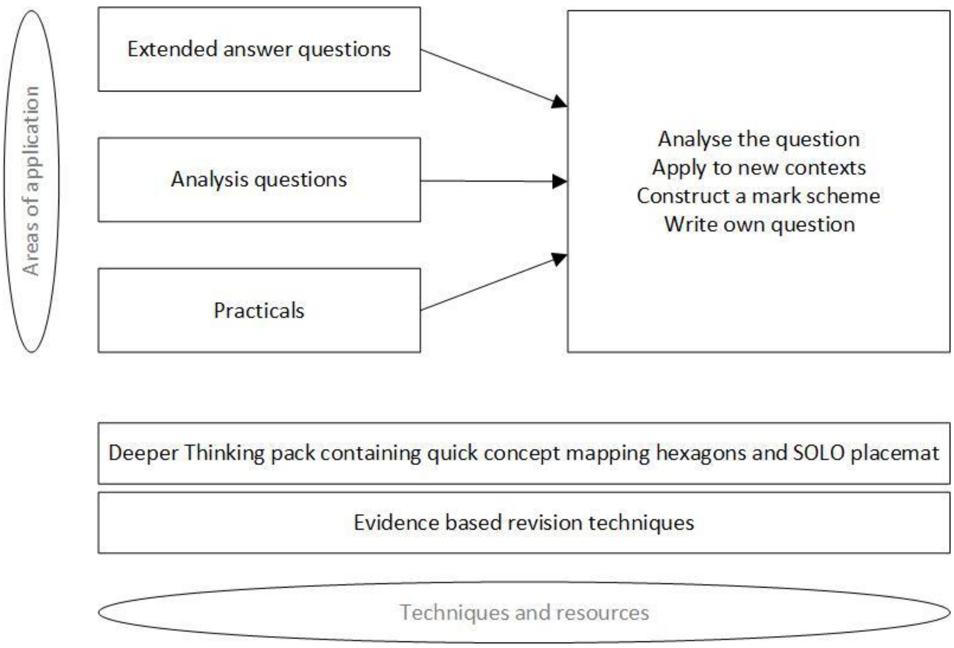
- extended abstract—being able to reconceptualise the information and apply it to a different question (Biggs and Collis, 1982).

Deeper Thinking encourages pupils to understand and then move up the SOLO levels. Although the SOLO Taxonomy is arguably the most visible part of the programme, Deeper Thinking involves the combination of the SOLO Taxonomy with:

- developing an understanding in pupils of command words used in examinations (for example, 'describe' and 'explain');
- quick concept-mapping—based on the use of blank laminated hexagon tiles that can be written on with dry-wipe pens and flexibly placed next to one another; pupils populate the tiles with scientific concepts, before discussing how they link together; and
- an approach to revision informed by Dunlosky et al.'s (2013) analysis of the utility of learning strategies.

This is summarised in **Figure 1**.

Figure 1: What is 'Deeper Thinking'?



Teachers are provided with a set of classroom packs (approximately 15) so pupils can work in pairs, with one pack per pair. Each classroom pack provides 22 blank laminated hexagons, three dry-wipe pens for populating the hexagons, a cloth to clean the hexagons, a rubric, and a SOLO Taxonomy placemat all held in a plastic folder (see Appendix 1).

The SOLO Taxonomy placemat summarises the SOLO levels in a language that is more pupil friendly than the academic terms used by Biggs and Collis (1982) whilst the rubric provides a template in which they can write notes that justify the level of their answer. The packs are typically used in relation to extended answer or practical mock exam questions.

5. What (procedures)? The programme involves training teachers to use the Deeper Thinking techniques within their science classes. In the intended full intervention, catch-up training is scheduled to take place in the second year of implementation to accommodate changes of staffing.

6. Who? Within each school there is a Deeper Thinking lead who coordinates the intervention and science teachers who deliver the intervention to pupils. A team of trainers from the Carmel Education Trust deliver the science teacher training.

7. How (mode of delivery)? The Deeper Thinking lead is trained in key elements of the approach through a one-day training event. All schools then receive three departmental twilight training sessions (two hours each) to be attended by all the science teachers taking part. Teachers are also able to access an online learning portal that provides further materials such as videos of exemplar lessons demonstrating the Deeper Thinking techniques. Teachers are also encouraged by Carmel to upload reflective paragraphs here to capture their learning.

In a typical Deeper Thinking session applied to an extended answer question, some variant on the following process typically takes place:

- Pupils are presented with an extended answer question from an exam paper.
- They rate themselves on the SOLO Taxonomy placemat diagram by placing their pen on the stage that best represents their current ability to answer the question.
- The teacher leads a short discussion on the wording of the exam question to reinforce what it is asking them to do and how it fits with the SOLO Taxonomy placemat and the hexagons.
- Pupils then engage in quick concept-mapping where they work in pairs to populate the blank hexagons with key words that relate to the question. Each pupil pair works with 22 hexagons.
- In the same pairs, the pupils then discuss links between the hexagons in relation to the extended answer.
- The teacher visits each group and has a discussion on how the hexagons link together; this involves asking pupils questions rather than giving them answers.
- There is some class-level discussion on the insight that is being generated. Sometimes pupils look at the concept maps of other groups and note down what they might be missing.
- Pupils then individually write an answer to the extended answer question.
- They rate themselves again on the SOLO Taxonomy placemat diagram by placing their pen on the stage that best represents their current ability to answer the question.
- The exercise concludes with some form of acknowledgement from the teacher and from the pupils themselves that they knew more than they thought and were able to answer a difficult exam question.

The extent to which the lesson focuses on the wording of the question, mark schemes, and alternative contexts is at the teacher’s discretion.

It is also recommended that a variant on this is applied to a science practical; this tends to be done in the lesson following the practical so the pupils can reflect on the activity rather than as the practical is taking place.

A further session where pupils explore evidence-based revision techniques informed by the learning strategies reviewed by Dunlosky et al. (2013) is also provided. This does not involve the hexagons but aims to increase pupils’ metacognitive awareness with respect to the activities they may be engaging in when revising for their science GCSE. During these sessions, the teacher leads a discussion on the different techniques that they use to revise for exams. The teacher then moves on to discuss the relative utility of the different learning strategies evaluated by Dunlosky et al. (2013). This process aims to flag to the learners that some of the techniques they may be relying on, such as highlighting and re-reading text, could be replaced with other learning strategies such as practice testing and interleaved practice.

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

9. When and how much (dosage)? Pupils receive the intervention at regular intervals (approximately once every two to four weeks) over the two-year period. The practicals are distributed throughout the course and practice of extended answer questions takes place at least half-termly.

10. Tailoring. Schools fit the programme around their usual work in class relating to the required practicals and practice examinations. The aspect of the intervention delivered (that is, the focus on extended answer questions or practicals) is partly dependent on the aspect of the curriculum being taught.

Background evidence

This pilot project was jointly funded by the Wellcome Trust and the EEF as part of the Improving Science Education round.

Each of the three key components of the Deeper Thinking intervention is based on research evidence, as discussed below.

The SOLO Taxonomy

Developed by Biggs and Collis (1982), SOLO Taxonomy breaks down stages of learning into five steps (prestructural, unistructural, multistructural, relational, and extended abstract). Pupils are taught these five steps and are encouraged to apply them to their scientific knowledge and to structure their responses through them. In addition, pupils are taught a series of connected metacognitive approaches, including concept-mapping and constructing success criteria to be used when appraising their work.

Previous research primarily applies the SOLO Taxonomy to undergraduate students rather than pupils of secondary school age (for example, Carew and Mitchell, 2002). This may partly reflect that the extended abstract level of the SOLO Taxonomy requires higher-level metacognitive skills that may be more appropriate for university students. Nevertheless, Minogue and Jones (2009) use the SOLO Taxonomy as a way of measuring the impact of a computer-based learning environment amongst middle school science students in the U.S., suggesting it can be a useful framework for working with this age group.

Revision skills

The use of evidence-based revision techniques is informed by the work of Dunlosky et al. (2013) who distinguished between ten different learning strategies. Whilst Dunlosky et al. (2013) did not focus primarily on revision techniques, the learning strategies that they evaluated can be closely aligned to the revision strategies applied by GCSE pupils:

- elaborative interrogation—explaining why a fact/concept is true;
- practice testing—self-testing or taking practice tests of to-be-learned materials;
- self-explanation—explaining how new information is related to already known information;
- summarisation—writing summaries of to-be-learned text;
- highlighting/underlining—marking potentially important portions of texts while reading;
- distributed practice—a schedule of practice that spreads out study activities over time;
- rereading—restudying text material again;
- key word memory techniques—using keywords to aid the recall of information;
- imagery for text—forming mental images of text materials; and
- interleaved practice—a schedule of practice that jumps between alternating topics.

Dunlosky et al. (2013) conclude that not all learning techniques are equally effective for improving pupil learning and assign them a utility rating ranging from low to high. Low utility techniques include summarisation, highlighting/underlining, keyword mnemonics, imagery for text, and rereading. Techniques rated as having moderate utility are elaborative interrogation, self-explanation, and interleaved practice. Practice testing and distributed practice are considered to have high utility. The Deeper Thinking intervention encourages both teachers and pupils to consider how they are revising and setting revision tasks and to adopt the higher utility approaches.

Quick concept-mapping

The use of the laminated hexagons is based on the rationale that quick concept maps are an effective learning tool. The 'quick' aspect comes from the way the Deeper Thinking hexagons are laminated, which allows for amends to be made quickly as the hexagons can be wiped clean and edited. Furthermore, the individual hexagons can be physically moved next to one another, which is relatively quick compared to a hand drawn concept map diagram that may require more effort to rework any changes. A meta-analysis (using international data) by Nesbit and Adescope (2006) suggests that the use of concept maps is associated with an increase in knowledge retention and transfer when compared to other activities such as reading text passages. The authors also conclude that concept maps may suit collaborative and cooperative learning as they make economical use of text that can be viewed by a small group. They cite van Boxtel et al. (2002) who found that concept-mapping was a successful collaborative learning tool amongst physics students aged 15 to 16, with learning outcomes positively related to the quality of student interaction and level of elaboration in discussions. Meanwhile, data from a review of 35 higher education medical papers suggests that concept maps provide four key functions: they promote meaningful learning, provide an additional resource for learning, enable instructors to provide feedback, and assess learning (Daley and Torre, 2010). Whilst the present study draws on a different sample demographic, there is no evidence to suggest that these findings would not apply to younger learners within a science context.

Existing evidence for the intervention

More generally, the Deeper Thinking intervention touches on several of the recommendations in the EEF's Improving Secondary Science guidance, namely:

- helping pupils direct their own learning by explicitly teaching pupils to monitor and evaluate their learning (via the SOLO Taxonomy placemat that requires the learners to reflect on what they think they know about a topic before and after the lesson);
- supporting pupils to retain and retrieve knowledge by providing opportunities for pupils to retrieve and elaborate on the knowledge they have previously learnt (via the quick concept-mapping hexagons);
- using practical work purposefully and as part of a learning sequence (via the sub-focus on required practicals and exam questions that apply to them);
- developing the language of science, including reference to the links between words; and
- using structured feedback to develop pupils' thinking, which requires finding out what pupils understand.

The intervention has been developed by the Carmel Education Trust over several years. In 2017/2018, a small pilot intervention of its current format was tested in six schools. This earlier evaluation, carried out by the Trust, included an exploratory analysis of pre- and post-tests which suggested a potentially positive impact on science outcomes.

Policy and practice context

Whilst the evaluation did not find any recent government policies relating to the SOLO Taxonomy, a 2014 government report by the National College for Teaching and Leadership[1] acknowledged that the SOLO Taxonomy in combination with the materials developed by Pam Hook were widely used in schools in New Zealand and when adopted in a selection of U.K. schools, was found to be a useful learning approach. Specifically:

'The materials and approach were found to be valuable both to assist teachers in their planning and pupils in their active engagement in learning activities, leading to enhanced peer and self-assessment skills being developed by the pupils and more differentiated approaches for personalising learning' (p. 13, Lilly et al., 2014).

The SOLO Taxonomy is a relatively well-known approach and is likely to be used in some schools already; this is anecdotally supported by some simple internet searches. Nevertheless, the evaluation is unable to ascertain how widespread its use is, particularly amongst GCSE science pupils.

The Deeper Thinking intervention also targets extended answer exam questions, some of which focus on practical activities. This has an accompanying policy context when one considers that in 2015 Ofqual introduced the requirement for science GCSEs to include mandatory practical activities (eight for single science subjects, 16 for combined science). These must cover the apparatus and techniques specified in selected subject content for biology, chemistry, and physics (Ofqual, 2015).

The reformed GCSEs in biology, chemistry, physics, and combined science have assessment objectives with the following weightings:

- 40% of the qualification relates to demonstrating knowledge and understanding of scientific ideas, techniques, and procedures;
- 40% to applying knowledge and understanding of scientific ideas, enquiry, techniques, and procedures; and

[1] In 2018 the majority of the functions of the National College for Teaching and Leadership were absorbed into the Department for Education.

- 20% in relation to analysing information and ideas to interpret and evaluate, make judgements and draw conclusions, and develop and improve experimental procedure (Ofqual, 2017).

Given this context, the evaluation considers the number of extended answer exam questions available on selected GCSE science papers to check that the intervention could realistically impact on pupil attainment outcomes as measured by GCSE grade.

The rationale for conducting the evaluation

The evaluation aims to determine whether the pilot demonstrates evidence of promise, whether it is feasible, and whether it is ready for trial.

The EEF is funding the Deeper Thinking programme partly on the basis that it deploys self-regulation and metacognitive strategies—approaches that derive strong support from evidence. As the EEF's Improving Secondary Science guidance report explains, several large correlational studies, alongside intervention studies, show strong links between self-regulation and attainment in science. Evidence also suggests that low prior-attainers benefit more than high prior-attainers, so explicitly teaching these strategies may help to close the attainment gap. The EEF also recognises that, more broadly across other subjects, several systematic reviews and meta-analyses exploring the effect of metacognition and self-regulation have consistently found large positive impacts. Several metacognition projects evaluated by the EEF have also found positive impacts.

The intervention is well developed in that it has already engaged in some piloting activities and the materials have been developed over several years and have been refined over this time.

Research questions

The current pilot and its evaluation had three key aims relating to the intervention. These focused on establishing (a) evidence to support the Theory of Change ('Is there evidence of the expected change happening?'), (b) to assess the feasibility ('Does delivery happen as intended?'), and (c) scalability ('Is it replicable and affordable?'). These aims are split into finer-grained research questions:

Evidence to support Theory of Change and evidence of promise

1. Can the Deeper Thinking intervention be summarised by a Theory of Change (ToC) model that is based on appropriate evidence?
2. Is the intervention theory of sufficient quality—plausible, doable, testable, and meaningful?
3. Are there any unintended consequences of the intervention? Are there anticipated indirect impacts on pupils that may not be picked up in the separate analysis of GCSE papers?
4. Is there evidence of change to Year 10 teachers' teaching practice because of the Deeper Thinking intervention?
5. Do the Year 10 teachers perceive there to be changes in pupil behaviour because of the Deeper Thinking intervention?
6. Is the intervention sufficiently innovative relative to the counter-factual?

Feasibility

7. Is the implementation happening as intended? Can facilitating and hindering factors be identified?
    - Is the programme attractive to schools?
    - How do teachers think the intervention can maximise transfer to the exam?
    - How can the CPD sessions maximise teacher engagement?
    - How does the implementation vary by the three areas—practicals, analysis, and extended answers?
8. What is the ideal way of delivering the Deeper Thinking intervention?
    - What do 'ideal' practitioners and schools look like?

Scalability and readiness for trial

9. How suitable is the Deeper Thinking intervention to progress to efficacy trial?
    - Is it clearly defined and scalable?
    - Will the intervention be practicable and attractive for schools beyond the immediate pilot group?

10. What might be the main design options for the efficacy trial?
    - What effect size is likely assuming only a direct impact on those GCSE paper questions that are specifically targeted? *
    - At what level should randomisation be performed in order to find the optimal balance between sample efficiency and avoid violation of causal assumptions (that is, ‘stable unit treatment value assumption’, SUTVA)? *
    - Given the proposed model and assumptions, what sample design yields relevant effect sizes? *
    - Will GCSE science be a suitably precise and targeted measure to pick up a measurable impact on student attainment? *
    - Are there secondary outcome measures that can be addressed by an efficacy trial?

The four design questions marked with an * are answered in the accompanying exam paper analysis research report.

Ethical review

All schools that were involved in this pilot project were asked to sign a memorandum of understanding (MoU) which was designed by the Carmel Education Trust and AlphaPlus Consultancy Ltd. It stated the roles, responsibilities, and obligations of being a pilot school and detailed the ethical conduct of the evaluation activities and the data-handling protocol. The MoU is included in Appendix 2.

The evaluation included an internal AlphaPlus ethics review, which is based on the British Educational Research Association (BERA) Ethical Guidelines.

Teacher Information sheets were provided by the evaluation team for use by the delivery team when explaining to the school what was involved in taking part (see Appendix 3). This document aimed to provide teachers with details of the evaluation and the ability to raise any concerns and potentially opt out at the earliest stage in the research when the MoU was being signed. This aimed to avoid a scenario where the school signed the MoU but had not gained informed consent from the teachers who would be taking part.

Consent for the online survey was obtained at the beginning of the survey. Consent for the case study visits was obtained during the organisation of the visits when liaising with the Deeper Thinking lead; at this point the teacher information sheets were re-issued to the schools. Teachers also provided verbal consent at the beginning of the interviews or observations when the aims of the data collection and storage of the data were reiterated. None of the participants raised any concerns about the ethical conduct of the research or about the basis of their participation during the evaluation. Eight of the twelve schools agreed to take part in a case study visit. Some visits included classroom observations. Teachers were the focus of the observations and the researchers did not speak directly to the pupils.

Project team

Table 1 details the team members who worked on this project and summarises their roles and responsibilities.

Table 1: Project team members

Name	Role in project	Organisation	Responsibilities
Andrew Boyle	Project director	AlphaPlus Consultancy Ltd	Research director for this project.
Dr Hayley Limmer	Project leader / lead on process evaluation	AlphaPlus Consultancy Ltd	Senior researcher / lead evaluator. Supported by Clair Metcalfe on the case study data collection.
Ben Smith	Assessment research lead	AlphaPlus Consultancy Ltd	Led the exam paper analysis.
Clare Dowland	Project manager	AlphaPlus Consultancy Ltd	Led project start-up and legal activities.
Dr Kathy Seymour	Survey administration and analysis	AlphaPlus Consultancy Ltd	Responsible for administering and analysing the survey and report writing.
Prof Steven Morris	Senior advisor on judgement regarding proceeding to full trial	Manchester Metropolitan University	Advisor on effect sizes for the exam paper analysis and trial recommendations.
David Bailey	Project director	Carmel Education Trust	Programme director for the delivery team.
Dr Dorothy Warren	Deeper Thinking Project Lead	Carmel Education Trust	Leading the Deeper Thinking activities and data collection.
Sarah McGee	Research and Development Project Officer	Carmel Education Trust	Schools coordinator.

Methods

Recruitment

Twelve schools were selected to take part in this pilot. The schools were selected from the North East and North Yorkshire regions as this allowed them to access training at the Carmel Education Trust.

The schools were selected by Carmel Education Trust. This was partly due to geographic considerations as schools had to attend central CPD session for leaders, which took place in Darlington (this made the journey time approximately one hour).

Within that area, the following criteria were applied:

- schools involved in pre-pilot work related to this project were excluded;
- schools involved in strategic school improvement projects where similar approaches were used were also excluded;
- a range of small, medium, and large schools were chosen; and
- schools with higher proportions of disadvantaged (FSM) pupils—quite prevalent in the geographic area—were approached first but schools with lower proportions of FSM/disadvantage were not excluded from the pilot.

Recruitment was through advertising through existing relationships and networks and included emails, attending cluster meetings, and discussions with strategic groups. A flyer was produced and disseminated via these routes. The pilot was also advertised on CET and EEF websites.

Several schools that agreed to take part were excluded at the recruitment stage because they were taking part in another EEF study. Beyond exclusion for involvement in a concurrent trial that had some overlap with the Deeper Thinking programme, there were no specific criteria for selection but schools with a high proportion of pupils eligible for free school meals were preferred, in line with EEF's aims.

The MoU was signed by the headteacher, executive principal, or manager and the staff member nominated as the Deeper Thinking lead, often the head of science. The intervention was intended to be delivered by all teachers of Year 10 science to all Year 10 science pupils. A list of all relevant Year 10 teachers' email addresses was provided by the school for the distribution of the teacher survey. One school signed the MoU and then withdrew from the pilot before the training or data collection commenced. It is not clear why it withdrew.

Data collection and analysis

The data collection comprised several activities including a Theory of Change workshop, interviews with teachers, the developer, and teachers who had used the intervention in an earlier trial, teacher surveys, an exam paper analysis, and case study visits. Each method had a specific purpose and linked directly to a research question, as summarised in Table 2, Table 3, and Table 4.

The case study data was collected by the project leader with assistance from another researcher, who was a former teacher so well placed to carry out classroom observations and speak to teachers. The focus group and interview schedules were designed to be as open-ended as possible to minimise researcher bias; similarly, the survey questions were designed to be non-leading and adhere to methodological best practice (such as avoiding double barrelling and ambiguous wording) to minimise bias.

Theory of Change workshop

The first workshop was attended by the Carmel Education Trust delivery team, a member of the EEF, and two members of the AlphaPlus evaluation team. The workshop was steered by the Evidence Based Practice Unit guidance (Wolpert, et al., 2016). During the workshop the participants examined:

- the target—who the intervention is for;

- the intervention—what exactly the intervention involves;
- change mechanisms—how the intervention leads to the outcomes that the developers expect; the developers were asked to bring a description of how each piece of academic literature relates to specific components of their intervention (plus reflections on the quality of the paper) to the workshop;
- outcomes—what the developers hope will happen as a result of the intervention, the speed of expected change, and what can be measured in the pilot evaluation; and
- moderators—a detailed discussion identifying the factors that will influence whether the intervention leads to the outcomes the developers identify.

