

PROJECT TITLE	Evaluation of pilot of the Deeper Thinking intervention
DEVELOPER (INSTITUTION)	Carmel Education Trust
EVALUATOR (INSTITUTION)	AlphaPlus Consultancy
PRINCIPAL INVESTIGATOR(S)	Andrew Boyle and Hayley Limmer
STUDY PLAN AUTHOR(S)	Andrew Boyle and Hayley Limmer
PUPIL AGE RANGE AND KEY STAGE	KS4 Year 10
NUMBER OF SCHOOLS/ SETTINGS	12
NUMBER OF PUPILS	Relative to the size of the Year 10 year group

Study plan version history

VERSION	DATE	REASON FOR REVISION
1.0	01/11/18	

Intervention

The SOLO Taxonomy and Deeper Thinking intervention aims to improve pupil outcomes on two areas of GCSE assessment – the assessed practical tasks and the extended answer questions (those worth four marks or more) in GCSE exams. It aims to do this by using the SOLO Taxonomy to help pupils assess their current understanding of a topic at the start of a lesson or learning sequence, pupils then use a structured mind mapping approach to make links between key vocabulary as they are learning a topic and then assess their understanding against the SOLO Taxonomy at the end of the lesson or learning sequence.

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 also structure their extended answer questions so that they include all of 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 sessions (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.

1. Brief name: Deeper Thinking: Revision programme for improving extended answer questions and 'required' practicals in GCSE Science.

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 and questions around 'required' practicals (typically 4-6 marks). 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 don't attempt extended answer questions. Those who attempt the questions often don't 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. Our 'Deeper Thinking' approach is based on approaches that work together to improve these responses.

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

4. What (materials): Participants will receive materials to support the approaches, including templates and guidance documentation.

5. What (procedures): The intervention involves training teachers in the use of approaches that are identified as moderate or high utility by Dunlosky et al. (i.e. practice questions, interleaved practice, distributed practice, elaborated interrogation, self-explanation), the SOLO taxonomy identified by Biggs and Collis and concept-mapping to support the connection of ideas. The programme involves training teachers to use these approaches over time. Training for in-school leaders/coordinators will take place to lead the implementation in school. Catch up training will take place in year 2 to accommodate changes of staffing.

6. Who: Science teachers and teaching assistants take part in delivering the intervention to pupils. Science leaders coordinate the intervention in school.

7. How (mode of delivery): A programme will be implemented in schools to help pupils develop reasoning skills in their work around 'required' practicals in lessons linked to these. Teachers will apply their training, using the approach with pupils to identify variables/factors that are relevant and

that may confound, to identify relationships and to transfer their learning to other similar contexts to help them answer unfamiliar questions. When completing extended answer questions, teacher will apply their training to support pupils to analyse questions. Through this, pupils will develop a greater understanding of success criteria around extended answer questions, leading them to be more secure in peer and self-assessment and self-remediation.

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

9. When and how much (dosage): Pupils receive the intervention approximately once every 2-4 weeks around their work in class relating to the required practicals and practice examinations. The intervention lasts two years. Twilight training takes place over terms 1 and 2 and implementation can begin after the second training session. Leadership training takes place in year 1.

10. Tailoring: Schools will fit the programme around their usual work in class relating to the required practicals and practice examinations. This will be subject to basic effective practice being followed (i.e. the practicals will be distributed throughout the course and practice of extended answer questions will take place at least half-termly)

Research questions

Evidence to support theory of change/evidence of promise (Is there evidence of expected change happening?)

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 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? (to be refined in the ToC workshop but may include the following examples: spreading reviewing over time, increasing peer and self-assessment).
6. Is the intervention sufficiently innovative relative to the counter-factual? (Given the absence of a control group teachers will be asked to reflect on differences compared to the previous year of teaching.)

Feasibility (Does delivery happen as intended? What does ideal delivery look like?)

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 3 areas practical's, 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 / readiness for trial (is it replicable and affordable?)

9. How suitable is the Deeper Thinking intervention to progress to efficacy trial?
 - Is it clearly defined and scaleable?
 - 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 (i.e. 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?

Methods

Recruitment

SCHOOLS

Up to 12 secondary schools in North East will be recruited. Carmel have expressions of interest from around 9 already, with a mix of sizes, performance and type.