The workshop built a Theory of Change model by using post-it notes attached to an A1 paper template of the Evidence Based Practice Model (see Appendix 4). Following the workshop, the evaluators produced a Theory of Change model that summarised the workshop content; this was shared with the EEF and the Carmel Education Trust. The Theory of Change model was revisited once the evaluation data was collected and analysed. The model was slightly revised—for example, moving an outcome from long-term to short-term—and then circulated to the delivery team for discussion; there was also the option of holding a second Theory of Change workshop if required. The relatively minor changes meant that a second face-to-face Theory of Change workshop did not take place as the changes were negotiated online.

Interviews with two teachers who had previously used the Deeper Thinking intervention and discussions with the developer

Discussions with the developer took place throughout the evaluation; several of these were face-to-face as they coincided with school CPD sessions. The evaluator also carried out telephone interviews with two teachers who had taken part in an earlier pilot (see Appendix 5 for a copy of the interview schedule). These teacher contacts were provided by the Carmel Education Trust. This insight was used to develop the teacher survey and case study research instruments, for example, the interviews with previous users of the Deeper Thinking intervention highlighted the need to ask teachers about different ability groups, both during the case study visits and in the teacher survey.

A pre and post teacher survey collected in January 2019 and June 2019

Following the Theory of Change workshop and interviews with previous users of the intervention, the teacher survey was developed. The survey was piloted with a former teacher for content and length.

The survey was distributed to 90 teachers in the first wave (January 2019) and 83 teachers in the second (June 2019). One school with seven teachers withdrew from the study prior to the training and has been excluded from pre and post pilot comparisons. Two reminder emails were sent to teachers who had not responded to the survey at one-week intervals during the three-week survey windows. Response rates are fully presented in Table 8 in the Participants section of this report, but 68 teachers responded in the first wave, 38 teachers in the second, and responses were received from all participating schools in both waves. There were 30 individual teachers identified as having responded in both waves. Ten teachers responded to the email for the second wave explaining that they had not taken part in the pilot; these were distributed across several schools. In addition, one of the case study conversations with a lead teacher reported that despite their intention of rolling it out across all Year 10 teachers, this was not possible, and they were the only teacher at their school who was able to take part. On this basis, we may not expect a high response rate, and the achieved response rate for the June survey was 52%, which is considered too low to make any generalisations beyond the sample who responded and, as such, the survey data is to be interpreted with caution. We have deliberately refrained from any statistical testing or inferences and report only descriptive means and percentages with the number of teachers per cell explicitly stated. The data is still considered to be valuable to evaluation but has been treated in a similar vein to the qualitative data.

The survey asked teachers to report on pupils of different ability levels separately. It defined 'high ability' as grade 7–9/A*–A; 'medium ability' as grade 6–4/B–C; and 'lower ability' as grade 3–U/D–U. The intention was to report on data at this level, however, the low response rate meant that this was not possible, so the data was pooled. See Appendices 6 and 7 for the survey questionnaires.

Teacher reflection data

The Carmel Education Trust provided the evaluators with teacher reflection data collected via their online portal. This data was coded in NVivo and common themes were identified.

Observations of the four teacher CPD sessions

An evaluator attended the leadership day and one of each of the three twilight CPD sessions offered to teachers taking part in the pilot. Following each session, the evaluator made notes on the content and how it was received by the attendees; this provided important context for the evaluation and allowed the evaluation team to fully understand the case study and survey comments on the CPD element of the Deeper Thinking Programme. This data informed the delivery of the case study interviews as the interviewer was aware of issues such as the sessions containing a substantial amount of theory and taking a full two hours at the end of the school day. The Carmel Education Trust also provided the Evaluation Team with a summary of the CPD evaluation forms collected at the end of each training session (see Appendix 8).

Case study visits or consultations with nine schools

Six case studies took place face-to-face and three took place over the phone as it was difficult for the schools to accommodate a visit during the exam period. The evaluation team aimed to arrange eight face-to-face visits, but several schools were unable to accommodate a face-to-face visit but could spare time for a telephone interview. Four visits included a lesson observation, a fifth scheduled observation was cancelled on the day due to clashes with a GCSE exam. In the other schools, teachers were unable to accommodate an observation session due to scheduling during the exam period.

The lesson observations took the form of a non-participant observation. They were semi-structured and followed a drafted checklist to guide the observation of activities whilst also allowing for unanticipated avenues to be documented (see Appendix 9). The case study visits were carried out by the Project Leader and a supporting researcher who is a former teacher. The data gathered during the case studies was compared to the Theory of Change model (University of Sheffield, n.d.). To try and minimise the impact of the observations on teacher behaviour, teachers were reassured that it was the intervention, not them, that was being examined and that any data collected was not to be shared with their school and that they and their school would not be identified in any reports resulting from the research. We requested classes of varying pupil ability across the schools and the observations covered high-, medium-, and lower-ability classes. The lesson observation notes were coded in NVivo.

During the visits the evaluators spoke to eight lead teachers and 11 teachers. The lead teacher interviews were semi-structured with the basic questions established and appropriate prompts listed in advance (Olsen, 2018). This included asking for reflections on how the teacher was managing the intervention at a school level, facilitating and hindering factors, and their thoughts on the CPD training (see Appendix 10 for the interview schedule). The Deeper Thinking lead was interviewed separately from the other teachers who were delivering the intervention to allow them the opportunity to discuss how delivery may vary across the different science teachers at their school.

The teacher consultations included a focus group, two paired interviews, and eight individual interviews. These followed a semi-structured schedule which asked teachers to reflect on CPD training and putting it into practice (see Appendix 11 for the focus group schedule). Focus groups were the preferred mode of data collection (as the interactions between participants encourage the production of more fully articulated accounts; Wilkinson, 1998) but often it was not practical for schools to facilitate this form of data collection; instead they offered individual interviews or interviews with pairs of teachers.

The teacher discussions were recorded, transcribed, and—alongside the teacher reflections and free text survey responses—analysed in NVivo. The analysis was carried out on a line-by-line basis with overarching headings of:

- reasons for taking part;
- set-up costs;
- training sessions;
- implementation;

- impact on teaching practice;
- counterfactual activities;
- impact on students; and
- transfer to the exam.

Over 500 references were coded across approximately 100 Nodes (that is, codes or topics) which broke these overarching categories down into more granular categories. The analysis was loosely based on a grounded theory approach whereby the data is broken down to a fine level and then built back up again into a higher-level summary.

An analysis of GCSE science exam paper questions

This is presented in an accompanying research paper titled 'GCSE Science as an Outcome Measure: Capacity of the Deeper Thinking Intervention to Improve GCSE Grades'. This may be found at: https://educationendowmentfoundation.org.uk/public/files/GCSE_Science_paper.pdf.

Relating the methods of data collection to the research questions

Tables 2 to 4 present the research questions broken down into those that refer to promise, feasibility, and scalability/ readiness for trial, respectively, and they indicate which data collection methods were intended to address which questions and what the intended output was in each case.

Table 2: Summary of the 'promise' research questions and corresponding methods of data collection

Research question	Data collection method	Output
Can the Deeper Thinking intervention be summarised by a Theory of Change (ToC) model that is based on appropriate evidence? Is the intervention theory of sufficient quality (Plausible, Doable, Testable and Meaningful)? Are there any unintended consequences of the intervention?	Theory of Change workshop part 1. Attended by the Carmel Education Trust delivery team, a member of the EEF, and AlphaPlus. Follow-up Theory of Change consultation post data collection to reflect on results.	Detailed logic model diagram and accompanying documentation. Document to be shared with developers. Revisit documentation with developers at the end of the project.
Is there evidence of change to Year 10 teachers' teaching practice because of the Deeper Thinking intervention?	Telephone interviews with 2–3 teachers who have used the Deeper Thinking intervention already.	Survey instruments developed with input from previous teachers and developers.
	Teacher survey data.	Survey measures of teacher reports of change in practice.
	Case study observations to corroborate teacher self-report data. The face-to-face data collection provides the opportunity to get more in-depth data than that afforded by the survey.	Teacher reflections on delivery and impact of the intervention in focus group/interview. The focus groups offer a chance to get insight into teacher consensus.
	The developer is already collecting teacher reflections; in the interest of being light touch, utilise these rather than a diary.	Concrete examples from teacher reflection data and observations.
Do the Year 10 teachers perceive there to be changes in pupil behaviour because of the Deeper Thinking intervention?	Indicators of pupil behaviour refined in the ToC workshop.	Indicators of interest.
	Teacher survey data.	Teacher responses to survey measures perceived changes in pupils.
	Case study visit, focus group. and interview data.	Teacher insight into perceived impact on pupils developed in more detail in the case study visit.
Is the intervention sufficiently innovative relative to the counter-factual?	Targeted review of similar interventions.	Judgment of points of similarity and difference.
	Teacher survey.	Teacher reflections relative to the previous year.

Table 3: Summary of the 'feasibility' research questions and corresponding methods of data collection

Research question	Data collection method	Output
Is the implementation happening as intended? Can facilitating and hindering factors be identified?	Teacher survey data. Case study visit focus group and interview data.	Identification of factors that facilitate and hinder teacher delivery. Teacher reflections of potential direct and indirect impact on pupils. List of what went well and where improvements can be made if proceed to efficacy trial.
Is the programme attractive to schools?	Case study interviews with the Lead Practitioner.	Insight into the perceived attractiveness of the intervention.
How do teachers think the intervention can maximise transfer to the exam?	Teacher focus group and interview.	Teacher insight, which may be too complicated for capture in survey.
How can the CPD sessions maximise teacher engagement?	CPD session observation. Teacher focus group and interview. CPD session feedback forms collected by Carmel. Teacher survey data.	Teacher reflections of CPD gathered in the focus groups and interviews. Final reflections from the end of the pilot based on a single survey question asking if they would change the CPD, considering their experience.
How does the implementation vary by the three areas: practicals, analysis, and extended answers?	Teacher focus group and interview. Teacher survey data.	Ratings and insight gathered against the three sub-areas.
What is the ideal way of delivering the Deeper Thinking intervention? What do 'ideal' practitioners and schools look like?	Case study interviews with the Lead Practitioner. Teacher focus groups.	Insight from individual teacher and school perspective.

Table 4: Summary of the 'scalability/ readiness for trial' research questions and corresponding methods of data collection.

Research question	Data collection/analysis method	Output
How suitable is the Deeper Thinking intervention to progress to efficacy trial? Is it clearly defined and scalable? Will the intervention be practicable and attractive for schools beyond the immediate pilot group?	Given all the above data, a judgement will be made by considering the evidence in relation to the success measures set out in the logic model. These will be reflected on during the second Theory of Change workshop. Targeted review of similar interventions to inform how innovative the programme is.	Final recommendations and report.
What might be the main design options for the efficacy trial?	Analyses of statistical power; suggestions as to sample size necessary for trial.	Information to inform the overall judgment of whether to progress to efficacy trial.
Will GCSE science be a suitably precise and targeted measure to pick up a measurable impact on pupil attainment?	Exam paper analysis and grade simulations. Examine the likely effect size or range of effect sizes that it is reasonable to expect given prior evidence and the analysis of GCSE papers.	Detailed discussion of what the outcome measure for an efficacy trail might look like.
Are there secondary outcome measures that can be addressed by an efficacy trial?	Theory of Change workshop. Case study visits could also provide insight into unanticipated secondary measures.	Additional measures that the efficacy evaluators may want to incorporate.

Timeline

The timeline for the delivery activities is presented in Table 5 and for the evaluation activities in Table 6.

Table 5: Timeline of delivery activities

Date	Activity
October–December 2018	School recruitment
January 2019	CPD leadership day
January–February 2019	Carmel collection of pre-test data (independent of the AlphaPlus evaluation)
February–April 2019	Three twilight sessions delivered to each school
February–June 2019	Teachers asked to practice the techniques
February–June 2019	Teachers asked to submit reflections into Carmel online portal
July 2019	Carmel collection of post-test data (independent of the AlphaPlus evaluation)

Table 6: Timeline of evaluation activities

Date	Activity
October 2018	Theory of Change workshop
November 2019	Telephone interviews with two teachers who took part in the earlier pilot
December 2019	Baseline teacher survey data collection
January–March 2019	Exam paper analysis
March 2019	Discussion with the developer on implementation so far
January–March 2019	Observations of CPD sessions offered to teachers
May 2019	Analysis of teacher reflections data
April–May 2019	Case study visits/consultations (n = 9)
June 2019	Post-intervention teacher survey
June 2019	Theory of Change revisited*
July–August 2019	Trial recommendations

*Evaluators offered the delivery partners the opportunity to attend a second workshop, but the relatively minor amendments meant that this was not needed as the discussion took place online.

Findings

Participants

As documented in the 'Recruitment' section of the report, 90 teachers across 12 schools took part in the evaluation. The number of teachers who received a survey varied considerably by school, ranging from 1 to 11. With a larger sample size or higher response rate (or both) the analysis may have taken into account the relative weight of schools but given the low response rate this was considered inappropriate.

The schools taking part were heterogeneous when considering the average percentage of children eligible for free school meals ('FSM pupils'), Progress 8 score, percentage of children whose first language is not English ('EAL pupils'), Ofsted ratings, and urban/rural status. To avoid disclosure of school identities, an aggregate summary of the individual schools relative to the national average is provided (see Table 7). The average percentage of FSM pupils across the schools was 19%, the Ofsted ratings varied from 'outstanding' to 'special measures', and there was also a mix of urban and rural schools.

Table 7: Selected school demographics

	National-level mean	Deeper Thinking Schools	
School-level (categorical)		n/N (missing)	Mean (%)
'Outstanding' or 'good' Ofsted rating	75%*	11 (1)	64%
Pupils whose first language is not English	16.9%	12 (0)	3.5%
Percentage of children eligible for free school meals	14.1%	12 (0)	18.6%
School-level (continuous)		n/N (missing)	Mean (SD)
Progress 8 score	-0.03	12 (0)	-0.35 (0.38)

Source: https://get-information-schools.service.gov.uk/
*Source: https://public.tableau.com/views/Dataview/Viewregionalperformanceovertime?:showVizHome=no

Two surveys were administered, one at the beginning of the pilot in January 2019 and one at the end in June 2019. In both the pre- and post-pilot surveys, at least one response was received from all participating schools. Table 8 shows the responses to each survey. Of the 38 respondents to the June survey, 30 also completed the January survey.

Table 8: Survey responses

Survey	No. of teachers invited to complete	No. of teachers who completed	No. of teachers who withdrew	Response rate (%) *
Pre pilot: January 2019	90	68	7	82%
Post pilot: June 2019	83	38	10	52%

*The response rate is calculated as the proportion of those who responded from the number who were invited to complete minus any withdrawals (withdrawals were instances where teachers contacted us to let us know that they had not participated in the pilot).

Evidence to support Theory of Change

Research question 1: Can the Deeper Thinking intervention be summarised by a Theory of Change (ToC) model that is based on appropriate evidence?

The inputs, activities, outputs, short- and long-term outcomes, and impact were explored using the template provided by the Evidence Based Practice Unit as documented in the data collection section of this report. This is based on the academic papers described in the 'Background evidence' section and what the delivery team expect will take place as the Deeper Thinking intervention progresses. The resulting Theory of Change model is presented in Figure 2.

The inputs to the model include teacher training, ongoing support, and the Deeper Thinking materials. Teachers delivering the intervention include the lead teachers, keen science teachers (classed as those who are typically extremely enthusiastic about pedagogical interventions), and regular science teachers. Pupils participating in the intervention may be grouped by science ability and behaviour, in terms of their ability to remain on-task in class.

Intervention activities are summarised in the ‘Intervention’ section of the Introduction chapter. To recap, the teachers participate in training to familiarise themselves with the Deeper Thinking approach. They then support pupils, typically working in pairs, to answer a science exam question whilst using the Deeper Thinking classroom packs that include a SOLO Taxonomy placemat and laminated hexagons. Pupils use the packs to identify, explore, and link their knowledge on the area in question. As the academic year progresses, pupils are expected to move on from analysing and answering exam questions to constructing mark schemes and writing their own questions. They are also encouraged to adopt revision strategies informed by Dunlosky et al.’s (2013) review of learning strategies.

The outputs of the model at teacher level are that the teachers provide scaffolding to pupils and gain a better understanding of pupils’ science misconceptions. Short-term teacher outcomes include:

- identifying gaps in their own knowledge;
- changing the way they set home revision to better align with the learning approaches identified as higher utility by Dunlosky et al. (2013);
- making more links in class;
- being able to address pupils’ science misconceptions in real time; and
- potentially reduced marking workload during hexagon-led lessons.[2]

Longer term outcomes include potential changes to schemes of learning to allow for more links and practicals, more synoptic assessment, a planned sequence of revision, and an overall increase in confidence as a science teacher. The overall impact anticipated at a teacher level is teaching that is more focused on linking ideas and grounded in revision best practice.

At a pupil level, the outputs fall into five areas:

- recognising existing knowledge and identifying gaps in knowledge;
- making more links between science content;
- developing their communications skills;
- becoming more sensitive to the language of assessment questions and the language needed to answer them; and
- learning the intervention steps so they can use the techniques unsupervised.

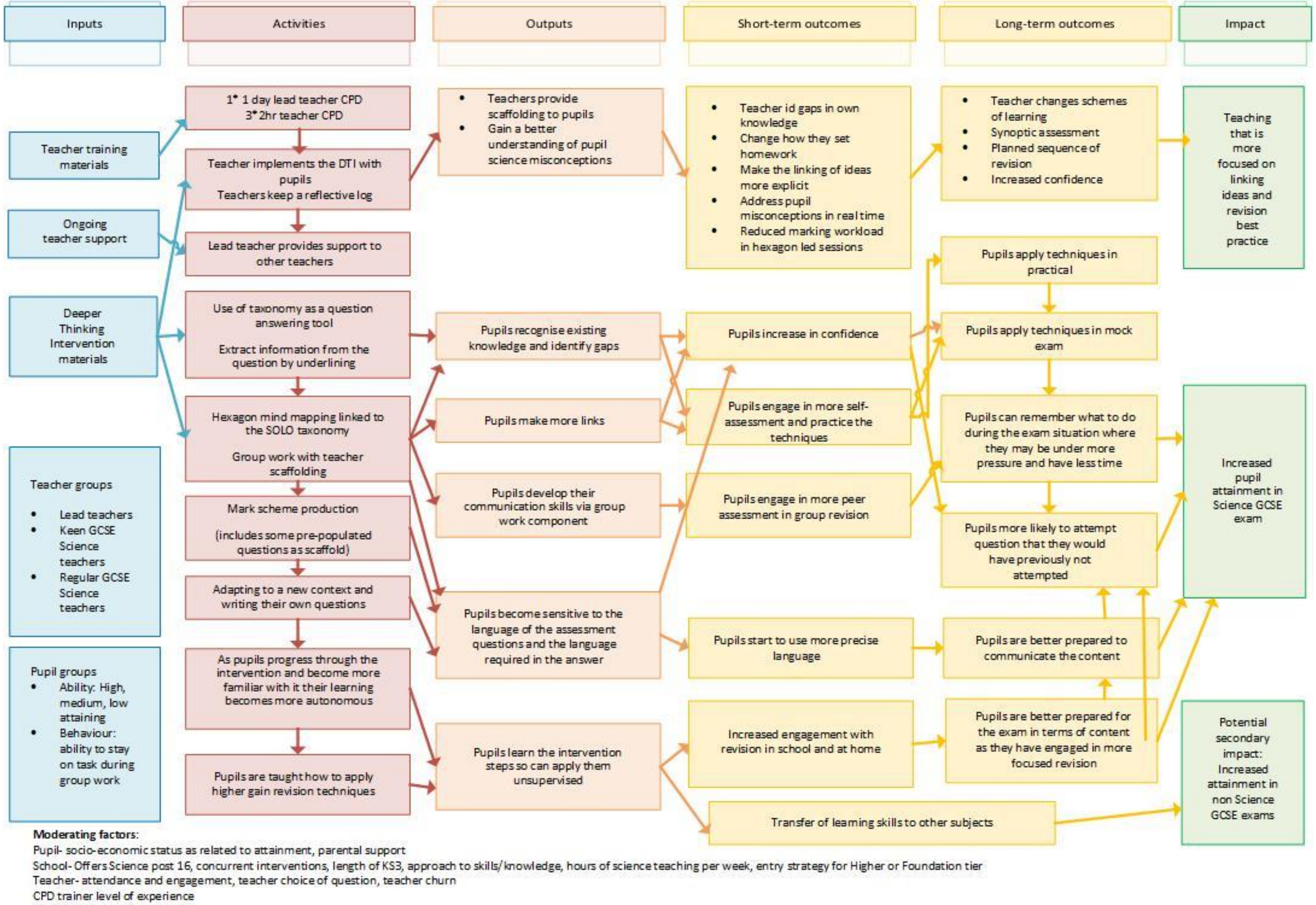
It is anticipated that these outputs lead to an increase in pupil confidence,[3] engagement in more self-assessment, more peer assessment, the use of more precise language, and an increased engagement with revision in school and at home. In the longer term, it is anticipated that pupils will be able to apply their techniques to practicals and mock exams and ultimately use them in the GCSE exam. Pupils will be better prepared to communicate the science content and be better prepared for the exam as they will have engaged in more focused revision. The ultimate impact is anticipated to be found in GCSE science attainment.

The pathways between the pupil-level outputs and outcomes are complex in the sense that they overlap. Whilst the Deeper Thinking intervention is linear in that pupils are only likely to write their own questions and mark schemes after they have engaged in analysing an existing question, the outcomes such as confidence can be linked to several outputs.

[2]These two outcomes were identified during the evaluation and were added to the model at the end; in a sense they were unanticipated outcomes.

[3]Pupil confidence was moved from a long-term to short-term outcome as it was flagged as an outcome by teachers during the relatively short evaluation window and appeared to come immediately after using the hexagons rather than having a lag as originally anticipated.

Figure 2: Final Deeper Thinking Theory of Change



Research question 2: Is the intervention theory of sufficient quality (plausible, doable, testable, and meaningful)?