It is important that the sample covers a range of schools in order that different school level factors affecting implementation and ideal delivery conditions can be understood.

As part of the recruitment Carmel will monitor publicly available information published by the DfE regarding the make-up of school science departments including:

- Progress 8 score
- Pupils eligible for free school meals at any time during the past 6 years
- Pupils whose first language is not English
- Ofsted rating

The baseline teacher survey will collect more detailed information:

- Role at school
- Number of hours dedicated to science teaching in Y10 per fortnight
- Specialist science teacher
- Science related degree

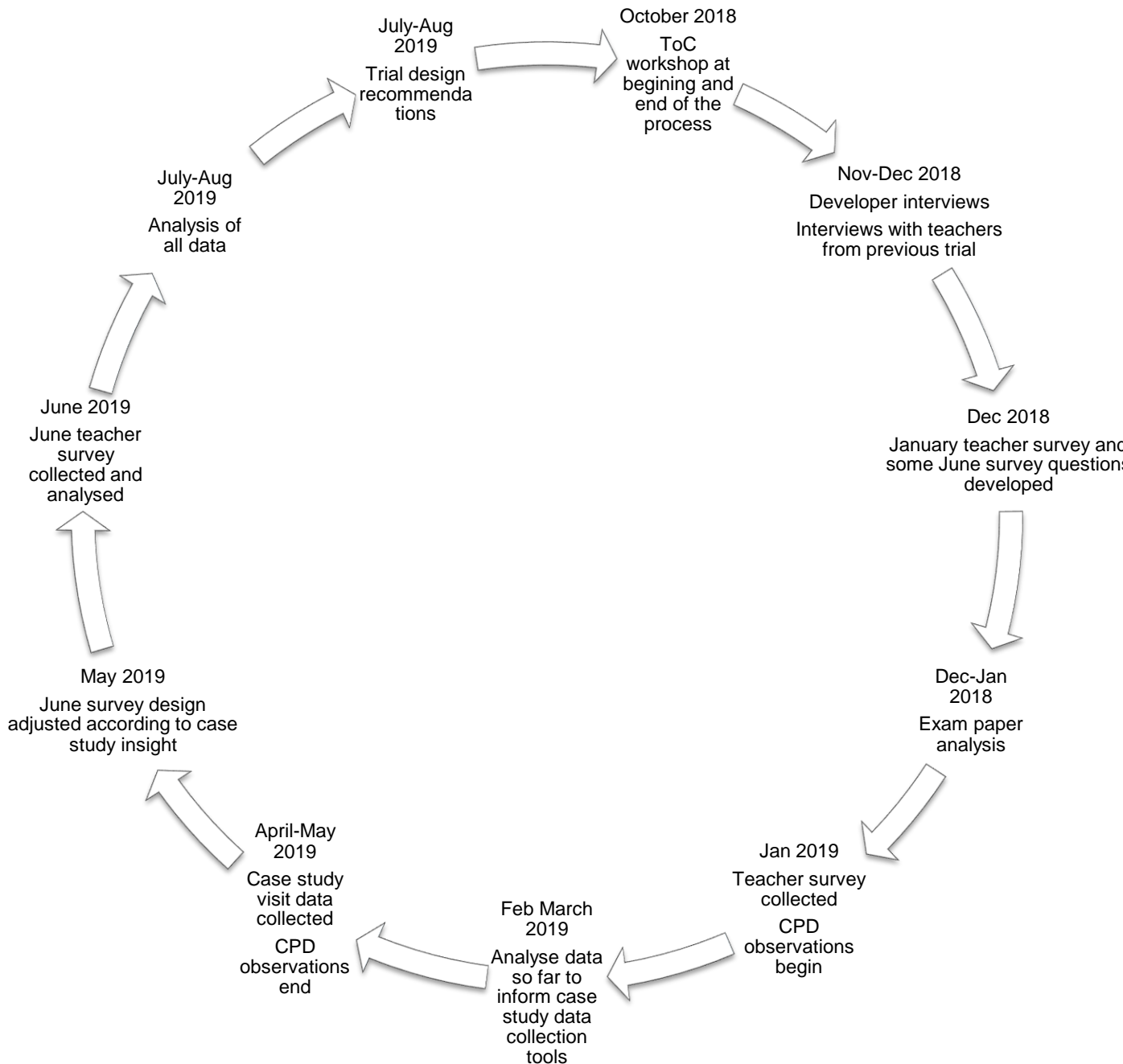
PUPILS

Within schools taking part all Year 10 pupils will be exposed to the intervention. No data will be collected directly from pupils.

Data collection

We will collect data drawing on the ToC, to test it in situ and on the basis of the data collected, we will refine and develop this theory. We will carry out qualitative and quantitative activities to gain an in-depth yet broad picture of the intervention. We will capture changes to teaching and pupil learning behaviour whilst documenting factors that teachers perceive to facilitate and hinder implementation.

Figure 1: High level summary of the data collection



Theory of change (ToC) workshop

We propose holding a face-to-face workshop with Carmel to draw out key assumptions held about the Deeper Thinking product and how it should be delivered. This workshop will follow the EEF template provided by the Evidence Based Practice Unit¹ and develop a logic model by systematically examining:

- The target- who is the intervention for.
- The intervention-what exactly the intervention involves.
- Change Mechanisms- how will the intervention lead to the outcomes that the developers expect. The developers will be 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.
- Moderators- a detailed discussion identifying the factors that will influence whether the intervention leads to the outcomes the developers identify.

The workshop will build a logic model scrutinise it for coherence, checking it connects together properly and conclude by developing the measures that will be applied to each element. Given it is a relatively small intervention we will focus on outcomes and assigning an appropriate unit of analysis. The discussion at set-up suggested that the intervention may have a subtle impact so selecting appropriate measures is expected to take considerable time.

In this workshop, we will address specific questions like:

- Precisely what does Carmel mean by 'deeper thinking'?
- How is the approach distinctive from other attempts to improve pupils' thinking skills, meta-cognition and awareness of exam marking criteria?
- What do developers envisage are appropriate teaching and learning activities that flow from their intervention?
 - Do they want to train pupils to interpret mark schemes in detail? If so, does this represent a genuine learning aim, or is it merely 'exam coaching'?
- In the development team's opinion, what outcome measure is appropriate to show benefit both overall and within the pilot phase?
 - How will teacher feedback be classified and what will be classed as a success?
 - Is it GCSE science (as a whole), and – if so – is the improvement in deeper thinking going to be sufficient to increase GCSE grades?
 - If some other measure is the intended outcome (e.g. performance on extended answer questions – EAQs) and/or questions using knowledge from practicals, is this reasonable?
 - If EAQs and practicals-derived questions are only a small part of new specification GCSEs, can Solo still be seen as an intervention that improves pupils' science learning?

By the end of this workshop, developers and evaluators should have a shared understanding of the inputs, outputs and intended outcomes of the intervention and the pilot phase. This ToC will underpin the evaluation and evidence collected will be scrutinised in relation to this document. As such it will be a key input into decisions regarding viability for trial.

¹ M Wolpert et al., 'EBPU Logic Model' (London: Evidence Based Practice Unit, n.d.), https://www.ucl.ac.uk/evidence-based-practice-unit/sites/evidence-based-practice-unit/files/pub_and_resources_resrouces_for_profs_logic_model_2017.pdf.

Telephone interviews with teachers and the developer

The mid academic year start-up date (Jan 2019) offers the ability to carry out telephone interviews in late 2018 to inform the design of baseline surveys administered in Jan 2019.

- As discussed in the set-up meeting if possible these will take place with 2-3 teachers who took part in the earlier pilot. Telephone interviews are recommended to minimise the impact on teachers' schedule, they can be flexibly rescheduled which is more difficult with face-to-face meetings. In our experience of working with teachers, we often find they require such flexibility particularly with relatively short interviews as they may be required to deal with emergencies. This information will be used to develop the teacher surveys and the case study visit research tools (interview schedules and observation note structures).
- We will also consult with the developer to discuss the final ToC outputs to establish the teaching and learning practices they expect to see as a result of the intervention, aim to inform instrument development. We will schedule monthly catch-ups with the developer to discuss the progression of the pilot. These meetings will be captured as field notes. If telephone meetings are difficult to arrange email communication may be used. The meetings will be structured around key questions dependent on the phase of the delivery.