Plausible

The theory behind the intervention as considered during the Theory of Change workshop is considered to be of sufficient quality because it is based on well-established academic literature (Biggs and Collis, 1982; Dunlosky et al., 2013) and is relatively simple and transparent. Nevertheless, some teachers may consider some of the techniques used by the Deeper Thinking intervention to be too close to their 'business as usual' activities, which may decrease plausibility for this group.

Doable

The human, material, and financial resources required to implement the intervention were all reported to be very low. Almost all the lead teachers consulted in the case study interviews reported no set-up or delivery costs and it was only when prompted that they reflected that the CPD time required during the training was a limited resource. As discussed later in the report, most schools reported that they would have liked to have received more classroom packs (ideally one per teacher).

Testable

The impact of the intervention can be tested in terms of attainment in GCSE science. The testing becomes more complex when one considers that the intervention combines three key areas: the SOLO Taxonomy, quick concept-mapping, and Dunloskian revision strategies. It is assumed that these are all to be used collectively and any impact is based on the combination of outputs and outcomes that they produce. If, however, one was to try and isolate the impact of these individual subcomponents then this may prove very challenging to disentangle. For example, using high value revision techniques may account for any improvement in attainment irrespective of whether they apply quick concept-mapping techniques in the exam, or vice versa. Furthermore, if one part of the intervention had a negative effect or there was a negative effect in combination, then this might lead to an overall finding of no effect even if other components were positive.

Meaningful

There is nothing to suggest that the theory that underpins the intervention is not meaningful.

Research question 3: Are there any unintended consequences of the intervention?

Case study discussions at two schools (case studies 1 and 9) suggested that using the Deeper Thinking techniques could reduce marking and planning. In the instance of marking, the teacher discussions in the classroom replaced the need to mark books.

Teacher 1: 'There is the other side to it that it reduces our marking, which is always a positive.'

Teacher 2: 'It's all writing stuff down. You might give them an extended piece of writing, but you can give them that on a worksheet and then actually you're quality marking that worksheet without going through all the rigmarole of books. So, it reduces our workload, it works well' (paired teacher interview, case study 1).

Another teacher reflected that they would previously spend considerable time planning their revision lessons but recently they spent five minutes picking out three difficult exam questions and asked the pupils to work through them with the hexagons (writing the links on post-it notes). They noted that some of the pupils commented that it was the most useful revision session that they had attended and another member of staff who dropped in commented on the high levels of pupil engagement in the room.

The evaluation has not identified any anticipated indirect impacts on pupils.

Research question 4: Is there evidence of change to Year 10 teachers' teaching practice because of the Deeper Thinking intervention?

Teacher interviews suggested that the intervention had encouraged teachers to check the assumptions that they held about pupil knowledge and identify where knowledge that they assumed was 'taken for granted' was not there. Others reported making more links between science topics and focusing more on practicals and writing style than they had done previously. The examples provided below are not necessarily identified more than once across the case studies but when considered as a collective could be interpreted as a small portfolio of evidence of 'promise'. The case study data also suggested that teachers widely used the Deeper Thinking techniques, or selected components of it, with groups of pupils who were beyond the scope of the pilot, namely those in other year groups and studying non-science subjects.

SOLO Taxonomy

A recurring theme throughout the case study discussions, observations and the teacher reflection data was that the Deeper Thinking intervention was potentially helping teachers to identify pupils' science misconceptions as a result of the method of delivery. During the delivery of the intervention, teachers typically ask pupils to work in pairs to populate[4] and discuss links between the hexagons. As this is taking place, the teacher works their way around the classroom discussing the outputs with the pupils. The observations suggest that these were high quality discussions that allowed the teachers to identify and address misconceptions in real time.

Teacher: 'And it allows you to have the conversations with the kids that you don't normally have. So, you can see, and you can address their misconceptions very, very quickly from those conversations that you're having' (teacher interview, case study 1).

'I spoke to groups and was able to pick out any misconceptions by challenging certain links of why students had linked some hexagons when perhaps the science was incorrect. For example, one group had placed "cations" next to "anode", so I was able to unpick a misunderstanding in the scientific literacy and therefore the fundamental understanding of electrolysis' (teacher reflective note).

'[The intervention] allows misconceptions to be identified very early and quickly corrected by simply moving a hexagon' (teacher survey response).

If the pupils were completing a comparable task writing key words and links in their workbooks, this would potentially require the teacher to mark this work outside the lesson, which would in turn result in a lag in feedback. The in-class discussions potentially allow for common misconceptions to be identified in real time and allow for class level feedback to be applied easily. However, there is some evidence of a linked, unintended consequence in that teacher discussions can take considerable time during which some pupils become disengaged. Thus, there sometimes appears to be a trade-off between in-depth discussions that can identify misconceptions and the time this takes during which disengagement can occur.

The SOLO placemat asks pupils to rate their current knowledge; one teacher reflected that the intervention had made them rethink their assumptions of pupil knowledge surrounding the science practicals.

'We presume that pupils have prior knowledge of things … Then we're going to use this and we're going to do this in the practical. But actually, that's not the case. And sometimes we're coming in to a practical from a place that they already don't understand. And it's [the Deeper Thinking pilot] allowed the team, I think, to really look at their practice and look at how we're delivering practicals and what the pitfalls are. And I think that we can bring that into our teaching, not only using the hexagons but when we're doing the practicals for the first time to try and come in from a standpoint that the pupils have a much clearer understanding of what they're doing. And we're not presuming that they have so much prior knowledge than we do already' (lead teacher interview, case study 1).

Quick concept maps

Whilst some teachers reported that they already used techniques that encouraged pupils to think about the links between key words, others reported that the Deeper Thinking pilot had encouraged teachers in their department to be

[4] In some instances, such as with lower-attaining groups, teachers may provide prepopulated hexagons.

more link- rather than fact-orientated (both were identified across different focus groups but did not feature in every single one).

'[Before taking part in the Deeper Thinking pilot, the teacher was already] asking students to explain connections between key words / concepts' (teacher survey response).

'I think people are thinking more about the linking, because we came from a course that was very much fact-orientated' (lead teacher interview, case study 2).

Every case study revealed that teachers were also using some of the Deeper Thinking techniques—particularly the hexagon concept-mapping—with groups of pupils beyond the Year 10 science pupils selected for the pilot. This included pupils in sixth form, Years 8, 9, and 11, and took place across different subjects, for example, one teacher reported using it with their Health and Social Care pupils.

'I've taken it through into sixth form, and the sixth formers, they quite like it, because they write their own exam questions based on how they've put the (quick concept-mapping) cards together. So, they can see all the links and then they've made up questions that they think an examiner could think of, realising how many links there are ... it's really quite cross-curricular, which is quite nice' (Teacher, case study 5).

'I completed a hexagon lesson with Y9 top set' (Teacher reflective note).

This suggests that many teachers considered there to be an inherent value in the quick concept-mapping component of the intervention as there was no onus on them to do this. One school (case study 9) reported that Deeper Thinking techniques were being built into standard departmental practice during a department redesign as the teachers considered the hexagon mapping techniques to be such effective revision and recap tools.

Analysing exam questions and mark schemes

One teacher reported that, following the training, they had mapped out a practical using the hexagon techniques as part of their lesson planning (teachers are not required to use the hexagons themselves in this way; this particular teacher had chosen to apply the technique to their personal work). This activity had made them realise how many ways the exam board could ask pupils about the practical. They reflected that in future they would put the practicals at the centre of their teaching.

'I had over 50 hexagons just from one question, one required practical, and I think that was quite a shock to my class that there's only one required practical in this topic but actually they could ask me about 50 different things just based around this one practical ... so it was quite useful for me as well, because I haven't really thought about kind of making the practical the centre of the module and teaching everything based on that practical and that made me sort of think, that's probably what we need to be changing in our school so we're focusing on how it links to this required practical rather than anything else' (lead teacher interview, case study 2).

Another teacher reported that the Deeper Thinking intervention had made them realise that their pupils needed longer to plan and structure extended-answer questions. Teachers in another case study suggested that the Deeper Thinking intervention had encouraged them to focus their teaching on writing and exam techniques beyond what they were doing before, particularly with lower-attaining groups.

'Even if I'm not using the hexagons, I'll get them to focus on the key words, and try to put them into sentences, and just get them to write for longer and encourage them … I've started to put that into my lessons a bit more, and I've realised the benefit, in the fact that they're going to have to write more in the exams. So, it's got me focused more on writing longer answers for pupils' (teacher interview, case study 3).

Evidence based revision techniques

Given the pilot was focused on Year 10 pupils, there was little qualitative evidence that teachers were applying the Deeper Thinking intervention component that focuses on encouraging pupils to use practice testing and distributed practice learning techniques that are defined by Dunlosky et al. (2013) as high value (see Background evidence section for further details).

As part of the pre and post survey, teachers were asked to rate the usefulness (or expected usefulness where they were aware of the technique but did not use it themselves) of a range of learning techniques examined by Dunlosky et al. (2013) and applied to GCSE science revision. They rated each on a scale of one to ten, where one corresponded to 'not at all useful' and ten indicated that it was 'extremely useful'. As the Deeper Thinking intervention training covered the relative value of different revision techniques, it was hypothesised that the post-pilot teacher ratings would be better aligned to the Dunlosky et al. (2013) ratings than their pre-pilot counterparts. However, the survey data suggests that there was no notable change in the mean usefulness rating applied to the learning techniques rated in the pre- and post-intervention surveys (see Appendix 11, Table 18).

A further point for consideration is that the Deeper Thinking intervention to some extent assumes that pupils will be proficient at the more effective revision techniques once they are explained, yet nudging behaviour in this way may require more nuanced training.

Research question 5: Do the Year 10 teachers perceive there to be changes in pupil behaviour because of the Deeper Thinking intervention?

The Theory of Change workshop identified three broad areas where we may expect to see evidence of teacher perceptions of changes in pupil behaviour:

1. identifying and linking scientific knowledge;
2. communicating scientific knowledge; and
3. engaging in GCSE science.

Identifying and linking scientific knowledge

In their written comments in the survey, a small number of teachers reported noticing an impact on pupils' abilities to recognise their existing science knowledge. The qualitative case study data suggests that some pupils initially struggled to identify their knowledge and whilst this was not exclusive to the lower-attaining groups, it was more prevalent amongst them. One teacher reflected that the intervention provided lower-attaining learners with a 'crutch':

'I would say it's going to be the ones that are down the bottom end of the group just because it gives them a little bit more, almost a crutch I would say, something that they can lean on in terms of, "Oh I don't know how to answer this", and if they manage to calm themselves down today—"Well let's just stop and think about what I do know". Whereas the ones that know what they're doing and they're confident—the higher end ones—they just fly through it anyway' (teacher interview, case study 6).

A reluctance amongst some learners to write answers when they are unsure their response is correct was reported as a widespread challenge, which made the laminated hexagons and ability to quickly erase and amend mistakes appealing to pupils.

The survey data suggests that there was some overall improvement in the ability of pupils to identify gaps in their knowledge. Of the 30 teachers who responded to both the January and June surveys, 23% agreed or strongly agreed that at least one group of their learners could do this pre-pilot; this increased to 34% post-pilot (see Appendix 11, Table 14). However, given the low response rate to the survey, this may simply be a maturation effect unrelated to the intervention.

Teacher reflections, case study discussions, and survey comments suggested that many teachers perceived the hexagon quick concept-mapping aspect of the Deeper Thinking intervention had helped some of their pupils to make links between different elements of their scientific knowledge.

'Definitely an impact. Yes, I think being able to show them that just learning a lot of facts isn't the way to pass the exam any more is probably helping, certainly the lower sets, so, like, "I don't have to know everything, I just have to

understand how it links together." And the technique is the same for answering kind of any type of question' (teacher interview, case study 2).

'The hexagons were brilliant at helping students build links between the three sciences and between individual topics within each one' (teacher survey response).

This is supported by the survey data that shows an increase in link-making ability post-pilot. The percentage of teachers agreeing or strongly agreeing that their Year 10 science pupils, if asked, could make links between different pieces of science knowledge increased from 16% to 34% (based on the 30 teachers who responded to both surveys; see Appendix 11, Table 14). This is to be expected as this was an aim of the intervention. However, one may expect such a change to have taken place under 'normal' conditions so the absence of a control group limits the inferences that can be drawn from any pre- and post-pilot comparisons.

Communicating scientific knowledge

The Theory of Change suggests that improvements in pupils' abilities to communicate in groups may be an output of taking part in the intervention. This is partly supported by the case study data; for example, one teacher reflected that pupils engaged in quite heated debates about the placement of their hexagons. Classroom observations and reflection data also noted that pupils were generally engaging well in small groups (usually pairs) discussing their hexagons. There were also several qualitative survey responses from teachers suggesting that they considered the intervention to promote good peer to peer discussions.

'Children get into quite in-depth arguments about why they think that should be there and why they think that's there. And, because they're being pushed and challenged by other members of their group, and by the staff that's in the room as well, they get more out of it, so at the end they feel more confident' (teacher paired interview, case study 3).

'Facilitates excellent peer-peer discussions when groups want to place the hexagons differently' (teacher survey response).

Nevertheless, the survey did not find any notable increase (over the duration of the pilot) in the percentage of teachers who agreed or strongly agreed that their Year 10 science pupils overall communicate well when carrying out group work.

The most notable change between the pre- and post-pilot surveys in relation to the extent to which their pupils were able to communicate scientific knowledge broadly was an increase of 14 percentage points (representing an increase from 17% in the pre-pilot survey to 31% in the post-pilot survey) in the proportion of the 30 teachers who responded to both surveys, agreeing or strongly agreeing that their pupils were able to understand the language they need to use when answering exam questions (see Appendix 11, Table 15). The case study data provided by several teachers reported that the pilot had improved their pupils' ability to understand the wording of exam questions and the language needed to answer them.

'It's [improved their] technique for writing down the answers and also realising that they can read the questions and get quite a bit of information from the questions and that they are capable themselves of actually doing some of this. So, I think I wouldn't say that by reading these exams—because bearing in mind I've only really concentrated on exam technique—I wouldn't say that they have suddenly got more knowledge. But I think they've certainly got a clearer idea of what they've got to write in the questions' (teacher interview, case study 4).

'Students are reading questions more carefully to look for information and key words' (teacher survey response).

Case study 3 provided one of the few examples of where the technique had been used with practicals and the teacher considered that the exercise had allowed the pupils to develop a better understanding of the vocabulary needed to answer practical questions and the need to cite the equipment terminology. Only eight practicals are required over the GCSE period so it is understandable that they would be relatively infrequent during the pilot period.

'The children have all understood the practicals a bit more as well. So, we've done the practicals, they've listed the equipment they've used and said how they've used it. That's helped them understand it and get a grasp of the six-

mark questions, on how to write the method, and things like that. So, they have really enjoyed it' (teacher paired interview, case study 3)

The data also suggests that the structure of some pupils' responses was improving as a result of the pilot; several teachers described how pupils used the hexagons to consider the order of their answers. In some schools the pupils were also considered able to produce longer answers more quickly than they would have done previously, although it is not clear whether this is due to practicing exam questions or due to employing the Deeper Thinking techniques.

'Yes. They [the pupils] definitely say that they've ... that it makes answering the questions easier when they can see how they can kind of put the information out in the order that they need to write about then they can write the question because they can see the order it goes in. It's helped them kind of visualise the links and things, so that's been good' (teacher interview, case study 2).

'The students can write a lot longer, they've got a lot more stamina, when it comes to writing. Before, especially with the lower sets, one or two lines maybe, for a six-mark question. Now, they'll quite happily write six or seven sentences, and they're getting faster at doing it' (teacher interview, case study 3).

Engagement in GCSE science

Triangulating the survey and case study data presents a mixed picture on pupil engagement in science lessons that used the Deeper Thinking techniques. Some teachers considered pupils to be more engaged in such lessons. For example, teachers consulted in case study 7 also reported that pupils remained on task and did not become disengaged. They also provided examples of their pupils asking to use the hexagons in a lesson where they were not being used. All the teachers consulted in the case studies perceived most of their pupils as enjoying using the Deeper Thinking techniques; for some pupils this was reported to be partly because they found it useful.

'Behaviour for learning was much better with my Year 10 class during this lesson' (teacher reflection data).

'Students are more positive, and they can see immediate improvement, so they have stuck in with it' (teacher reflection data).

'I have gone around and asked them [the pupils] …"Do you find this useful?" And the ones that you wouldn't expect ... I've got a couple of girls in there who are ... either in on report or in trouble or they're excluded, but they're, "Oh yes, really useful"' (teacher interview, case study 7).

'The pupils enjoyed the kinaesthetic nature of the task and trying out a new teaching resource' (teacher reflection data).

On the other hand, some teachers reflected that it was difficult to use the Deeper Thinking techniques with pupils who exhibited more challenging behaviour:

'With my triple class, they're more than happy to do it. They'll do as they're told; they find it quite useful. But with my low ability, it's now time that they can chat about other things rather than the task, so that side of things makes it more difficult and it's much easier for me to just hold two words up at the front and say, "How do these link together?", rather than you working in groups to try and sort it out because that's quite difficult with the behaviour that's in there. So, swings and roundabouts. If there's a couple of children off, then I can get more out of it. But it doesn't lend itself to classes with more challenging behaviour' (teacher interview, case study 5).

This was corroborated by lessons observations where the evaluator noted that learners did not always stay on-task if they were given a long time to work through the quick concept-mapping. Teachers are faced with a balancing act whereby they work their way around the different groups discussing what they have produced, but whilst they are doing this, the rest of the class is trusted to be discussing their hexagons in pairs. Whilst this did tend to occur with well-behaved classes, there were instances, particularly in some of the classroom observations, where engagement was intermittent.

The case study data suggests that some pupils may have become more confident in their science ability as the intervention encourages them to recognise their knowledge:

‘Some of them have definitely said they’ve enjoyed doing it because it’s given them an idea about how it can move things around and gives them a bit more confidence in the fact that they can just write something down. And go, “Oh yes, I know that!”, and they can write it down and go, “Oh actually I know a bit more than I think.” So, if nothing else it’s been a bit of a confidence boost for them’ (teacher interview, case study 6).

For others, confidence was associated with attempting to answer an extended-answer question when they would have previously left it blank. For example, case study 2 highlighted that some pupils with strong perfectionist tendencies may refuse to answer difficult questions for fear of making a mistake. The concept-mapping allows them to generate a structure, which in turn makes it easier to start answering a question. Having the confidence to attempt questions was also reported by a small number of survey respondents.

‘So, there's a couple of girls in my Year 10 class that are quite unwilling to begin a task. You know the kinds that don't want to write anything in case they mess it up in the first sentence. So, they particularly are more willing to have a go at a question now because they've had a chance to kind of map it out’ (teacher interview, case study 2).

‘Improved confidence with answering new contexts and attempting longer answers to get at least some marks rather than leaving blank’ (teacher survey response).

The pre- and post-pilot survey data shows that there has been a slight increase in the proportion of ‘agree’ and ‘strongly agree’ responses given against the statements referring to pupils’ engagement in continual self-assessment of their science knowledge and their confidence in their own science ability—although both represent increases from very low figures in the pre-pilot survey (see Appendix 11, Table 16).

Research question 6: Is the intervention sufficiently innovative relative to the counterfactual?

Targeted review of similar interventions

The Deeper Thinking intervention combines quick concept-mapping, the SOLO Taxonomy, and evidence-based revision techniques. Whilst the evaluators have not found a comparable intervention that brings these three elements together, there are teaching resources that draw upon them separately.

Online searches using ‘SOLO Taxonomy + Intervention + GCSE’ and ‘Evidence-based revision + Intervention + GCSE’ yielded a wide variety of results, many of which were not relevant. ‘GCSE’ was then removed from the filter to broaden the return results and these were manually sifted to identify those that could be applied to GCSE pupils. The resources that applied to analysing exam questions and mark schemes were identified by the lead evaluator as they previously worked for an awarding organisation.

SOLO Taxonomy

HookED (2017) provides a variety of free resources for teachers to introduce the SOLO Taxonomy as a model of learning. There are over 40 downloadable packs and posters that include concept-mapping hexagon templates and a SOLO Taxonomy functioning knowledge rubric template similar to those used in the Deeper Thinking intervention. These are not specifically tailored to GCSE science exam questions but list the key command words alongside the stages of the SOLO Taxonomy.

Quick concept maps

In addition to printable hexagon templates, HookED also provides a SOLO Taxonomy hexagons app which allows pupils to engage in virtual concept-mapping (Hook Education Ltd, 2019). Concept-mapping tools are widely available, and teachers and pupils potentially have access to a large number of online resources (Educators Technology, 2018). Whilst these are not interventions per se, they offer some of the functionality of the Deeper Thinking approach. Furthermore, they may offer pupils the opportunity to efficiently capture their hexagon representations, which some teachers found to be lacking with the use of the laminated hexagons because they did not necessarily provide pupils with a record of their learning, unless they captured it in some way, for example, by taking a photo. Nevertheless, introducing a virtual element to the quick concept-mapping may reduce usage within the current pilot due to a lack of computers in science laboratories and potential barriers in terms of log-in details.

Analysing exam questions and mark schemes

The Deeper Thinking intervention includes the analysis of exam question wording and encourages pupils to learn and apply the language used by examiners. This, in isolation, is not specifically innovative when you consider that teachers can freely access resources to help them do this on the awarding organisation websites; for example, AQA provides teachers with a subject-specific vocabulary glossary and a list of command words as part of its teaching resources for science trilogy (AQA, 2019).

Developing science vocabulary

Teachit (2019), in partnership with the Oxford University Press, have produced a pack of classroom activities for science teachers to promote ways to build scientific vocabulary amongst pupils; this includes using word banks, making links between key terms, exploring etymology and morphology, using talk to widen vocabulary, avoiding common mistakes and misunderstandings, and understanding the vocabulary for exams/assessments. Independent tasks include providing pupils with command words from the exam board and, based upon the tasks set in class, asking pupils to write their own questions using these words. There are also sections on extended writing and experimental write-up and mark schemes. Whilst these resources do not reference the SOLO Taxonomy, they do encourage pupils to construct concept maps.

Evidence-based revision techniques

There is some evidence that interventions and resources are available to teachers within the area of evidence-based revision techniques; for example, The Learning Scientists provides free materials for teachers and pupils based around six strategies for effective learning: spaced practice, retrieval practice, elaboration, interleaving, using concrete examples to understand abstract ideas, and dual coding—that is, combining words and visuals. Each strategy has a variety of teaching resources including an accompanying poster and PowerPoint presentation which references supportive academic literature (The Learning Scientists, 2019).

The Durrington and Sandringham Research Schools have designed and delivered a three-day training package on evidence-based learning strategies focusing on the findings of Dunlosky et al. (2013) (ResearchSchool.Org, 2019).

In summary, the evaluation found that comparable resources are available to teachers to deliver the individual components of the Deeper Thinking intervention. However, they do not bring all the components together in a single intervention. Furthermore, there is a lack of research evidence evaluating the impact of these resources either separately or in combination with one another.

Any future trialling may also wish to consider the extent to which teachers in schools that are considering taking part are engaging in one or more of these activities already.

Evidence of teachers already engaging in similar activities

Case study discussions and the teacher survey suggest that many teachers would ordinarily be using some of the Deeper Thinking techniques. This is supported by the post-pilot teacher survey which shows that 39% of teachers (who responded to the survey question) reported that before taking part in the Deeper Thinking pilot they were engaging in similar activities with their Year 10 science pupils (see Appendix 11, Table 11). These activities were broadly aligned with concept-mapping, linking techniques, and analysing exam questions and mark schemes. The SOLO Taxonomy aspect was not specifically referenced except by one teacher in the survey and three teachers during case study visits.

'I already used some hexagons in starters before topics and as revision. I used the language of SOLO in all my classes, and I had been using some other SOLO techniques like describe maps and sequencing maps' (teacher survey response).

Several teachers reported they would be using mind maps in lessons but noted that they found the hexagons an improvement on this as they were more flexible. Pupils could easily amend the laminated hexagons with the drywipe pens provided; the hexagons could also be moved as the pupils saw fit making it easy to add or remove links between the individual pieces of knowledge.

‘They [the students] just rub it out; they’re on laminated card so if they feel a little bit, “Oh oh”, they scrub it out and start again’ (lead teacher, case study 6).

‘Thinking about how to do long answer questions, we’ve got templates and you can give kids key words and stuff. You do a mind map. I’m a mind map person, so I just get them to throw it all on a mind map. But the benefit of the hexagons are you can move them around, so it’s not fixed. So, “Oh, it doesn’t belong there” or “You don’t want that.” Whereas with a mind map, you’ll have a massive cross through it—you might have to start again. So, you’ve got the flexibility of moving the words around on the hexagon’ (teacher, case study 3).

Making links between the content of the mind maps was also embedded into some existing practice. For example, one teacher reported that they would ordinarily carry out mind mapping on A3 paper, which, alongside identifying key words, involved a discussion of how they linked together. The teachers considered their ‘business as usual’ activities to have the additional advantage of providing pupils with something concrete that they could take away with them, which is not the case with the Deeper Thinking technique. They also reported that using the SOLO Taxonomy placemats was too time-consuming and restricted their ability to fit in enough pupil feedback at the end of the lesson (something they reported was possible with their simpler version).

Teachers in one case study school discussed the use of scaffolding sheets that specifically targeted extended answer questions and presented the question, key words, and elements of the mark scheme on a single sheet of paper. Many teachers also noted that they would normally work with pupils to analyse question wording to help them learn how to break down exam questions. For example, one teacher described how they previously spent time talking pupils though how they (the teacher) would approach answering a specific exam question including all the considerations they would make along the way. This was also reported by some teachers who completed the June teacher survey who were asked to describe what similar activities they were doing already. For example one teacher reported:

‘Already use exam questions on a regular basis (including those for practical activities). Often break them down step by step for the pupils and use of model answers/mark schemes to help them with their learning. Have also asked pupils to come up with mark schemes for questions I have given them. (Nothing new)’ (teacher survey response).

The pre- and post-pilot surveys also provide some insight into the extent to which teachers were already engaging in some of the teaching activities encouraged by the Deeper Thinking programme (albeit caveated with caution regarding the small sample of only 30 teachers who responded to both surveys, which limits the extent to which this data can be generalised beyond the respondent population). Table 12 (Appendix 12) shows that 43% (n = 13) of teachers reported asking pupils to analyse exam or assignment questions at least once a week prior to the pilot and 50% (n = 15) reported doing the same post-pilot. Meanwhile, Table 13 (Appendix 12) indicates that 33% (n = 10) of teachers who completed both surveys were discussing science mark schemes at least once a week before the Deeper Thinking pilot and this increased to 57% (n = 17) post-pilot. The evaluators acknowledge that the absence of a control group limits any inferences that can be drawn from pre- and post-pilot data. It may be that teachers would ordinarily increase activities relating to mark schemes as pupils progress through Year 10, so we might be reporting ‘business as usual’ rather than an impact of the pilot. One could also speculate that February to May is a relatively short period within which to detect a change in activities that are exam orientated in year one of a two-year course.

Feasibility

Research question 7: Is the implementation happening as intended? Can facilitating and hindering factors be identified?

The intervention would ordinarily run across two academic years with the second year focused on practising the techniques learnt in the first year. The revision techniques are also expected to become more central in the second year as the final exams approach. The pilot was condensed between late January (when the leadership CPD training took place) and June (when the post-intervention survey took place). There was a four-month window from February to May within which the science teachers were trained in how to apply the Deeper Thinking Techniques and put them into practice.

As summarised in **Figure 1**, each of the three areas (practicals, analysis, and extended answer questions) has four sub-activities (analysing questions, applying to new contexts, constructing mark schemes, and writing their own questions) and implementing all of these has proved to be beyond the scope of the pilot.

Data from the post-pilot survey shows that 100% of teachers who completed the survey applied the Deeper Thinking techniques to extended-answer questions, 84% applied them to practical work, and 76% applied them to analysis questions at least once during the pilot (see Appendix 11, Table 19). Analysing the exam questions was by far the most frequently used technique; all but two survey respondents (95% of respondents) reported that they had used this at least once in relation to extended-answer questions, while 82% had done so in terms of practical work and 76% for analysis questions. In contrast, constructing mark schemes and writing their own questions were less frequent activities; the former was used at least once by between 34% and 50% of respondents across the three areas, while the latter was used by between 13% and 47% across the three areas.

This raises some concerns regarding the overall level of teacher engagement with the intervention particularly when one considers the teachers who did not respond to the survey. The case study visits suggested that the volume of activities for the pilot was too great to fully implement and would need a full two-year trial to be implemented properly. When asked, the teachers reported the limited time-frame as the reason for having not yet tried or fully implemented a technique. They noted that the sessions linked to practicals were linked to timetabling of these activities, which did not necessarily overlap with the pilot. Similarly, the revision techniques and mark schemes did not feature in any of the sessions observed.

The case study discussions suggested that some teachers were using the concept-mapping hexagons but not the SOLO Taxonomy placemat or, where they do use both, the pupils do not fully grasp the SOLO Taxonomy component. Furthermore, nearly half of the respondents to the June survey (n = 17) reported that they never used the SOLO Taxonomy placemat and the concept-mapping hexagons together or did so less than half the time, which suggests that the intervention was not always implemented as intended (see Appendix 11, Table 20). This highlights that there is potential for dissonance between the theory presented in the teacher training and how it is applied in practice by the teachers.

Facilitating factors

Several facilitating factors were identified across the evaluation activities. These tended to refer to aspects of the intervention's design and how it was presented and perceived within the school, rather than to the specific conditions, circumstances, and contexts that had facilitated implementation.

Simplicity of using the hexagons

The developer was of the opinion that the simplicity of the intervention and absence of any requirement for technological resources made it relatively straightforward to implement. When asked to reflect on which of the aspects of the Deeper Thinking intervention had gone well, many teachers (at least ten), without prompting, identified the use of the hexagons as a concept-mapping tool.

'Use of hexagons … a novel and different approach to breaking down and analysing long answer/practical exam questions. Have also used them with my Year 9 lower-ability class to try to link chemical structures and their properties together' (teacher survey response).

Teacher buy-in

The case study visits suggested that teachers were enthusiastic about the intervention and were committed to applying it where time allowed. However, not all teachers were consulted during these visits and the low survey response rate suggests that there may be a sizable proportion of the teachers who did not engage with the intervention—or at least had not engaged enough to partake in the evaluation.

How the intervention was 'sold' to the pupils

The Deeper Thinking lead at one school described how they had carefully explained the Deeper Thinking intervention to the pupils and 'marketed' it as something that was solely for the pupils' benefit.

Providing sufficient time

Allowing teachers the time to attend training, reflect upon and plan how they will implement the training, and to share good practice with each other was also a key facilitator to successful implementation.

Barriers

Barriers were identified at the pupil level, and, more generally, in terms of practicalities and the scheduling of practical sessions, as noted elsewhere.

Pupil-level barrier: content knowledge

Teachers across the case studies, reflections, and qualitative survey responses identified a lack of initial scientific knowledge among pupils as a barrier to applying the content to other contexts, mark schemes, and question-writing. Several teachers reflected that mark scheme production was a high-level skill that was beyond all but the highest-attaining pupils; they considered it more important to familiarise pupils with concept-mapping and making links.

'Overall the activity seemed to work well if the students had a good understanding already, but when the students were struggling with the basics, jumping to linking seemed to be a bit too much for them' (teacher reflection data).

Pupil-level barrier: science ability

The training sessions reassured teachers that the Deeper Thinking techniques had achieved very good results with lower-attaining pupils but acknowledged that it takes more practice and longer to achieve results among this group and suggested that prepopulated hexagons can be a useful tool in achieving this. Nevertheless, the evaluation data suggests that many teachers have found using the intervention with lower-attaining groups to be challenging.

'Don't think the sessions actually enabled teachers to use the hexagons to write mark schemes/own questions effectively, especially for lower end groups' (teacher survey response).

'I would say it [the Deeper Thinking intervention] was very good at being targeted at higher ability but not so good at giving ideas for the lower ability' (teacher interview, case study 5).

Pupil-level barrier: behaviour

Similarly, some teachers reflected that the Deeper Thinking techniques worked well for well-behaved groups but noted that the intervention needed further development for use with more disruptive classes. For example, one teacher reflected that they had tried the intervention with a lower-attaining Year 10 group with more challenging behaviour and found a substantial percentage of the class were unable to remain on task. The Deeper Thinking techniques suggest that pupils work in small groups to carry out quick concept-mapping. However, the teacher reported that this particular class (a lower-attaining group with challenging behaviour) used the time to avoid working. The teacher adapted the session by reducing the number of hexagons to two, holding them up, and asking the pupils to discuss the two cards and to make links between them. This was not suggested in the training so was something the teacher had developed.

Pupil-level barrier: understanding the language of SOLO

Some teachers reported that the language of the SOLO Taxonomy is not very pupil friendly. This could be interpreted as a barrier to implementation as it suggests that some pupils do not understand what they are being asked to do.

'It was interesting to find that, as a SOLO practitioner, the class basically confessed they didn't really know what I was talking about when I used SOLO terms and they never thought to ask! Bit of a wakeup' (teacher reflection data).

'I wouldn't say they understand what SOLO Taxonomy is and engage with it from that point of view, but in terms of the actual activity, they really like it' (teacher, case study 7).

'I did not find the SOLO Taxonomy sheet useful. The pupils didn't engage well with them and seemed to miss the point of using them. They seemed to rate their confidence of a topic according to where they thought I expected them to be, rather than reporting their confidence in a genuine manner' (teacher survey response).

Practical barrier: running out of time

A small number of teachers also reported that they found the Deeper Thinking techniques difficult to manage within the allocated lesson period. This was also evident in the classroom observations and may be more prevalent if teachers attempt to implement all of the aspects recommended rather than 'cherry picking' the bits they found to be most appropriate/useful. Future teacher training would benefit from offering guidance on how to condense or break down the intervention into smaller chunks.

'It took them a long time to write out the cards and then when I asked them to group them, if they had repeating words, they grouped these together. They found it much more difficult to group them. I wondered if it was the topic. We also ran out of time' (teacher reflection data).

Applying the techniques in an exam

Not all teachers considered the intervention to be transferable to the exam; there were several examples from the survey where teachers discussed how they or their pupils did not think the hexagons would be useful when answering exam papers.

'The majority of students did not see the value of the technique, and the students themselves questioned how it would help them in an exam' (teacher survey response).

'My higher-ability class didn't see the point in using the hexagons as "they can't use them in an exam". This class is very focused on outcome and struggled to understand the relevance and importance of the process. The class argued that they would rather do something that will help them when they are in the exam too' (teacher survey response).

How can the CPD sessions maximise teacher engagement?

Data from the Leadership Day, the three twilight training session evaluation forms, session observations, and the June survey suggests that the majority of teachers rated the CPD sessions highly on the day of the session (12 people attended the Leadership day, 78 attended the first twilight session, and 79 attended the second and third twilight sessions).

In the June teacher survey, 77% (n = 28) of respondents considered that the training they received was helpful or very helpful for delivering the intervention with their Year 10 pupils, and 17% (n = 6) rated it as unhelpful or very unhelpful (see Appendix 11, Table 17). This question was included in the survey as a sense check as teachers can be positive about training sessions at the point of delivery but can re-consider their ratings once they have implemented the intervention. The teachers were also asked to fill out the evaluation sheets whilst the trainer was still in the room, which may have led to some social desirability bias.

The intervention may maximise teacher engagement if it were to make the following changes:

Resources tailored to pupils of varying ability

Several teachers reported that the implementation was more challenging with lower-attaining science pupils and that they would have preferred to have received more tailored guidance for working with this group. This could include modelling a lesson with lower-attaining pupils as well as a higher-attaining class.

‘You have to take everything like that [the training demonstration] with a pinch of salt because it was a small class in a nice school with very high abilities, so it's entirely different than me doing it with some of my, you know, bigger classes that are less able than that’ (teacher interview, case study 7).

‘It would be good if it was modelled, I suppose, with a lower ability as well, to enable staff who were less experienced and less confident at using it, to be able to see how it could be implemented that way’ (teacher interview, case study 8).

Shorten and streamline the twilight training sessions

Several teachers reflected that the twilight sessions were too long and too research-heavy and would benefit from being streamlined to make them shorter and sharper and to better engage teachers; condensing the content to one hour was recommended. Two lead teachers suggested they could have delivered the training rather than having an external speaker. However, one teacher recommended that external trainers should be used as their input was more likely to be accepted than that of colleagues.

‘I think the useful bit of the [twilight] session—that we actually take and then use—was probably done in an hour, 45 minutes, rather than, you know, two hours’ (lead teacher interview, case study 7).

Minimise teachers’ CPD homework

Observation data also suggested that some teachers struggled to complete ‘homework’ set between the CPD twilight sessions, for example, reading academic papers or keeping reflective logs. A further trial may wish to condense these resources into the CPD sessions to avoid variation in compliance.

Provide session modelling for non-lead teachers

The Deeper Thinking leads were positive about observing the Deeper Thinking techniques in action; for example, one teacher reflected that whilst a video of the observation was available online, there was additional value in being there in person where you could discuss it with other teachers and the pupils.

‘Rather than watching a video or a PowerPoint, literally seeing him teaching it and using the skills in front of you, I thought that was really useful, really good’ (lead teacher interview, case study 3).

Whilst all schools were offered, and took part in, lesson-modelling with their pupils, not all teachers were available to attend the session, thus not all of the non-lead teachers were provided with observation sessions. There are videos available, but it is unclear how many teachers were actively using these resources. The survey suggests that all teachers would benefit from observing a live session for each component of the intervention so they could see how to bridge the theory and practice, particularly when the intervention moves from the extended-answer questions to mark schemes and practicals.

‘There was not enough time [in the training] given to the actual techniques to be employed in the classroom. There needed to be less time spent on the theory behind it. More example and model lessons [are] needed and observations or co-teaching with participating staff’ (teacher survey response).

Provide checklists, instructions, and templates

It was also reported that some teachers would find it useful to be provided with a checklist reflecting the preferred order of delivery. In a case study interview, one teacher reported that they had produced such a PowerPoint themselves, which they felt aided their delivery structure and clarity.

‘I had made myself, like, a PowerPoint which had the kind of steps. So I couldn't muddle up the steps when I was explaining it to them, so my explanation to them was a lot clearer which meant that they could follow each step and we could do one step at a time’ (teacher interview, case study 2).

This was also evident in the survey data where several teachers reported in their free text responses that they would have liked to have received reference materials in the training sessions. During the training sessions attended by the evaluators, teachers were not provided with paper copies of handouts summarising the session (instead they were directed to the online portal where the resources could be accessed). In a further trial, it is recommended that teachers are offered the option of paper copies of the training materials. One teacher reported that it would have been useful to have received some guidance resources for use when delivering Deeper Thinking sessions.

‘I feel that I would have benefited from handouts with written instructions and ideas on how to integrate the activities into lessons as this is something that I could refer back to after the session to reflect upon and guide me’ (teacher survey response).

Provide examples from alternative contexts

One lead teacher also suggested it would have been useful to have had more discussion beyond the concept-mapping component. Similarly, another case study teacher reported that it would be helpful to have a set of alternative contexts that they could apply to the current GCSE specifications. One stage of the programme is to encourage learners to think of alternative contexts that might be used in the exam, but clearly teachers are also required to provide examples if the pupils do not generate them themselves. Four free text survey responses also requested more examples of how to apply the intervention (in addition to five requests to observe live sessions).

‘[We need] examples because sometimes we’re genuinely racking our brains thinking, “How can we do that?” Some questions that lend themselves [are] dead easy to do in some of them; some don’t’ (teacher interview, case study 4).

Summary of suggested improvements to the CPD sessions

In summary, the data suggests that the CPD sessions could maximise teacher engagement by:

- providing resources tailored to pupils of varying ability, particularly low ability and disengaged contexts;
- shortening and streamlining the twilight training sessions;
- minimising the volume of teachers’ CPD homework;
- providing session-modelling for non-lead leachers;
- providing checklists, instructions, and templates for use when planning and delivering Deeper Thinking classes; and
- providing examples from alternative contexts.

Research question 8: What is the ideal way of delivering the Deeper Thinking intervention?

The case study data suggested that the intervention would ideally be delivered over a longer period (as opposed to the condensed pilot period), which would allow for all the activities to take place. Teachers would be provided with one class pack of resources each and this may be adapted to include a physical take-away for learners.

Schedule the activities at a slower pace

The evaluation flagged that the intervention, whilst relatively simple, has many sub-components and activities that many teachers considered needed to be implemented over a much longer period of time than the six-month pilot. Several teachers suggested that it would be useful to introduce the Deeper Thinking techniques to pupils before Year 10 so they could focus on applying them to content in Year 10.

Provide the schools with one hexagon pack per teacher

The data from case studies one, three, and nine suggests that ideally the intervention would distribute a class hexagon pack to each teacher taking part, not just one or two per department. Teachers flagged that as science classes run concurrently, so packs need to be booked in advance, which may reduce use. In some cases, teachers prepared their own packs using their own teaching budgets but in a future trial ideally teachers would not need to do this.

Teacher 1: 'I'm using it quite a bit with Year 10s and with Year 9s. I think the biggest thing … initially was resources of only having the one pack. So, I think we've taken the initiative in that I've created a pack of just my classroom so that I can be using them all the time. And I think you've done the same?'

Teacher 2: 'I did that as well.'

Teacher 1: 'And I think that's the look at each classroom is going to have a set eventually which will make it much more easier because you can then adapt your teaching and use spontaneously as well' (teacher interview, case study 1).

'We really have managed. We book them through the technician, if you want them. So, she does quite a bit of negotiating—"Oh, no, they've already been booked." "Alright, I'll get onto the next lesson." But we've made it a bookable resource. I don't think it's caused much hassle' (teacher interview, case study 3).

Consider how pupils can capture the hexagon work so they have a physical takeaway

There was some suggestion that the pupils lacked a physical copy of their hexagons once the session had ended; it may be useful if pupils could be encouraged to take a photograph or similar to create a record of the activity. This could possibly be most useful for the lower-attaining groups who some teachers considered would forget the content and techniques of the Deeper Thinking lessons.

What do 'ideal' practitioners and schools look like?

Based on the classroom observations and observations of the teacher CPD sessions, the following ideal practitioner characteristics were identified:

- the ability to reinforce resilience and confidence alongside delivery, to be able to balance addressing misconceptions and recognising and celebrating achievement at the end of the session;
- enthusiasm and the ability to carefully consider how they 'sell' the Deeper Thinking techniques to the pupils;
- the ability to control pupils lacking the skills or capacity to work independently during group work and ensure they remain on task whilst simultaneously having discussions with other groups of pupils; and
- have very good questioning skills as modelled by the Carmel project director.

Regarding the latter: the questioning techniques employed rarely gave pupils any additional information but encouraged them to retrieve and link the information themselves. This may be partly because the model lessons took place with high-attaining pupils who may be well suited to this technique. The inclusion of more reference to these soft skills in the training could be a useful addition to encourage teachers to mirror this behaviour.

At a school level, ideal schools:

- provide time for teacher CPD training;
- provide/protect time for Deeper Thinking practitioners to share learning from the delivery of the intervention with one another, rather than with the delivery partner via the online portal; and
- encourage the use of the techniques in earlier years so the techniques are already familiar to pupils prior to GCSE study

Scalability and readiness for trial

Research question 9: How suitable is the Deeper Thinking intervention to progress to efficacy trial?

Is it clearly defined and scalable?