School case study visits

School case study visits will be carried out mid-way through the project. These will involve one day for each of the eight visits. During this time the evaluators will carry out the following activities (for each of the eight visits):

- One * Year 10 Science lesson non-participant observations. This will be semi-structured and follow a drafted checklist to guide the observation of activities whilst also allowing for unanticipated avenues to be documented. The purpose of this method is to examine how the Deeper Learning intervention is delivered in practice and collect data independently of participant perceptions. It will be important to see the intervention in action to facilitate the interpretation of the survey and interview data. This data will be compared to the ToC model that will outline the original plans for change and allow us to note what does and doesn't happen.² To try and minimise the impact of the observation on teacher behaviour, teachers will be reassured that it is the intervention, not them, that is being examined and that any data collected will not be shared with their school and during any reporting they and their school will not be identified. We will request classes of varying pupil ability across the eight schools, so we can observe equal quotas of high, medium and lower ability classes.
- One * 1hr focus group with Year 10 Science teachers. This will follow a semi-structured schedule and will ask teachers to give reflections on CPD training and putting it into practice. Focus groups are preferred over interviews as the interactions between participants encourage the production of more fully articulated accounts. The focus groups will be recorded and transcribed³. A single focus group with the Year 10 science teachers would be our preference but we recognise that this may not be possible due to scheduling restrictions and teaching commitments. We may carry out shorter individual or paired teacher interviews if this is more feasible. Ultimately the evaluators will fit around teacher commitments on the day of the case study visit.
- One *1hr interview with the Deeper Thinking leadership person. This will follow a semi-structured interview schedule that will include reflections on how the teacher is managing the

² University of Sheffield, 'Observations', n.d., <https://www.sheffield.ac.uk/lets/strategy/resources/evaluate/general/methods-collection/observation>.

³ S Wilkinson, 'Focus Group Methodology: A Review', *International Journal of Social Research Methodology* 1, no. 3 (1998): 181–203, <https://www.tandfonline.com/doi/abs/10.1080/13645579.1998.10846874>.

intervention at a school level, facilitating and hindering factors and their thoughts on the CPD training. This basic scaffolding is filled out with appropriate prompts that are listed in advance⁴. The Deeper Thinking lead will be interviewed separately from the teacher focus group as they may need to reflect on the level engagement across the teachers, which would not be appropriate in a focus group scenario where they may not feel comfortable discussing this topic with the participants (fellow teachers, line managers, etc.). The interviews will be recorded and transcribed.

- In addition to the 8 case study visits the evaluators will also attend one of each of the four CPD training sessions offered by the developer. During the set-up meeting Carmel discussed collecting immediate feedback forms that the evaluators can analyse these if this is appropriate. The evaluators will ask teachers to reflect on the CPD training in the case study interviews/focus groups and the June survey on the basis that the teachers will have had chance to implement content from the CPD sessions and on that basis comment on the extent to which it met their needs when putting the intervention into practice.

Each of these activities will provide rich examples that can complement the survey data not repeat it. They will be crucial to establish if the implementation is happening as intended. The surveys are also required to be as short as possible to minimise the burden on teachers, the focus groups and interviews allow for additional and developed information to be collected that is not possible in a short survey.

Capturing ongoing teacher reflections on implementing the Deeper Thinking intervention

The evaluation team initially planned to consult the current intervention teachers at the beginning of the study and ask them to keep a reflective diary over the pilot. During the set-up meeting the developers described how reflection forms part of the delivery model meaning this data is already collected. In the interest of keeping the intervention as light touch as possible, we will not proceed with this activity. Instead we will ask the developers to share some of the reflection data that they are already collecting. We want to avoid duplicating data collection in the survey and case study visits, so we will study the existing templates as part of the data collection.

Teacher survey

A small longitudinal Census Survey will also be carried out. It is a census in the sense that we want to collect data from all Year 10 Science teachers involved in the study. To do this, we need the developer to provide the survey sampling frame of Year 10 Science teachers. This will include the following three fields: School name; teacher name and teacher email address.

The main purpose of the survey is to gather pre and post intervention measures which can be considered alongside the qualitative case study data. We will design a concise survey asking teachers to document changes in practice as a result of the intervention. The exact measures included will be determined largely by the ToC workshop