As identified earlier in the report, the Theory of Change underpinning the intervention is relatively complex with many causal pathways. The survey data suggests that the extent to which teachers implement the whole intervention, not just selected parts of it, may need to be carefully considered in a future trial. It was assumed that the teachers would use the SOLO Taxonomy placemat and the quick concept-mapping hexagons together as advocated in the training and guidance. However, in practice, the quick concept-mapping hexagons were often used in isolation, which suggests that any future trial should monitor teacher usage of the micro components of the intervention or alternatively focus on a single aspect, potentially the hexagon tool. The mechanisms by which this monitoring could be done reliably and robustly are not clear since there are no 'naturally occurring' records of usage as we might expect from, say, use of an online resource which could automatically record usage levels therefore monitoring would rely on teachers' self-reports which could cause issues around accuracy and demands on teachers' time.

The evaluation also found that many teachers would have preferred more structured lesson-planning resources that they could use when planning and delivering Deeper Thinking lessons. For these teachers, more concrete definitions beyond those discussed in the CPD sessions would have been useful.

Provided the delivery partner can scale up the delivery of the training and ensure that lesson demonstrations are still provided, there are no anticipated issues around scalability of the intervention itself. There was some evidence from the case study visits that the teachers preferred the demonstrations delivered by the senior members of the Carmel team over the third-party trainers. As with all scaled-up interventions, careful consideration should be given to who is recruited to deliver nationwide training if the intervention progresses to further trial.

It is also recommended that a hexagon pack is provided to each teacher taking part, which would have some implications for the mass production of such packs in view of an efficacy trial.

Will the intervention be practicable and attractive for schools beyond the immediate pilot group?

As highlighted earlier in the report, the intervention is considered relatively low demand in terms of its practical implementation. There are no technological requirements in terms of access to IT or any equipment required beyond the hexagons, which makes the intervention relatively simple to implement. As highlighted in the facilitating factors section, most teachers were happy to use the hexagons, which could be a positive factor in terms of stability as it indicates a willingness to use at least some components of the intervention.

There is no demand on the learner's family as no parental participation is required and the pupils do not need access to computers at home or mobile devices to take part.

The main practical limitation raised by teachers during the evaluation is being able to fit the intervention into their lessons when the curriculum content is demanding.

When asked during the case study visits, many teachers reported they would recommend the Deeper Thinking approach to other teachers. For example, the following teacher considered the programme to be beneficial to both the pupils and teachers. For teachers, the benefit comes from being more in tune with the style of current exams, which may have changed since they trained as a teacher.

'As teachers it's made it clearer to us the format of the exam, how the examiners are thinking. That's made it easier to teach it … I think just all around it's given the kids a technique that they can use in any question, but for me, mostly, it's made me think more about having teaching to match the style of the exams now' (teacher interview, case study 2).

Research question 10: What might be the main design options for the efficacy trial?

The design options for any future efficacy trial are discussed below.

Implementation and process evaluation recommendations

Considerations relating to establishing usual practice and gauging and monitoring fidelity would be central to the design of the implementation and process evaluation.

Defining usual practice at baseline

The evaluation has flagged that the individual components of the intervention already exist in isolation, and some teachers reported that they would already be engaging in similar teaching methods. It will be important to capture the extent to which teachers are already engaging in one or more of these activities.

It is anticipated that it may be difficult to identify and quantify the nature and extent of activities that overlap with the Deeper Thinking intervention when recruiting schools as these activities are taking place at an individual teachers' discretion and are not necessarily coordinated at a school level. This could be a problem when confirming and maintaining absence of treatment in the control schools.

It is proposed that when recruiting schools for the trial, data is collected on the extent to which GCSE science teachers currently engage in activities that involve the three central tenants of the intervention:

- SOLO Taxonomy;
- quick concept-mapping; or
- Dunloskian revision strategies.

In an ideal scenario, the teachers in the control schools would not use any of these techniques, but it is expected (based on what the current evaluation has found) that many teachers use a variant of concept-mapping. It would be unethical to impose a restriction among control group schools on the use of quick concept-mapping and revision strategies in line with the Dunloskian ratings. Therefore, it is recommended that the sampling approach for an efficacy trial would exclude schools where the SOLO Taxonomy is already used in conjunction with quick concept-mapping and evidence-based revision strategies in science teaching.

An efficacy trial would need to capture usual practice at baseline. It is proposed that this would take the form of a teacher survey that asks them to report the frequency with which they use SOLO Taxonomy, quick concept-mapping, the analysis of exam-question wording, constructing mark schemes, and writing their own exam questions. It will also be important to collect data on the extent to which teachers ordinarily set revision tasks in line with the strategies identified by Dunlosky et al. (2013) as high or medium utility. One could argue that the departmental approach to effective teaching techniques in general will influence the likelihood that a student adopts them during revision. Ideally, this could also be factored into the baseline although in practical terms it may be preferable to monitor this during the trial.

Fidelity to the intervention's subcomponents

The results from Table 20 (Appendix 11) suggest that any future trial will need to consider the fidelity of teachers delivering the intervention in its entirety. The intervention is predicated on teachers using the SOLO Taxonomy techniques and the quick concept-mapping together, but the survey and interview data from this evaluation suggest that in practice they were often used independently in this pilot. Any further trial should monitor the use of the SOLO Taxonomy, quick concept-mapping, and revision techniques together and separately. If a school's implementation is mainly focused on the use of laminated hexagons as a novel way of quick concept-mapping because they are not using the SOLO Taxonomy placemat, then the intervention school would be more similar than anticipated to the control group schools, among which it is assumed (based on what the current evaluation has found) that there is a high likelihood of teachers engaging in some similar activities in the form of mind mapping activities.

This presents potential issues relating to *how to collect the necessary data* on which techniques and activities are used (whilst minimising the burden on participants in the trial), *from whom* (pupils or teachers), and *how to overcome issues of data accuracy* with self-reporting systems such as surveys or logs. Alternatively, the intervention could be redesigned to focus on fewer theoretical aspects, for example, to just look at the impact of laminated hexagons in promoting the relational level of the SOLO Taxonomy. However, as discussed elsewhere, the intervention brings together components that already exist in isolation, so this may weaken its unique selling point.

Impact design recommendations

Overall, the key recommendations for the impact evaluation design—if this were to proceed to an efficacy trial—are as follows:

- randomise at school level;
- use GCSE as a primary outcome (but tier and awarding organisation would also need to be factored in);
- analyse pupil ability as a moderator of attainment outcomes;
- explore the effect of dosage (the number of times a pupil is exposed to the intervention) on outcomes; and
- measure fidelity to the programme as intended.

These are briefly discussed below.

Randomise at school level

In broad terms, we assume that the Deeper Thinking programme will be targeted at a pupil cohort entering Year 10 from September in a given school year and that the programme will run for two years (that is, throughout GCSE for this cohort). Further, that the trial will be a cluster randomised controlled trial with randomisation of whole schools to intervention and control groups. A cluster randomised controlled trial would avoid interference between units that could occur if control and intervention pupils or classes were located in the same school.

GCSE as a primary outcome

GCSE grade could be considered as the primary outcome for an efficacy trial although ideally it would be preferable to use overall marks as these allow for more differentiation in the outcome variable (note that results in marks would be converted into points scores for reporting purposes, as we are ultimately concerned with the intervention's capacity to improve GCSE attainment). Marks are not provided in the NPD so would need to be requested from the school or awarding organisation(s) separately. This would require a high level of cooperation from all schools (in both the control and intervention groups) and if this is not secured, there is a risk of high levels of missing data.

The exam paper analysis that formed part of this evaluation suggests that an efficacy trial should factor in whether the pupils are entered for Foundation or Higher tier papers since it may be more likely that there is an impact on grades amongst pupils entered for the Higher tier relative to the Foundation tier. This would require the schools to provide data on the exam paper entry patterns for individual pupils. Again, this approach has its inherent risks in terms of securing cooperation from schools or risking high levels of missing data. There is a case to be made that ideally a future trial would be specific to a single awarding organisation (AQA have the largest market share in GCSE science) as this would avoid the need to standardise between different qualifications.

The exam paper analysis considered combined (double award) science as an outcome measure. Whilst this is the most common science GCSE sat nationally, a large minority of the cohort sit separate science GCSEs (that is, in biology, chemistry, and physics). The grade outcome from double science is a 'combined' grade, such as 5-5 or 9-8, whilst the separate sciences will result in three grade outcomes; for example, 9, 9, and 8 across the three sciences. Numeric grades for single-award GCSEs convert straightforwardly into points scores, and double award grades convert onto the same points score scale, simply utilising the 0.5 decimal values between each grade to account for the 'mixed' grades; 9-8 for instance would convert to 8.5 in points score. This means that it is possible to measure the capacity of the intervention to improve the average science grade students achieve, regardless of the science GCSEs sat, by averaging the three points scores resulting from the separate sciences.

Analyse pupil ability as a moderator of attainment outcomes

The teacher data collected over the duration of the pilot suggests that the pupils' attainment is a key consideration when implementing the intervention in terms of delivery method and also the function of the intervention. It is recommended that the entire Year 10 cohort within each school takes part to avoid selection bias in the sample, for example, by schools selecting only higher-attaining classes. The Key Stage 2 scores for individual pupils would also be required for inclusion in any statistical analysis as this will provide a way of controlling for prior attainment.

Explore the effect of dosage on outcomes

The case study discussions also flagged that where learners have separate teachers for the three sciences, they may be exposed to the intervention on several occasions over a given week whilst a learner with a single-science teacher may only be exposed to the intervention once during the same period. Given the fact that the resources do not need IT access, this means there will be no way of automatically collecting usage data (for example, by auditing logins to an online platform).

It could be recommended that exposure data is collected from pupils rather than teachers, but this presents a data collection burden on schools and the inevitable missing data from pupils who do not complete the survey will present a challenge to the statistical analysis. Pupils would also need to be able to accurately recall the use of different elements of the intervention, which itself could be difficult as the intervention is complex. Ideally, pupil data would be sense checked against teacher data, but this may be tricky where pupils have several science teachers.

If pupils can be matched to teachers, data might be collected through periodic teacher surveys (which inevitably rely on respondents' recall, which might become less accurate the greater the time lag between surveys), or on some more systematic way of gathering data, for example, through asking respondents to keep logs or diaries of their use of the resources (which could be too burdensome on teachers, which might generate a high attrition rate). It is anticipated that measuring dosage is likely to present a significant challenge to a future evaluation, which presumably will want to include a measure of dosage as a predictor variable.

Sample size calculations for efficacy trial

Based on the analysis of possible distributions of grades and underlying marks attained at GCSE science under different scenarios presented in the accompanying research paper, Table 9 presents a range of plausible effect sizes related to assumed improvements in marks on average across the population of five, ten, and 15 marks respectively, subsequent to exposure to the intervention. In order to compute effect sizes, these mark gains were converted into points scores as reported in the NPD. (The accompanying research paper provides full details of the methodology used.)

The calculations presented in Table 9 were conducted using the PowerUp statistical program available in R. They are based on an assumption that sample effect size estimates will be obtained from a three-level hierarchical linear model where grade point is the dependent variable. The assumed covariates in the model are an intervention group membership indicator and a baseline covariate extracted from KS2 results (this is assumed to be the combined maths and reading score at KS2). The baseline covariate is entered into the model as a departure from the class mean at the pupil level, as a mean departure from the school mean at the class level, and as a mean departure from the overall mean at the school level. The covariate is assumed to explain about half the outcome variance at the pupil level, 20% at the class level and 10% at the school level. These assumptions are based on an analysis provided by Demack (2019) which does not relate specifically to science but maths and English, which are used in the absence of any better information. For the intraclass correlation coefficient, it is assumed that 20% of the total variance is at the school level. This value is typically used in the planning of EEF-funded cluster randomised trials and can be considered conservative (see for example Morris, Seymour, and Limmer, 2019). Finally, it is assumed that on average each school contains six classes per year with an average of 30 pupils in each class.

Whilst the accompanying research paper only considers the more common double award science, as outlined above, the grade outcomes of the separate sciences can be converted into an average points score on the same scale as double science's points scores. Considering the separate sciences as opposed to double science would undoubtedly result in marginally different estimations in the below tables, but this fell outside the scope of this analysis.

Table 9: Estimation of minimum detectable effect sizes for use in sample size calculations

Assumed average improvement in marks (across the population)	Mean points score in the control group	Mean points score in the intervention group	Difference in means	Pooled standard deviation	Minimum detectable effect size
5	4.11	4.33	0.21	1.72	0.12
10	4.11	4.54	0.42	1.71	0.25
15	4.11	4.74	0.63	1.71	0.37

Based on these values, we can present a range of sample sizes; that is, the total number of schools randomised to intervention and control that, given our assumptions, would result in a trial capable of detecting a given MDES at a 95% level of statistical significance.

Table 10: Sample size estimates

Minimum detectable effect size	Total number of schools
0.12	450
0.25	110
0.37	50

As can be seen in Table 10, the assumed effectiveness of Deeper Thinking on attainment at GCSE is a crucial determinant of the required sample size for any trial. If we accept the lowest estimate of effectiveness (ES = 0.12), the requirement is for a sample of over 400 schools. A sample of this size is logistically impractical for an efficacy trial. If, however, it is plausible to expect the true impact of Deeper Thinking on the population to reach ES = 0.25 or even higher, then a trial of around 110 schools would be sufficient for a meaningful test of the intervention. A trial sample of this order of magnitude is more realistic both from practical and cost perspectives. As a result, our judgement is that if we are content to assume Deeper Thinking is capable of producing an increase in the average mark across the population of pupils of at least ten marks (out of 420 available for the double award specification) then, from the perspective of having a definable and measurable primary outcome, an efficacy trial would be worth undertaking.

Will GCSE science be a suitably precise and targeted measure to pick up a measurable impact on pupil attainment?

The accompanying research paper assessed how many extended-answer questions (EAQs) and practical questions there are on the AQA Science Trilogy paper as these are the questions targeted by the Deeper Thinking Intervention. This specification was selected as a majority of the national cohort sit double award science, and, of those, AQA has the largest market share; considering more than one specification was out of scope. It is worth noting that, whilst outcomes on both double science and the separate sciences can be put onto the same (average) points score scale and compared, in considering only combined science we are not dealing with the entire cohort (anecdotally, the separate sciences tend to be sat by more-able students in schools that deliver them alongside the double award science).

The analysis of the AQA Science Trilogy specification found that:

- for Foundation tier, approximately 42% of the items are EAQs and/or on practical content; and
- for Higher tier, approximately 62% of items are EAQs and/or on practical content.

The paper also estimated the proportion of pupils in each tier that would see an increase in their science GCSE grade if they gained additional marks on the extended answer and practical questions as a result of the Deeper Thinking Intervention. The analysis suggests that:

For the Foundation tier:

- around 3-4% of pupils fall one mark below a grade boundary; and
- around 18% of pupils fall within five marks of the next grade boundary.

For the Higher tier:

- around 5% of pupils fall one mark below a grade boundary; and
- around 26% fall within five marks of the next grade boundary.

If the Deeper Thinking intervention does result in gains, even of only a handful of marks on GCSE Science, it has the potential to improve the grade of at least a few pupils in a class of 30 (above what they would have achieved without the intervention). As an example, if five marks were gained, on average, across a 30-pupil class, where half the class are entered for higher and half entered for the foundation tier, we could expect, on average, 3.9 higher-tier pupils to gain a grade and 2.7 foundation-tier pupils to gain a grade—based on the expected proportions of pupils within five marks of a grade boundary. In any one class, these figures are extremely unlikely to pan out exactly, however; they reflect the average across an intervention population.

This analysis suggests that the Deeper Thinking intervention is less likely to impact the grades of higher-attaining, Higher-tier pupils as they have less marks remaining as 'potential gains'. It is also less likely to impact the grades of high performers on the Foundation tier as there is a capping effect.

That said, we do not expect the gain in marks and grades to be uniform for pupils performing at different levels. Any future analysis would need to consider school strategies in terms of entering 'borderline' candidates for Foundation and Higher papers and whether the intervention interacts with these patterns.

Are there secondary outcome measures that can be addressed by an efficacy trial?

As discussed in the section on unintended consequences, reducing science teacher marking workload could be considered to be a potential secondary outcome for consideration. However, it is unclear how prevalent this outcome is in the pilot as it was not explicitly measured. Whether it would be considered to be of enough importance to be operationalised in an efficacy trial would need to be explored further. The ongoing EEDI evaluation (Morris et al., 2019) considers teacher workload as a secondary outcome measure and could be used as a template for designing teacher survey questions to address this.

Conclusion

Summary of pilot findings

Question	Finding	Comment
Is there evidence to support the Theory of Change?	Mixed	There is some evidence of positive changes to teaching practice and perceptions of an impact on pupils, but some teachers were already using similar techniques.
Was the approach feasible?	Mixed	The programme has low costs, is easy to set up, and does not require access to IT. Fidelity to the programme implementation and delivery may have been a cause for concern as a substantial proportion of teachers reported that they were not consistently using the SOLO Taxonomy placement and quick concept-mapping hexagons together. Some teachers reported that strategies may support lower-ability learners, but others reported that it was a challenge to apply the Deeper Thinking intervention to lower-attaining pupils and those lacking the skills/capacity to work independently.
Is the approach ready to be evaluated in a trial?	No	The delivery partner would need to adapt the Deeper Thinking classroom packs and training materials to better prepare teachers of lower-attaining and disruptive classes.

Formative findings

Based on the evaluation of this pilot, there are a number of formative findings that should be considered if the intervention went to trial.

A key issue identified throughout the evaluation was implementation of the programme as intended. Even in the context of the short timelines, it was notable how few teachers represented in the survey data and during the case study visits had fully implemented the programme, for example, by using all resources: nearly two-fifths of post-pilot survey respondents reported that they used just the hexagons and not the SOLO Taxonomy placemats at least half of the time. The use of evidence-based revision techniques was also somewhat limited, however this is likely to be a factor of the year group that the intervention was piloted with—that is, Year 10 rather than Year 11—and we might expect to see greater use of these techniques if any future trial ran over the course of two years. On this note, some participants in the evaluation suggested that the Deeper Thinking techniques should be introduced before Year 10 so that pupils are familiar with them meaning that valuable Year 10 and 11 lesson time is not taken up with teaching new ways of learning but can focus on the GCSE science content.

Training might also need reviewing if the intervention is rolled out more widely. The Leadership Day CPD session was well-received and felt to be useful by all who completed the evaluation form administered by the programme developers; however, the resource demands of running this format of training across a much larger number of schools must be considered. There was evidence to suggest that the training sessions delivered by the more senior members of the Carmel Trust team were more positively received than those delivered by third party trainers; this would have implications for the larger scale delivery of the programme in terms of ensuring the quality of the training delivered regardless of the position of the deliverer. The case study visits indicated that some teachers felt that the twilight sessions offered to science teachers were too long, so there is scope for reducing the content of these (for example, by reducing the amount of 'theory' covered and focusing on the practicalities of delivery) which might make the training less resource-intensive and a more feasible option for delivery on a larger scale.

More specifically, the evaluation data suggests that the training sessions could be improved to maximise teacher engagement by:

- offering several live modelling sessions so all teachers could attend and offering a session on each area of the intervention—not just the hexagon mapping applied to an extended-answer question, which may be the most straightforward aspect of the intervention;

- providing more templates and checklists for teachers taking part so they have a teaching aid when they come to put the training into practice; and
- providing more resources and modelling in relation to lower-attaining or disengaged pupils.

On this final point, the evaluation found evidence that it was more challenging for some teachers to deliver the programme to lower-attaining and disengaged pupils. It is likely that a wider roll-out or largescale trial of the intervention will require adjustments to make the programme more suitable for these groups, particularly as they are arguably the groups who need it the most.

Interpretation

The evaluation found that the Deeper Thinking intervention could be summarised in a Theory of Change model (see Figure 2); however, this is quite complex due to the intervention combining the SOLO Taxonomy, quick concept-mapping, and revision techniques informed by Dunlosky et al. (2013). The intervention is plausible in the sense that it is based on established academic literature (Biggs and Collis, 1982 and Dunlosky et al, 2013). The human, material, and financial resources required to implement the intervention were reported to be low and there was nothing to suggest that the intervention is not meaningful. The impact of the intervention can be tested in terms of attainment in GCSE science but the testing assumes that the intervention is delivered as intended (that is, the teachers use SOLO Taxonomy and quick concept-mapping together alongside revision techniques informed by Dunlosky et al. 2013) as it would present a significant evaluative challenge to test the impact of the individual components.

There was some evidence of change to Year 10 teachers' teaching practice because of the Deeper Thinking intervention, namely making more links between science topics and focusing more on practicals and exam writing technique. Year 10 teachers also perceived there to be some changes in pupil behaviour because of the Deeper Thinking intervention. There was some evidence of an increase in pupils identifying and linking science knowledge and improvements in pupils' abilities to communicate scientific knowledge during group work and when answering example exam questions. There was mixed evidence in terms of the intervention improving engagement in GCSE science; whilst there were some positive accounts of perceived improvements in engagement during Deeper Thinking lessons, many teachers also raised concerns that the intervention needed tailoring to meet the needs of lower-attaining pupils and those lacking the skills or capacity to work independently. This is important when one considers that the intervention aims to teach self-regulation and part of the EEF rationale for pursuing the trial is that teaching self-regulation strategies has been found to benefit low prior-attainers more than high prior-attainers.

A targeted review of similar interventions found that comparable resources are available to teachers to deliver the individual components of the Deeper Thinking intervention, but they are not bundled into a single intervention. There was a lack of evidence evaluating the impact of these resources either separately or in combination with one another. The case study discussions and teacher survey found that many of the teachers would ordinarily be using some of the Deeper Thinking techniques, namely concept-mapping, linking techniques, analysing exam questions, and mark schemes, which raises some flags that the intervention may not be sufficiently innovative compared to current practices to lead to an effect on outcomes.

Within the four-month window for the pilot there was evidence that the intervention was not always implemented as intended. Whilst all the teachers who responded to the June survey reported applying the Deeper Thinking techniques to analysing extended-answer questions, less than half of the survey respondents reported constructing mark schemes and writing their own exam questions. Furthermore, there was evidence that teachers were not always using the classroom packs as intended. Specifically, many teachers reported using the laminated hexagons for quick concept-mapping whilst not using the accompanying SOLO Taxonomy placemat. This suggests that the Deeper Thinking developers may need to refine the CPD sessions to support the greater use of the placemats or they could narrow the intervention to focus solely on the use of the hexagons.

The ideal way of delivering the intervention would include scheduling the activities so they are distributed evenly over the two-year period and allowing for the intervention activities to be planned around the required practicals, providing schools with one hexagon pack per teacher, and adding a way of allowing pupils to capture the hexagon work so they have a physical takeaway.

The exam paper analysis suggests that GCSE science is a suitably precise and targeted measure to pick up a measurable impact on pupil attainment as 42% of items on the Foundation tier and 62% of items on the Higher tier of AQA Science Trilogy are extended-answer questions and practical questions, the questions that the Deeper Thinking intervention aims to address. The analysis also suggests that around 18% of pupils on the Foundation tier and around 26% of pupils entered for the Higher tier are within five marks of a grade boundary, which suggests that the Deeper Thinking intervention could improve the grades of pupils if it is, as intended, able to provide pupils with the techniques needed to gain five extra marks on extended-answer questions and practical questions.

Whilst the intervention is theoretically defined and scalable, the fidelity issues highlighted above mean that it is difficult to recommend progression to an efficacy trial without carefully specified design options that measure how teachers are using the individual components of the Deeper Thinking intervention as the pilot data suggests that many teachers are not using the classroom packs as intended. This may present a substantial evaluative challenge. An additional challenge will be posed by the need to monitor dosage as no automated online usage data is generated. These evaluative challenges are considered to be substantial and, when combined with the need to tailor resources and training to meet the needs of lower-attaining pupils, may mean progression to efficacy trial needs to be treated with caution.

Alternatively, it may be that the activity of using laminated hexagons as a way of encouraging pupil recall and link-making, on their own, could form a much simpler intervention, provided strategies are devised to assist with their use with pupils of varying ability and engagement. However, as previously discussed, it is the combination of the three elements that gives the Deeper Thinking approach a unique contribution.

The main limitations of the evaluation are as follows. As discussed above, the time frame for the pilot was extremely short and represented a very condensed version of what is intended to be a two-year intervention. This is likely to have affected the extent to which the programme has been fully implemented in schools, and the relevance of some elements of the intervention to the stage pupils are at in their GCSE programme (for example, the use of evidence-based revision techniques was not particularly appropriate in the timeframe and with the cohort involved in the evaluation).

Where the evaluation sought stakeholders' views, these focused on teachers' perceptions of the intervention and there was no opportunity to gather evaluation data from the pupils to seek their views on the intervention. This means we are relying on teachers' reports of the perceptions of and impacts on pupils. Given that most of the outputs identified as being within the scope of this pilot relate to the pupils themselves, it would have been useful to gather evidence directly from the pupils if the evaluation timelines and budget had permitted. However, it is acknowledged that this may have made it more onerous for the schools taking part in the trial as they would have been required to liaise with parents to gain consent.

The relatively low teacher response rate to the post-pilot survey is also a limitation. We cannot make reasonable assumptions about the nature of the non-respondents that will indicate how representative those who did respond are of participants overall. It might be that non-respondents all implemented the programme but did not have time to complete the survey, but it is also possible that those who did not respond had not implemented the programme at all, or felt that they had not done so to a sufficient level to make their responses useful to the evaluation. Those who contacted the evaluators to say they were not able to complete the survey, where they gave reasons, tended to indicate that this was because they had not been trained and/or had not used the intervention (for example, due to sickness or absence), but these were the minority and, despite repeated efforts, is was very challenging to get responses to the post-pilot survey. Another potential factor in the low response rate is the late stage at which some schools were recruited to the pilot; this might have led to teachers being signed-up who were never realistically going to be able to participate.

Future research and publications

Specific recommendations for the design of a larger-scale trial of this programme are provided above in the discussion of scalability/readiness for trial. Briefly, it is felt that provided the necessary adjustments are made to the resources and training, a future trial could be undertaken and could collect as a primary outcome measure pupil attainment (GCSE science performance), and that possible secondary outcome measures might include teachers' workload and pupil engagement. However, there was evidence from the evaluation to suggest that the intervention as it stands might not be suitable for lower-attaining and disengaged pupils, which, in effect, means it could widen rather than close the gap between higher- and lower-attaining pupils. This is contrary to the aims of the EEF and therefore, in its current form it is not felt that a recommendation to proceed to an efficacy trial can be made on the basis of the pilot evaluation.

If the intervention could be adjusted to sufficiently address these issues, the four key areas for consideration in the design of a larger scale trial would be:

- Firstly, establishing robust systems for gauging usage of the resources and techniques.
- Secondly, establishing how usage of the individual components differs from ‘business as usual’ in the schools;
- Thirdly, measuring any potential difference in experience of the programme for pupils of differing attainment levels;
- Finally, it would be important to establish the desirability of and mechanisms for measuring secondary outcomes, such as teacher workload.

References

AQA (2019) ‘Combined Science: Trilogy (8464) | Teaching resources’. https://www.aqa.org.uk/subjects/science/gcse/combined-science-trilogy-8464/teaching-resources

Biggs, J. and Collis, K. (1982) *Evaluating the Quality of Learning: The SOLO Taxonomy* (1st edn), New York: Academic Press.

Carew, L. and Mitchell, A. (2002) ‘Characterizing Undergraduate Engineering Students' Understanding of Sustainability’, *European Journal of Engineering Education*, 27 (4), pp. 349–361.

Daley, B. and Torre, D. (2010) ‘Concept Maps in Medical Education: An Analytical Literature Review’, *Medical Education*, 44 (5), pp. 440–448.

Demack, S. (2019) ‘Does the Classroom Level Matter in the Design of Educational Trials? A Theoretical and Empirical Review’, London, Education Endowment Foundation.

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

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J. and Willingham, D. T. (2013) ‘Improving Students’ Learning with Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology’, *Psychological Science in the Public Interest*, 14 (1), pp. 4–58.

Education Endowment Foundation (2018) ‘Improving Secondary Science, Seven Recommendations for Improving Science in Secondary Schools’. https://educationendowmentfoundation.org.uk/tools/guidance-reports/improving-secondary-science/

Educators Technology (2018) ‘Nine Great Concept Mapping Tools for Teachers and Students’ (blog). https://www.educatorstechnology.com/2018/01/9-great-concept-mapping-tools-for.html

Hedges, L. V. (2011) ‘Effect Sizes in Three-Level Cluster-Randomized Experiments’, *Journal of Educational and Behavioral Statistics*, 36 (3), pp. 346–380.

HookED (2017) Downloadable resources: http://pamhook.com/free-resources/downloadable-resources/

Hook Education (2019) ‘SOLO Hexagons’: https://itunes.apple.com/us/app/solo-hexagons/id1023237205

Lilly, J., Peakcock, A., Shoveller, S. and Struthers, R. (2014) ‘Beyond Levels: Alternative Assessment Approaches Developed by Teaching Schools’, National College for Teaching and Leadership. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/349266/beyond-levels-alternative-assessment-approaches-developed-by-teaching-schools.pdf

Morris, S. P., Seymour, K. and Limmer, H. (2019) ‘Research Protocol: Evaluating the Impact of EEDI Formative Assessment Online Platform (formerly Diagnostic Questions or DQ) on Attainment in Mathematics at GCSE and Teacher Workload’, *International Journal of Educational Research*, 93, pp. 188–196. https://doi.org/10.1016/J.IJER.2018.11.007

Nesbit, J. and Adesope, O. (2006) ‘Learning with Concept and Knowledge Maps: A Meta-Analysis’, *Review of Educational Research*, 76 (3), pp. 413–448.

Ofqual (2015) ‘GCSE Science: Decisions on Conditions and Guidance’. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/443980/gcse-science-decisions-on-conditions-and-guidance.pdf

Ofqual (2017) ‘Guidance GCSE, AS and A level Assessment Objectives’. https://www.gov.uk/government/publications/assessment-objectives-ancient-languages-geography-and-mfl/gcse-as-and-a-level-assessment-objectives

Olsen, W. (2018) *Data Collection: Key Debates and Methods in Social Research*, London: SAGE (pp. 33–38). https://doi.org/10.4135/9781473914230.

ResearchSchool.org.uk (2019) ‘Three Day Training programme: Improving memory’. https://researchschool.org.uk/durrington/event/training-programme-improving-memory

Teachit (2019) *Closing the Word Gap: Activities for the Classroom—Science*, Oxford: Oxford University Press. http://fdslive.oup.com/www.oup.com/oxed/Closing_the_Word_Gap_-_science.pdf?region=uk

The Learning Scientists (2019) ‘Six Strategies for Effective Learning: Materials for Teachers and Students’. http://www.learningscientists.org/downloadable-materials

University of Sheffield (n.d.) ‘Observations’. https://www.sheffield.ac.uk/lets/strategy/resources/evaluate/general/methods-collection/observation

van Boxtel, C., van der Linden, J., Roelofs, E. and Erkens, G. (2002) ‘Collaborative Concept Mapping: Provoking and Supporting Meaningful Discourse’, *Theory Into Practice*, 41 (1), pp. 40–46.

Wolpert M., Sharpe H., Humphrey N., Patalay P. and Deighton, J. (2016) *EBPU Logic Model*, London: CAMHS.

Wilkinson, S. (1998) ‘Focus Group Methodology: A Review’, *International Journal of Social Research Methodology*, 1 (3), pp. 181–203. https://www.tandfonline.com/doi/abs/10.1080/13645579.1998.10846874

Appendix 1: The Deeper Thinking Classroom Pack

The hexagons are cut out individually to allow for tessellation and laminated so they can be written on and wiped clean as required. Each pack contains 22 hexagons.

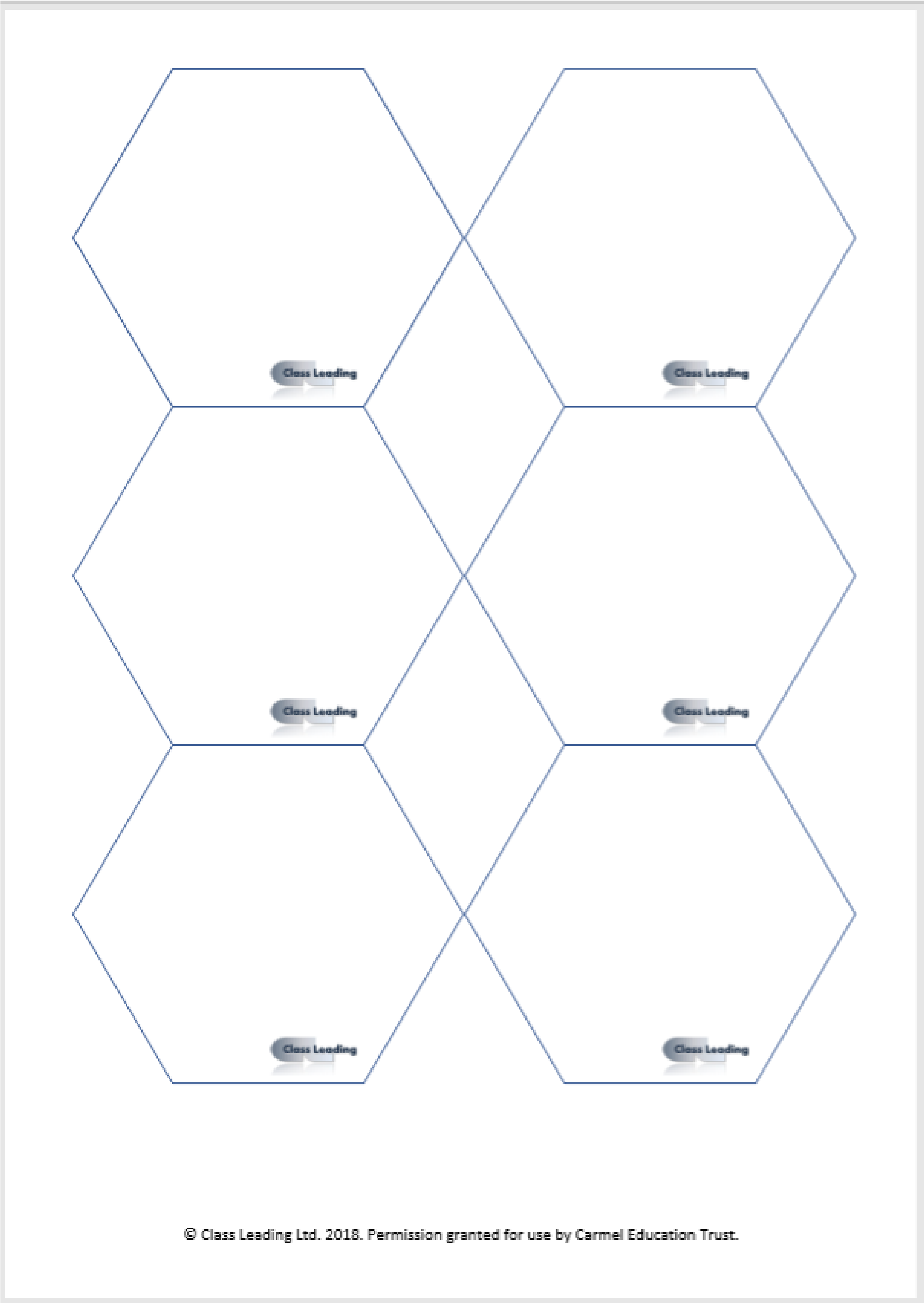


Figure 3: Quick concept mapping hexagons

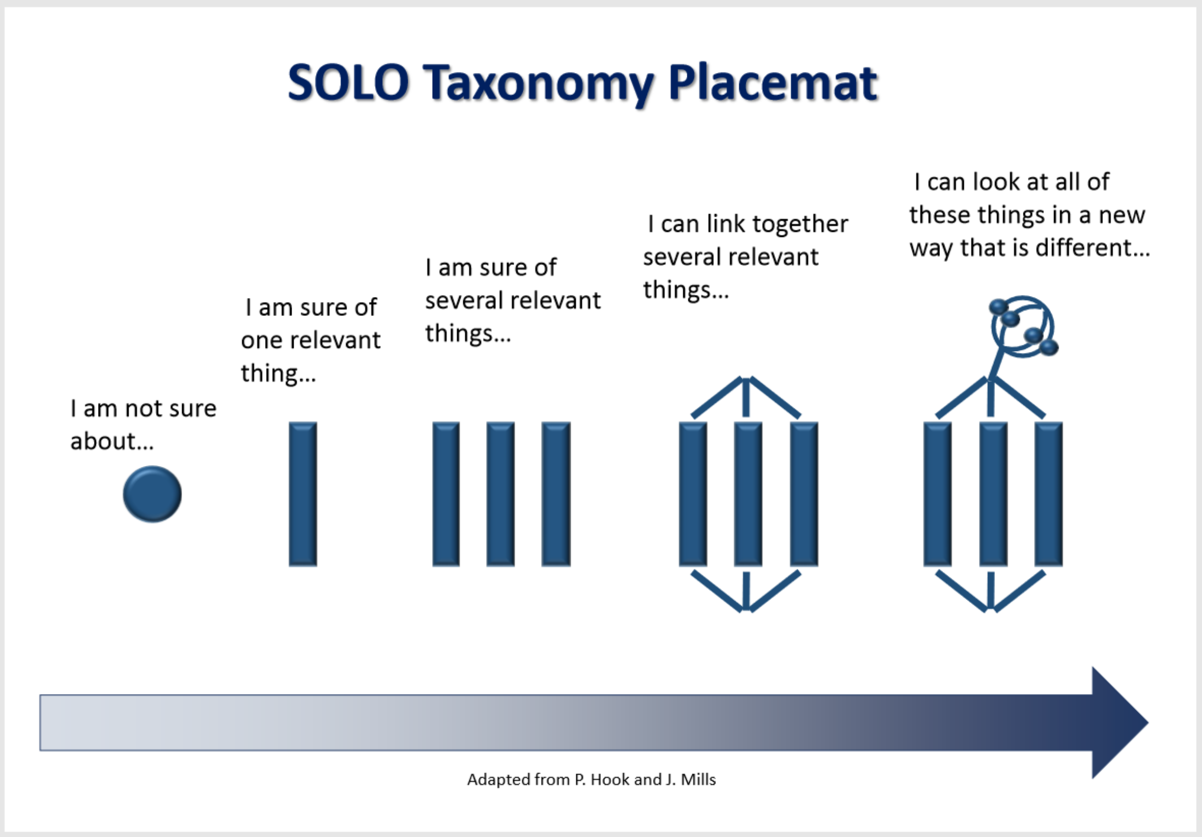
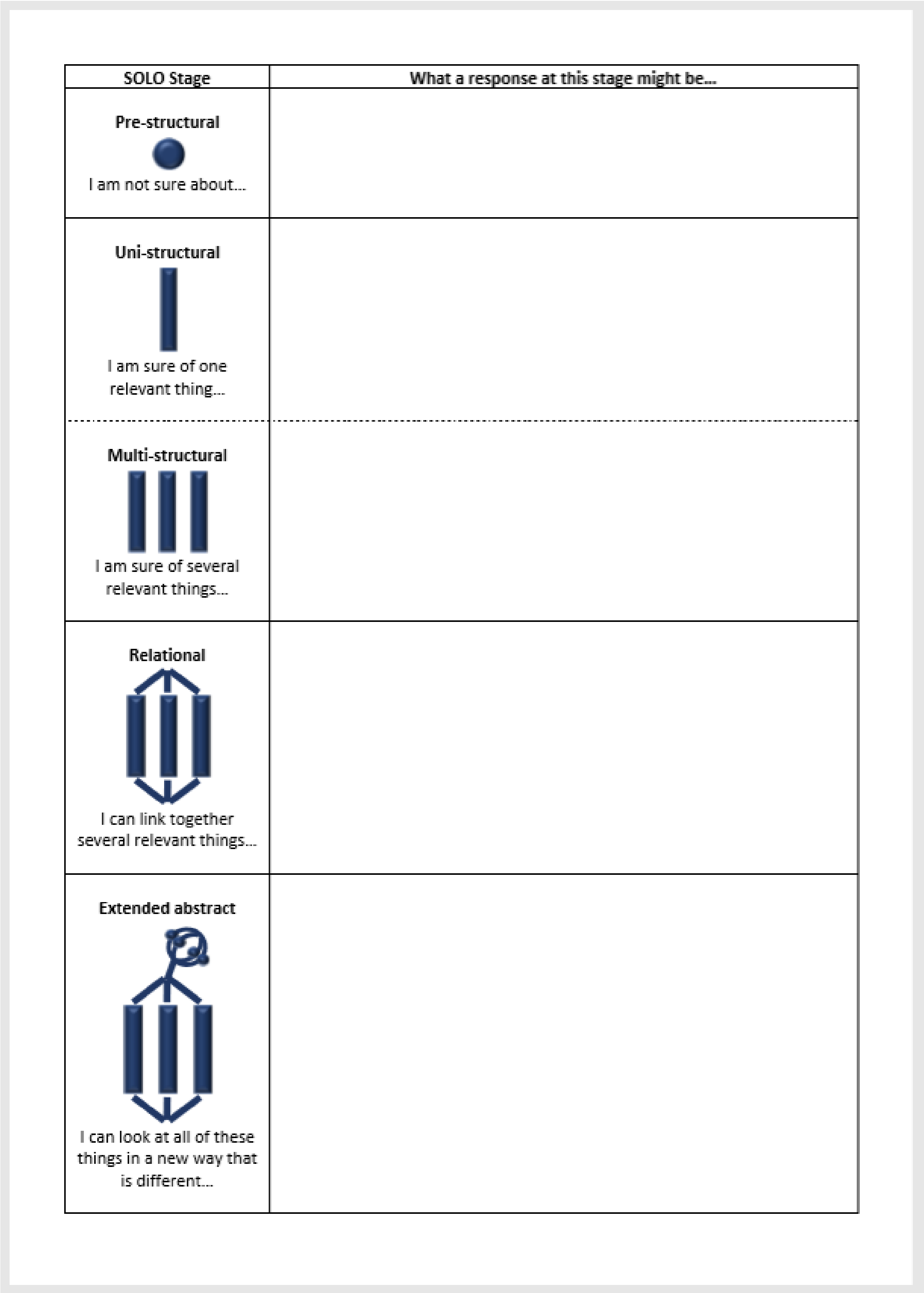


Figure 4: SOLO Taxonomy placemat, used by pupils



SOLO Stage	What a response at this stage might be…
Pre-structural I am not sure about…	
Uni-structural I am sure of one relevant thing…	
Multi-structural I am sure of several relevant things…	
Relational I can link together several relevant things…	
Extended abstract I can look at all of these things in a new way that is different…	

Figure 5: SOLO Taxonomy Rubric

Appendix 2: Memorandum of Understanding (MOU)

Agreement to participate in EEF's Deeper Thinking project evaluation

This MOU sets out the roles and responsibilities of schools participating in, and the parties involved in delivering and evaluating, the Deeper Thinking intervention and evaluation. Schools should read this MOU in full and direct any questions about the intervention to Carmel Education Trust and any questions about the evaluation to AlphaPlus. Schools should then sign and return it to Carmel Education Trust as per the instructions at the end of this document. Further information about how this research complies with GDPR is provided in the Privacy Notice on page 3.

The Deeper Thinking intervention is being delivered by Carmel Education Trust (CET) who we refer to as the project 'developers'. The project is funded by the Education Endowment Foundation (EEF). EEF are an independent charity that fund research to test various way of raising attainment in English schools. The project is being evaluated by AlphaPlus, an independent education sector consultancy. We refer to AlphaPlus as the 'evaluator'.

1. Aims of the project

The aim of this project is to evaluate a pilot programme of the Deeper Thinking intervention in up to 12 secondary schools in the north-east of England and assess whether it would be suitable to proceed to a full trial at a national level.

2. About the Deeper Thinking intervention

The Deeper Thinking intervention aims to improve pupil outcomes on two areas of GCSE science assessment – the assessed practical tasks and the extended response questions (which require detailed description, explanation or use of information) in GCSE exams.

Students will practise the technique of using hexagons as a concept-mapping tool, supported by SOLO taxonomy to encourage link making and increase the quality of the answer.

The objectives are that the pupils better understand success criteria and have a clearer understanding of the link between different ideas. This should allow them to better identify variables that are relevant and may confound during assessed practicals and structure their extended answer questions so that they include all the key ideas, can show the links between ideas and are not deterred from answering due to unfamiliar contexts.

Each school will identify a leader for the implementation of the project and this person will be trained in key elements of the approach through a one-day training event. All schools will then receive three departmental twilight training session (two hours each) that will be attended by all teachers teaching GCSE science. The intervention would then be delivered throughout Years 10 and 11, with teachers embedding the approach in their normal teaching.

3. The evaluation

As stated above, the EEF Deeper Thinking project evaluator is AlphaPlus Consultancy.

3.1 Who is included

Up to twelve schools will participate in the evaluation.

The evaluation will take place between October 2018 – August 2019 with process evaluation activities taking place from October 2018 to July 2019. It will be focussed on science classes from Year 10.

All Year 10 science teachers will be asked to complete a short online survey in January 2019 and June 2019.

The Deeper Thinking lead may be asked to take part in an interview.

3.2 Data collection

The evaluators will undertake the following data collection activities.

With pilot schools:

- Baseline teacher survey data collection (Jan 2019)
- Post-intervention teacher survey (June – July 2019)

In addition, eight of the twelve school will be asked to take part in a case study visit. This will last up to one day and will involve the following three activities:

- An interview with the Deeper Thinking lead (up to 1 hr)

- A focus group or paired interview with Year 10 science teachers (up to 1 hr)
- A classroom observation (up to 1 hr or as determined by the school).

The evaluators will work with the school to make this as flexible as possible.

An evaluator may also attend one of each of the four CPD sessions offered as they are due to attend one of each of the four sessions, in total, not per school. This is solely to observe how the Deeper Thinking training is provided by Carmel.

We will ask each school to provide the following information to enable us to contact all Year 10 science teachers to carry out the activities specified above:

- Teacher names
- School name and exam board for GCSE science
- Teacher email addresses and phone numbers

3.3 Data protection

All teacher personal data will be treated with strictest confidence by the evaluators in accordance with the requirements of the GDPR 2018. The evaluators will not share the personal contact information with any other parties as its sole use will be to contact teachers to arrange case study visits, send links to online surveys and to confirm that they are currently teaching Year 10 science classes.

Although we will observe up to one science lesson per school, we will not collect any personal information from or about students and we will emphasise that we are observing the intervention in action and not the performance of teachers or students.

We will make it clear to all teaching and school staff that their participation is voluntary, and they can withdraw from the evaluation at any point.

No school, teacher or student will be identified in any report arising from this evaluation. The information collected will be used for research purposes only and no information that can identify individuals will be used for any other purpose. Any personal data collected will be destroyed in accordance with the GDPR when it is no longer required.

AlphaPlus will provide all participants with a Privacy Notice (below) before each evaluation activity commences, for example the notice will be included at the beginning of each online survey and provided in advance of an interview.

Privacy Notice

Why we are collecting data?

The data sharing is necessary for the parties to undertake a research project into the effectiveness of the use of the Deeper Thinking intervention aimed at science students in Year 10.

This project is in the public's interest as the results will help assess the potential performance of Deeper Thinking in its pilot phase and to help decide if the intervention should move to a full trial to assess its impact on student achievement.

The intervention is designed to improve student outcomes on two areas of GCSE science assessment – the assessed practical tasks and the extended answer questions.

The collection and sharing of contact data and surveys, interviews and observations from teachers participating in the Project is necessary to assess the impact of the intervention.

The project is funded by the Education Endowment Foundation (EEF). EEF are an independent charity that fund research to test various way of raising attainment in English schools.

Who are we collecting data from?

We are asking schools to provide us with contact information for all Year 10 science teachers (names, phone numbers and email addresses) and we will use this information to invite teachers to participate in our evaluation activities in the form of

- • Baseline teacher survey data collection (Jan 2019)
- • Post-intervention teacher survey (June – July 2019)
- • An interview with the Deeper Thinking lead (up to 1 hr)

- • A focus group or paired interview with Year 10 science teachers (up to 1 hr)
- • A classroom observation (up to 1 hr or as determined by the school).

You have been asked to participate as you are a Year 10 science teacher in a school piloting the Deeper Thinking intervention.

Who is collecting data?

The Pilot intervention is being led by Carmel Education Trust (CET) and evaluated by AlphaPlus Consultancy Ltd.

CET will be the originating Data Controller for any teacher data they disclose to the parties under this project.

AlphaPlus will be the Data Controller in respect of any personal data of teachers which they process for the purposes of the project.

The parties rely on the below processing condition to process personal data under this agreement: The processing is necessary for purposes of legitimate interests (article 6(1)(f)).

What data is being collected?

- Teacher names
- School name and exam board for GCSE science
- Teacher email addresses and phone numbers

We will only use the data to contact you in advance of evaluation activities and to send online survey requests. Your personal information will not be used in our data analysis and we will not identify any individual in our findings.

The data will not be matched to other data sets or used for any automated decision-making or profiling.

How is the data stored?

Any data provided to us will be stored securely according to our data security policy –Only authorised project members will be able to access the data, which will be kept on a secure, password protected server. Any data transfers will use encryption.

AlphaPlus will destroy all personal data associated with this project, including data sets received from the Data Controllers.

Contact

If you have any questions about the evaluation, please contact Clare Dowland at AlphaPlus on 01962 840362 or clare.dowland@alphaplus.co.uk.

The project has received ethical approval from the AlphaPlus ethics committee.

4. Responsibilities for schools and evaluators

Responsibilities for schools and the evaluators are set out below.

4.1 Responsibilities for CET

Recruit 12 schools to the Deeper Thinking intervention pilot.

Distribute MOU to each school.

Act as first point of contact with the schools in the pilot to provide the evaluators with:

Contact details for all Year 10 science teachers – name, email address for the teacher survey.

Phone number, name and email address for the Deeper Thinking lead in the eight schools to be visited.

Deliver CPD for each school, with 3 sessions for science teachers and a full day for the nominated lead.

Provide ongoing support to teachers through an online platform.

If any existing materials from previous Deeper Thinking interventions are made available to the evaluators, CET will advise the schools and/or teachers of the reason for this and provide relevant information to them to ensure GDPR compliance.

4.2 Responsibilities for schools

Allow Year 10 science teachers to attend CPD on the Deeper Thinking intervention

Appoint a lead for Deeper Thinking in the school

Distribute teacher information sheets to Year 10 science teachers and Deeper Thinking lead

Provide contact information for all Year 10 science teachers and the Deeper Thinking lead to CET to share with the evaluators

Ensure teachers complete the two teacher surveys, as required by the evaluators.

If selected to take part in a case study visit, liaise with the evaluators to arrange a case study visit to each school and allow the evaluators to visit the school and observe a CPD session for Deeper Thinking; observe a Year 10 science lesson where the Deeper Thinking intervention will be utilised; interview the Deeper Thinking lead for the school; and facilitate a focus group with Year 10 science teachers who are utilising the Deeper Thinking approach.

4.3 Responsibilities for the evaluators

Answer any queries about the evaluation.

Collect, manage, store and analyse the data required for the evaluation.

Provide privacy notices to any individual participating in the evaluation.

Administer teacher surveys and telephone interviews.

Conduct case study visits in a small number of schools during Spring 2019.

Ensure all members of the evaluation team are appropriately trained and have full Disclosure and Barring Service (DBS) clearance.

Provide a report on the findings of the evaluation.

5. Timeline

2019

January	Baseline teacher survey data collection
April	Case study visits
June-July	Post-intervention teacher survey

6. AlphaPlus contact

If you have any questions about the evaluation, please contact Clare Dowland at AlphaPlus on 01962 840362 or clare.dowland@alphaplus.co.uk

We commit to the evaluation of Deeper Thinking as set out above.

Please sign two copies. Keep one and email the other to zak.horrocks@alphaplus.co.uk, or send by post to:

Zak Horrocks
AlphaPlus
Unit 109 Albert Mill,
50 Ellesmere Street, Castlefield,
Manchester,
M15 4JY

School Name ………………………………………..

Name of Head/Exec Principal/Manager ………………………………………..
Signature ……………………………………….. Date …………………………

Name of Deeper Thinking lead ………………………………………..
Signature ……………………………………….. Date …………………………

Project Director for AlphaPlus Andrew Boyle

Signature Date 20/11/18

Project Director for Carmel Education Trust ………………………………………..
Signature ………………………………………. Date ………………………..

Appendix 3: Teacher Information sheet

AlphaPlus Consultancy Ltd
109 Albert Mill
10 Hulme Hall Road
Castlefield
Manchester
M15 4LY

Date: November 2018

Dear Sir/Madam,

EEF Deeper thinking intervention evaluation

You have received this letter/email as your school has been asked to take part in a pilot Deeper Thinking Intervention, which targets Year 10 Science students.

This intervention is designed by the Carmel Education Trust and the pilot intervention and its evaluation is funded by the Education Endowment Foundation.

AlphaPlus Consultancy is responsible for carrying out the evaluation to assess whether the intervention should proceed to a larger efficacy trial.

Why we need your participation

To evaluate the Deeper Thinking intervention we need to collect insight from the teachers delivering the intervention.

What you may be asked to do

The planned evaluation activities include:

A short survey in January and June 2019 (each taking less than 15 minutes to complete)

12 schools are taking part and eight of these will be visited for one day in April-May.

If your school is involved in a case study visit you will be asked to take part in a focus group discussion about the intervention. If you are the lead teacher this will be a 1-1 interview.

During this day one Year 10 science class will be observed for approximately one hour. This may not be your lesson as only one observation will take place. The purpose of the observation is to see how the intervention is delivered in different settings, it is solely to appraise the intervention not the teacher delivering it.

What will happen to the information you provide?

The information you provide about your experiences of using the intervention will be used solely by AlphaPlus for the purposes of this evaluation, and it will not be shared with your school, the Carmel Education Trust, or anyone else.

The data collected over the evaluation will be summarised in a report that will be shared with the Education Endowment Foundation and Carmel Education Trust. The data will be analysed and written up in a way that means that neither you nor your school will be identified.

Although we will observe up to one science lesson per school, we will not collect any personal information from or about students.

Participation is voluntary, and you can withdraw from the evaluation at any point.

Data retention and sharing

All teacher personal data will be treated with strictest confidence by the evaluators in accordance with the requirements of the GDPR 2018. The evaluators will not share the personal contact information with any other parties as its sole use will be to contact teachers to arrange case study visits, send links to online surveys and to confirm that they are currently teaching Year 10 science classes.

No school, teacher or student will be identified in any report arising from this evaluation.

The information collected will be used for research purposes only and no information that can identify individuals will be used for any other purpose. Any personal data collected will be destroyed in accordance with the GDPR when it is no longer required.

I hope you will be willing to take part in these evaluation activities, but if you have any concerns please raise them with myself and your school.

Yours Sincerely,

Dr Hayley Limmer
Senior Researcher AlphaPlus

T: +44 (0) 161 249 9249
DD: +44 (0) 161 249 9263
hayley.limmer@alphaplus.co.uk

Privacy Notice

Why we are collecting data?

The data sharing is necessary for the parties to undertake a research project into the effectiveness of the use of the Deeper Thinking intervention aimed at science students in Year 10.

This project is in the public's interest as the results will help assess the potential performance of Deeper Thinking in its pilot phase and to help decide if the intervention should move to a full trial to assess its impact on student achievement.

The intervention is designed to improve student outcomes on two areas of GCSE science assessment – the assessed practical tasks and the extended answer questions.

The collection and sharing of contact data and surveys, interviews and observations from teachers participating in the Project is necessary to assess the impact of the intervention.

The project is funded by the Education Endowment Foundation (EEF). EEF are an independent charity that fund research to test various way of raising attainment in English schools.

Who are we collecting data from?

We are asking schools to provide us with contact information for all Year 10 science teachers (names, phone numbers and email addresses) and we will use this information to invite teachers to participate in our evaluation activities in the form of

•Baseline teacher survey data collection (Jan 2019)

•Post-intervention teacher survey (June – July 2019)

•An interview with the Deeper Thinking lead (up to 1 hr)

•A focus group or paired interview with Year 10 science teachers (up to 1 hr)

•A classroom observation (up to 1 hr or as determined by the school).

You have been asked to participate as you are a Year 10 science teacher in a school piloting the Deeper Thinking intervention.

Who is collecting data?

The Pilot intervention is being led by Carmel Education Trust (CET) and evaluated by AlphaPlus Consultancy Ltd.

CET will be the originating Data Controller for any teacher data they disclose to the parties under this project.

AlphaPlus will be the Data Controller in respect of any personal data of teachers which they process for the purposes of the project.

The parties rely on the below processing condition to process personal data under this agreement: The processing is necessary for purposes of legitimate interests (article 6(1)(f)).

What data is being collected?

- Teacher names
- School name and exam board for GCSE science
- Teacher email addresses and phone numbers

We will only use the data to contact you in advance of evaluation activities and to send online survey requests. Your personal information will not be used in our data analysis and we will not identify any individual in our findings.

The data will not be matched to other data sets or used for any automated decision-making or profiling.

How is the data stored?

Any data provided to us will be stored securely according to our data security policy. Only authorised project members will be able to access the data, which will be kept on a secure, password protected server. Any data transfers will use encryption

AlphaPlus will destroy all personal data associated with this project, including data sets received from the Data Controllers.

Contact

If you have any questions about the evaluation, please contact Clare Dowland at AlphaPlus on 01962 840362 or clare.dowland@alphaplus.co.uk

Appendix 4: Theory of Change template

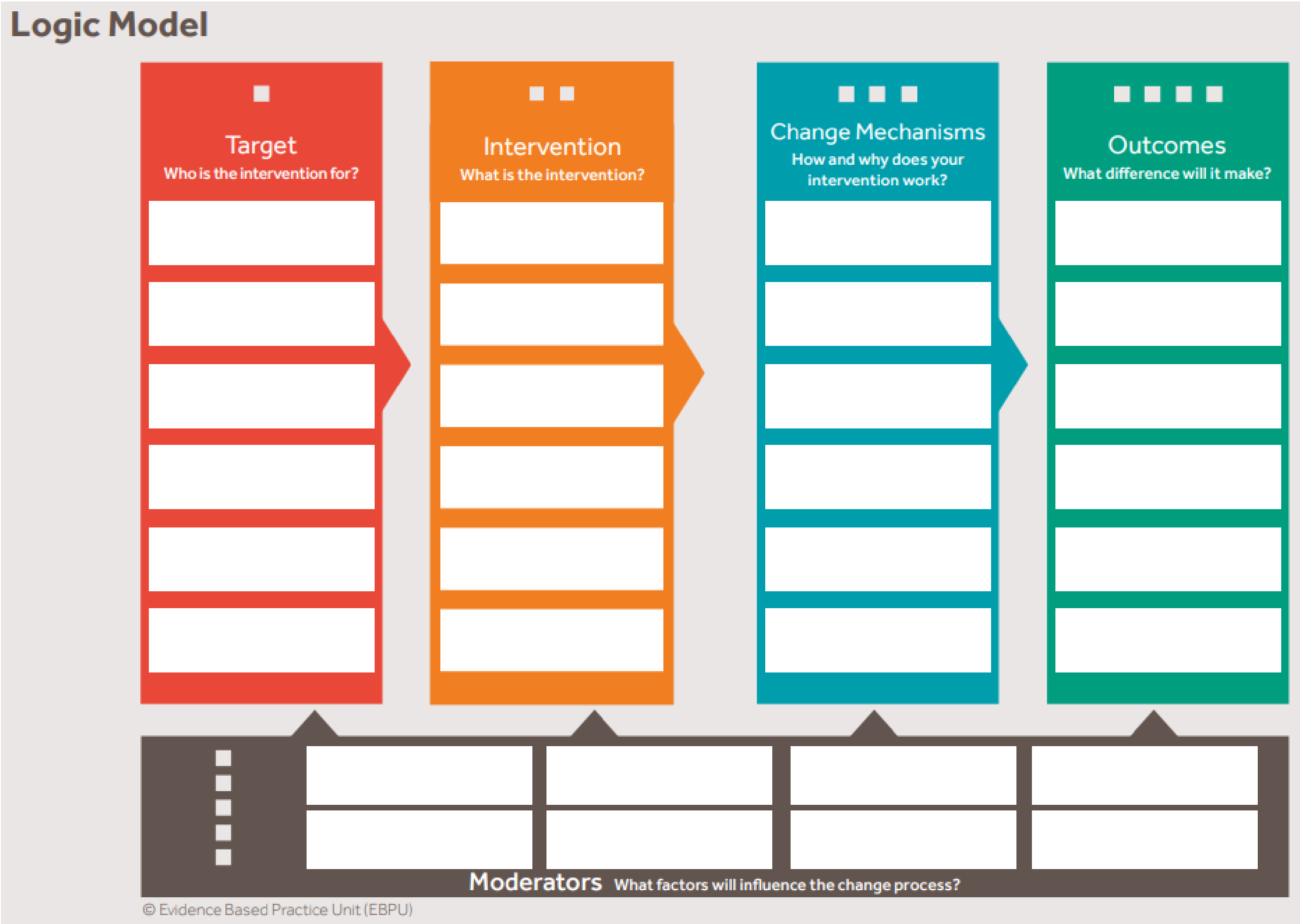


Figure 6: Evidence Based Practice Unit template used in the Theory of Change workshop

Appendix 5: Interview schedule for previous users of the Deeper Thinking Intervention

Quick introduction

Thank you for taking the time to speak to me

This conversation will be confidential, so I will not share anything you say with David or anyone at Carmel. I will simply use it inform the survey questions I am developing.

Is it OK if I record this conversation so I can concentrate on what you are saying rather than writing notes?

Main questions

I am very keen to hear how you found using the Deeper Thinking Intervention, what went well, what was difficult anything you would have changed.

To begin with please can you tell me why the school decided to take part in the pilot?

Activities

How long did it run for and what did you do over this period?

- Hexagon mind mapping
- Mark scheme production?
- Applied to exams? Practical assessments?

How did the teachers find the delivery of the intervention?

- What went well?
- What was more challenging?

Impact

What are your thoughts on whether the intervention had an impact on the students and staff taking part:

- Students? Please provide an example
- Did it impact on all students equally?
- Teachers delivering the intervention?

Final question

The intervention aims to impact on students in variety of ways it would be good to sense check these with you

- Improving student's ability to recognise existing knowledge and identify gaps
- Make more links between relevant knowledge
- Develop their communication skills while working in groups on the hexagon tasks
- Become more sensitive to the language of assessment and the language required in answers

Appendix 6: January Teacher Survey

The content of the online pre-pilot survey is reproduced below.

Solo baseline survey

Thank you for taking the time to complete this survey, your contribution is extremely valuable. This survey is confidential and is being collected for evaluation purposes only. It will not be shared beyond the evaluation team. Neither you nor your school will be identified in reports produced in relation to this study. By completing the survey you consent to your data being used for the evaluation. You can withdraw at any point by contacting Kathy Seymour at AlphaPlus (kathy.seymour@alphaplus.co.uk).

About your school

Does your school offer GCSE triple science?

- ❍ Yes
- ❍ No

About you

Do you teach year 10 science?

- ❍ Yes
- ❍ No

If No > skip to end of survey with message: Thank you for taking part, the survey applies to Year 10 science teachers so we do not need to ask you any further questions.

What is your role at the school? (select as many as apply)

- ❑ Head of science/head of department or equivalent
- ❑ Teacher
- ❑ A member of the school leadership team

Which science subjects do you teach to year 10? (select as many as apply)

- ❑ GCSE Combined Science - Synergy Foundation
- ❑ GCSE Combined Science - Synergy Higher
- ❑ GCSE Combined Science - Trilogy Foundation
- ❑ GCSE Combined Science - Trilogy Higher
- ❑ GCSE Chemistry
- ❑ GCSE Physics
- ❑ GCSE Biology

Did you study at degree level the science subject you are currently teaching?

- ❍ Yes
- ❍ No

Your year 10 science students

Do you teach the following Year 10 science groups?

- ❑ High ability (Grade 9 - 7 | A* - A)
- ❑ Medium ability (Grade 6 - 4 | B - C)

- ❑ Lower ability (Grade 3 - U | D - U)

For each of the ability groups they indicate they teach, the following set of questions was asked:

Considering the performance of your Year 10 science students over the past three months. To what extent do you agree or disagree with the following statements
My high/medium/low ability year 10 science students:

	Disagree Strongly	Disagree	Slightly Disagree	Slightly Agree	Agree	Agree Strongly	Not sure/ prefer not to say
If asked, recognise their existing science knowledge	○	○	○	○	○	○	○
If asked, identify gaps in their existing science knowledge	○	○	○	○	○	○	○
If asked, make links between different pieces of science knowledge	○	○	○	○	○	○	○
Communicate well when carrying out group work	○	○	○	○	○	○	○
Understand the wording of assessment questions	○	○	○	○	○	○	○
Understand the language they need to use when answering exam questions	○	○	○	○	○	○	○
Are engaged in science lessons	○	○	○	○	○	○	○
Engage in continual self-assessment of their science knowledge	○	○	○	○	○	○	○
Are confident in their own science ability	○	○	○	○	○	○	○

Science teaching

Considering your Year 10 science teaching over the past three months…
Did you ask students to analyse the specific wording of exam or assessment questions?

- ❍ No

- ❍ Yes - every lesson
- ❍ Yes - every week
- ❍ Yes - every month
- ❍ Yes - once or twice over the last three months
- ❍ Prefer not to say

If you did this, please provide any additional comments, for example: What did you ask the students to do? How did the students find the task? Did it have an impact on students?

____

____

____

Considering your Year 10 science teaching over the past three months…
Did you discuss mark schemes with students??

- ❍ No
- ❍ Yes - every lesson
- ❍ Yes - every week
- ❍ Yes - every month
- ❍ Yes - once or twice over the last three months
- ❍ Prefer not to say

If you did this, please provide any additional comments, for example: What did you ask the students to do? How did the students find the task? Did it have an impact on students?

____

____

____

How would you rate your confidence as a year 10 science teacher?

- ❍ Extremely unconfident
- ❍ Unconfident
- ❍ Slightly unconfident
- ❍ Average
- ❍ Slightly confident
- ❍ Confident
- ❍ Extremely confident
- ❍ Not sure/ prefer not to say

Please rate the following learning techniques in terms of their usefulness to year 10 science students (we do not expect you to be using every technique so feel free to rate them based on their expected usefulness)

	0= Not at all useful	10 = Extremely useful	❑ Not applicable

	0 1 2 3 4 5 6 7 8 9 10
Elaborative interrogation: explaining why a fact/concept is true	
Practice testing: self-testing or taking practice tests of to-be-learned materials	
Self-explanation: explaining how new information is related to already known information	
Summarisation: write summaries of to-be-learned text	
Highlighting/underlining: marking potentially important portions of texts while reading	
Distributed practice: a schedule of practice that spreads out study activities over time	
Rereading: restudying text material again	
Key word memory techniques: using keywords to aid the recall of information	
Imagery for text: forming mental images of text materials	
Interleaved practice: a schedule of practice that jumps between alternating topics	

If you have any further comments on learning techniques or anything in this survey please add them below

Appendix 7: June Teacher Survey

The content of the online post-pilot survey is reproduced below.

Solo follow-up survey

Thank you for taking the time to complete this survey, your contribution is extremely valuable. This survey is confidential and is being collected for evaluation purposes only. It will not be shared beyond the evaluation team. Neither you nor your school will be identified in reports produced in relation to this study. By completing the survey you consent to your data being used for the evaluation. You can withdraw at any point by contacting Kathy Seymour at AlphaPlus (kathy.seymour@alphaplus.co.uk).

About you

Do you teach year 10 science?

- ❍ Yes
- ❍ No

If No > skip to end of survey with message: Thank you for taking part, the survey applies to Year 10 science teachers so we do not need to ask you any further questions.

What is your role at the school? (select as many as apply)

- ❑ Head of science/head of department or equivalent
- ❑ Teacher
- ❑ A member of the school leadership team

Which science subjects do you teach to year 10? (select as many as apply)

- ❑ GCSE Combined Science - Synergy Foundation
- ❑ GCSE Combined Science - Synergy Higher
- ❑ GCSE Combined Science - Trilogy Foundation
- ❑ GCSE Combined Science - Trilogy Higher
- ❑ GCSE Chemistry
- ❑ GCSE Physics
- ❑ GCSE Biology

Did you study at degree level the science subject you are currently teaching?

- ❍ Yes
- ❍ No

Your year 10 science students

Do you teach the following Year 10 science groups?

- ❑ High ability (Grade 9 - 7 | A* - A)
- ❑ Medium ability (Grade 6 - 4 | B - C)
- ❑ Lower ability (Grade 3 - U | D - U)

For each of the ability groups they indicate they teach, the following set of questions was asked:

Considering the performance of your Year 10 science students over the past three months. To what extent do you agree or disagree with the following statements
My high/medium/low ability year 10 science students:

	Disagree Strongly	Disagree	Slightly Disagree	Slightly Agree	Agree	Agree Strongly	Not sure/ prefer not to say
If asked, recognise their existing science knowledge	○	○	○	○	○	○	○
If asked, identify gaps in their existing science knowledge	○	○	○	○	○	○	○
If asked, make links between different pieces of science knowledge	○	○	○	○	○	○	○
Communicate well when carrying out group work	○	○	○	○	○	○	○
Understand the wording of assessment questions	○	○	○	○	○	○	○
Understand the language they need to use when answering exam questions	○	○	○	○	○	○	○
Are engaged in science lessons	○	○	○	○	○	○	○
Engage in continual self-assessment of their science knowledge	○	○	○	○	○	○	○
Are confident in their own science ability	○	○	○	○	○	○	○

Science teaching

Considering your Year 10 science teaching over the past three months…
Did you ask students to analyse the specific wording of exam or assessment questions?

- ❍ No
- ❍ Yes - every lesson
- ❍ Yes - every week

- Yes - every month
- Yes - once or twice over the last three months
- Prefer not to say

Considering your Year 10 science teaching over the past three months…
Did you discuss mark schemes with students??

- No
- Yes - every lesson
- Yes - every week
- Yes - every month
- Yes - once or twice over the last three months
- Prefer not to say

How would you rate your confidence as a year 10 science teacher?

- Extremely unconfident
- Unconfident
- Slightly unconfident
- Average
- Slightly confident
- Confident
- Extremely confident
- Not sure/ prefer not to say

Student learning techniques

Please rate the following learning techniques in terms of their usefulness to year 10 science students (we do not expect you to be using every technique so feel free to rate them based on their expected usefulness)

	0= Not at all useful	10 = Extremely useful	❑ Not applicable
	0 1 2 3 4 5 6 7 8 9 10		
Elaborative interrogation: explaining why a fact/concept is true			
Practice testing: self-testing or taking practice tests of to-be-learned materials			
Self-explanation: explaining how new information is related to already known information			
Summarisation: write summaries of to-be-learned text			
Highlighting/underlining: marking potentially important portions of texts while reading			
Distributed practice: a schedule of practice that spreads out study activities over time			
Rereading: restudying text material again			
Key word memory techniques: using keywords to aid the recall of information			
Imagery for text: forming mental images of text materials			
Interleaved practice: a schedule of practice that jumps between alternating topics			

Reflections on the Deeper Thinking Intervention

Training

To what extent was the training you received helpful for delivering the intervention to your Year 10 students?

- ❍ Very unhelpful
- ❍ Unhelpful
- ❍ Helpful
- ❍ Very helpful
- ❍ Prefer not to say
- ❍ I didn't attend all the training sessions

Please use the space below to suggest any changes to the training which could maximise teacher engagement in the intervention

____

____

____

Implementation

How many Year 10 Science groups/classes have you regularly taught between January and June 2019?

- ❍ 1
- ❍ 2
- ❍ 3
- ❍ 4
- ❍ 5 or more
- ❍ NA / prefer not to say

How many times, if at all, have you used the Deeper Thinking techniques with your Year 10 students (since the beginning of the pilot)? Please select zero rather than leave it blank if you haven't used the technique.

Extended answer questions:

	0	1	2	3	4	5	6	7	8 or more	NA / prefer not to say
Analysing extended answer exam questions	○	○	○	○	○	○	○	○	○	○
Applying to new contexts	○	○	○	○	○	○	○	○	○	○
Constructing a mark scheme	○	○	○	○	○	○	○	○	○	○
Writing own questions	○	○	○	○	○	○	○	○	○	○

Practical work:

	0	1	2	3	4	5	6	7	8 or more	NA / prefer not to say
Analysing extended answer exam questions	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Applying to new contexts	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Constructing a mark scheme	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Writing own questions	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

Analysing analysis questions:

	0	1	2	3	4	5	6	7	8 or more	NA / prefer not to say
Analysing extended answer exam questions	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Applying to new contexts	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Constructing a mark scheme	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
Writing own questions	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

In terms of when you use the Deeper Thinking classroom packs, how often do you..?

	Always	More than half of the time	Half of the time	Less than half of the time	Never	NA / prefer not to say
Use the Solo placemat but not the hexagons	○	○	○	○	○	○
Use the hexagons but not the Solo placemat	○	○	○	○	○	○
Use both the Solo placemat and the hexagons together	○	○	○	○	○	○

Which aspects of using the Deeper Thinking techniques, if any, went well?

____

____

____

Which aspects of using the Deeper Thinking techniques, if any, were challenging?

____

____

____

Before taking part in the Deeper Thinking pilot were you engaging in similar activities with your Year 10 science students?

- ❍ Yes
- ❍ No
- ❍ Not sure / prefer not to say
- ❍ Not applicable

If yes, please describe what you were already doing

____

____

____

Impact

Please use the space below to provide any suggestions or thoughts on how the intervention can maximise transfer to the exam

____

____

____

Please use the space below to describe any impacts (positive or negative) that you have noticed as a result of the Deeper Thinking pilot, or to provide any further comments on the intervention overall

____

____

____

Appendix 8: CPD Evaluation form questions

The quality of the professional learning overall

How well the learning outcomes were met

The quality of the presenter

How relevant to my practice

How interesting/enjoying I found this professional learning

Response options:

Very Good, Good, Satisfactory, Poor.

Appendix 9: Observation template

Number of learners in the class: ______
Topic: ___________________________
Ability ___________________________

Learners	
Signs of engagement	
Are they working on the task in groups as directed?	
Do they appear to understand the task?	
Do they ask questions?	
Do they maintain momentum?	

Do some students appear to be struggling?	

Teacher	
How easy is the intervention to deliver?	
Facilitating Factors/ barriers	

Any unintended consequences for teachers or learners?

Anything else that might be useful

Appendix 10: Interview schedule for Deeper Thinking Lead Teacher

House-keeping

Thank you for taking the time to speak to me

Introductions

- Thank you for taking the time to speak to me.
- I am a senior researcher at A+. We have been commissioned by EEF to independently evaluate the Deeper Thinking pilot.
- I am independent from both EEF and Carmel and neither know which schools I am visiting for the case studies.
- The overall results will be written-up but no individual or school will be identified in the report.
- The school is not being evaluated. I am simply trying to build a picture of how this the programme pans out real world.
- It would be great to get an overarching senior view of the programme as implemented in your school.
- Do you have any questions before we begin?
- Please can I record this discussion? This is so I can focus on what you say rather than writing it down?
- AP: Set recorder running and say ‘This is recording X for case study number X’- don’t say the school name

Main questions

I am very keen to hear how your science department is finding using the Deeper Thinking Intervention with Year 10.

Is the programme attractive to schools?

To begin with please can you tell me why your school decided to take part in the pilot?

Have there been any costs to the school in terms of setting it up?

- Time
- Materials

Training

How did you find the first training session (the one-day event at Carmel?

- Leadership day
- Which bits were the most useful?
- Were there any parts you found less helpful?

What about the 3 twilight teacher sessions delivered in the school?

- Which bits were the most useful?
- Were there any parts you found less helpful?
- Was there anything missing?

How do you think the CPD sessions could maximise teacher engagement with the intervention?

Overall, was the training and guidance provided appropriate for your school?

Activities

What sort of activities have taken place in the school so far?

- Hexagon mind mapping
- Mark scheme production
- Practical work

*note we pick up impact in the next section

How often are teachers using it?

- Is this in line with what you expected based on the training?
- Do you have a sense of how what the ideal frequency of use might look like?

Across the team of science teachers how do you think they have found delivering the intervention?

- What went well?
- What was more challenging?
- Has there being any variation between teachers?

How does the implementation vary by the 3 areas it aims to tap into, these are:

- Science practicals
- Analysis questions
- Extended answer questions?

Are there any other facilitating or hindering factors that could be identified in terms of implementation?

Impact

What are your thoughts on whether the intervention had an impact on the students and staff taking part:

Teachers

- Has science teaching practice changed since taking part in the pilot?
- Have you noticed any differences between teachers?

Students

- Have you had any feedback from students?
- Do you think it is having an impact on learning/ behaviour?
- Please provide an example
- Do some groups of students benefit more than others? For example, thinking about ability group…

Have there been any unintended consequences from taking part?

Final question

Would you recommend the Deeper Thinking intervention to other schools? (do you think it would be attractive to other schools?)

Before we wrap up, is there anything else you would like to add about your schools experience?

Thankyou…

Appendix 11: Focus Group Schedule for Deeper Thinking Teachers

House-keeping

Thank you for taking the time to speak to me

Introductions

- Thank you for taking the time to speak to me.
- I am a senior researcher at A+. We have been commissioned by EEF to independently evaluate the Deeper Thinking trial.
- I am independent from both EEF and Carmel and neither know which schools I am visiting for the case studies.
- The overall results will be written-up but no individual or school will be identified in the report.
- The school is not being evaluated. I am simply trying to build a picture of how this the programme pans out real world.

- It would be great to hear how the programme is being implemented in your school.
- Does anyone have any questions before we begin?

- Please can I record this discussion so I can focus on what you say rather than writing it down?

- AP: Set recorder running and say ‘This is recording X for case study number X’- don’t say the school name

Main questions

I am very keen to hear how you are finding using the Deeper Thinking Intervention.

Training

To begin with please can you tell me how you found the training sessions you attended?

- Twilight teacher sessions (check how many)
- Any extra internal training?

Was there anything the Deeper Thinking team could do to the CPD sessions maximise teacher engagement?

Activities

How often have you used the Deeper Thinking techniques with your class so far?

- Hexagon mind mapping
- Mark scheme production
- Practical assessments
- Applied to mock extended answer exam questions

How did you find it?

- What went well?
- What was more challenging?
- Is it going as you expected?

How does the implementation vary by the 3 areas it aims to tap into, these are:

- Science practicals
- Analysis questions
- Extended answer questions?

Did the training prepare you properly for the implementation?

What about the Carmel portal? How are you finding that aspect?

Impact

What are your thoughts on whether the intervention has had an impact your teaching practice

- Please provide an example

How about your students? Do you think the Deeper Thinking Intervention is having an impact on Year 10 Science student behaviour?

- Please provide an example
- Did it impact on all students equally?

Were there any unintended consequences taking part in this pilot?

Do you think students will be able to apply Deeper Thinking techniques in the final exam situation?

- How do you think the intervention can maximise transfer to the exam?

Final question

Would you recommend the Deeper Thinking intervention to other teachers? (do you think it would be attractive to other schools?)

Before we wrap up, is there anything else you would like to add about your experience of taking part?

Thankyou…

Appendix 12: Survey Tables

The January Teacher survey was sent to 90 teachers and the June Teacher Survey was sent to 83 teachers. 30 teachers responded to both.

Table 11: Counterfactual activities

	No	Yes	Not sure/ prefer not to say	Total
Before taking part in the Deeper Thinking pilot were you engaging in similar activities with your Year 10 science pupils?	55% (21)	39% (15)	5% (2)	100% (38)

Source: Deeper Thinking June Teacher Survey. N=38 (all teachers who completed the June Survey).

Table 12: Frequency of analysing the wording of exam or assessment questions pre and post pilot.

Did you ask pupils to analyse the specific wording of exam or assessment questions?	Pre-pilot % (n)	Post-pilot % (n)
No	3% (1)	0% (0)
Yes - every lesson	10% (3)	13% (4)
Yes - every week	33% (10)	37% (11)
Yes - every month	37% (11)	30% (9)
Yes - once or twice over the last three months	17% (5)	20% (6)
Total	100% (30)	100% (30)

Source: Deeper Thinking January and June Teacher Surveys. N=30 (all the teachers responded to both surveys).

Table 13: Frequency of discussing mark schemes with pupils pre and post pilot.

Did you discuss mark schemes with pupils?	Pre-pilot % (n)	Post-pilot % (n)
No	3% (1)	0% (0)
Yes - every lesson	3% (1)	7% (2)
Yes - every week	30% (9)	50% (15)
Yes - every month	47% (14)	37% (11)
Yes - once or twice over the last three months	17% (5)	7% (2)
Total	100% (30)	100% (30)

Source: Deeper Thinking January and June Teacher Surveys. N=30 (all the teachers responded to both surveys).

Table 14: Extent to which teachers agree or strongly agree with statements about their Year 10 science pupils' perceived ability to identify and link their scientific knowledge.

My Year 10 science pupils…	% agree or agree strongly pre-pilot (n)	% agree or agree strongly post-pilot (n)	Percentage point change between pre- and post-pilot
If asked, recognise their existing science knowledge	41% (26)	44% (28)	3
If asked, identify gaps in their existing science knowledge	23% (15)	34% (22)	11
If asked, make links between different pieces of science knowledge	16% (10)	34% (22)	18

Source: Deeper Thinking January and June Teacher Surveys. N=30 (all the teachers responded to both surveys). This question was asked by pupil ability grouping (low- grades 3 to U, medium- grades 4-6 and high grades 9-7) but because the sample size was small the pupil groups were combined. Teachers taught between 1 and 3 groups meaning the n for each category is variable. The base for the percentages is the number of respondents who answered the question with reference to any of the three ability groups in each survey.

Table 15: Extent, to which teachers agree or strongly agree with statements about their Year 10 science pupils perceived ability to communicate in a group and understand the language used in assessment questions and answers.

My Year 10 science pupils…	% agree or agree strongly pre-pilot	% agree or agree strongly post-pilot	Percentage point change between pre- and post-pilot
Communicate well when carrying out group work	44% (28)	48% (31)	4
Understand the wording of assessment questions	22% (14)	22% (14)	0
Understand the language they need to use when answering exam questions	17% (11)	31% (20)	14

Source: Deeper Thinking January and June Teacher Surveys. N=30 (all the teachers responded to both surveys). This question was asked by pupil ability grouping (low- grades 3 to U, medium- grades 4-6 and high grades 9-7) but because the sample size was small the pupil groups were combined. Teachers taught between 1 and 3 groups meaning the n for each category is variable. The base for the percentages is the number of respondents who answered the question with reference to any of the three ability groups in each survey.

Table 16: Extent, to which teachers agree or strongly agree with statements about their Year 10 science pupils perceived engagement in science.

My Year 10 science pupils…	% agree or agree strongly pre-pilot	% agree or agree strongly post-pilot	Percentage point change between pre- and post-pilot
Are engaged in science lessons	59% (38)	53% (33)	-6
Engage in continual self-assessment of their science knowledge	14% (9)	25% (16)	11
Are confident in their own science ability	14% (9)	22% (14)	8

Source: Deeper Thinking January and June Teacher Surveys. N=30 (all the teachers responded to both surveys). This question was asked by pupil ability grouping (low- grades 3 to U, medium- grades 4-6 and high grades 9-7) but because the sample size was small the pupil groups were combined. Teachers taught between 1 and 3 groups meaning the n for each category is variable. The base for the percentages is the number of respondents who answered the question with reference to any of the three ability groups in each survey.

Table 17: Post-implementation appraisal of training 'To what extent was the training you received helpful for delivering the intervention to your Year 10 pupils?'

	Very helpful	Helpful	Unhelpful	Very unhelpful	Prefer not to say	Total
Teacher rating	19% (7)	58% (21)	14% (5)	3% (1)	6% (2)	100% (36)

Source: Deeper Thinking June Teacher Survey

Table 18: Frequency of teacher use of Deeper Thinking techniques

Learning Technique* and associated Dunlosky, et al. (2013) rating.	Pre-pilot teacher rating of usefulness: mean (SD) and (n)	Post-pilot teacher rating of usefulness mean (SD) and (n)
Practice testing (high)	7.5 (1.7) (n=29)	7.4 (2.0) (n=30)
Distributed practice (high)	7.0 (2.3) (n=27)	6.8 (2.2) (n=29)
Elaborative interrogation (moderate)	7.4 (1.8) (n=29)	7.0 (1.9) (n=30)
Self-explanation (moderate)	7.4 (1.6) (n=28)	6.8 (1.8) (n=30)
Interleaved practice (moderate)	7.2 (2.8) (n=27)	7.1 (2.4) (n=28)
Summarization (low)	6.0 (2.3) (n=28)	5.2 (2.5) (n=30)
Highlighting/underlining (low)	5.5 (3.1) (n=27)	4.7 (2.7) (n=29)
Keyword memory techniques (low)	7.9 (1.7) (n=29)	6.6 (1.9) (n=30)
Imagery for text (low)	6.6 (2.6) (n=27)	5.9 (2.4) (n=28)
Rereading (low)	5.4 (2.6) (n=28)	5.2 (2.4) (n=26)

*Learning techniques were presented to teachers in a randomised order to avoid presentation effects.

Source: Deeper Thinking January and June Teacher Surveys. 30 teachers responded to both surveys.

Table 19: Frequency of teacher use of Deeper Thinking techniques

How many times, if at all, have you used the Deeper Thinking techniques with your Year 10 pupils?	% of teachers who used the technique at least once (n)*	Mean times used per teacher during the pilot (if used =>1) (s.d)
Extended answer questions*	100% (38)	7.3 (5.5)
• **Analysing extended answer exam questions**	95% (36)	3.4 (2.3)
• **Applying to new contexts**	74% (28)	2.7 (1.7)
• **Constructing a mark scheme**	50% (19)	2.1 (1.6)
• **Writing own questions**	47% (18)	2.2 (1.6)
Practical work*	84% (32)	5.3 (4.0)
• **Analysing the exam questions relating to the required practicals**	82% (31)	2.6 (2.0)
• **Applying to new contexts**	58% (22)	2.1 (1.2)
• **Constructing a mark scheme for required practical exam questions**	39% (15)	2.2 (1.7)
• **Writing own required practical exam questions**	13% (5)	1.8 (0.8)
Analysis questions*	76% (29)	7.0 (5.3)
• **Analysing analysis questions**	74% (28)	3.5 (2.2)
• **Applying to new contexts**	53% (20)	3.0 (2.0)
• **Constructing a mark scheme for analysis exam questions**	34% (13)	2.2 (2.0)
• **Writing own required analysis exam questions**	18% (7)	2.1 (1.2)

* The figures quoted for the three overall technique categories show the number of teachers (and % of the overall respondent base of 38) who used at least one of the techniques in that category at least once during the pilot. The means show the average number of times these teachers used any of the four techniques during the pilot.

Source: Deeper Thinking June Teacher Survey. N=38. Note that in the survey question, the highest response category was '8 or more' times. In order to calculate a mean, this category was treated as a value of 8, therefore the means could be slightly under-estimated if anyone who ticked '8 or more' had in reality used it more than 8 times. Valid ns for the mean calculations are the same as the ns quoted in brackets in the middle column of the table.

Table 20: Using the Deeper Thinking classroom packs

'In terms of when you use the Deeper Thinking classroom packs, how often do you...?	Always	More than half of the time	Half of the time	Less than half of the time	Never	Total (n)
Use both the SOLO placemat and the hexagons together	11% (4)	29% (10)	11% (4)	29% (10)	20% (7)	100% (35)

Source: Deeper Thinking June Teacher Survey.

This document is available for download at https://educationendowmentfoundation.org.uk

Education Endowment Foundation

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn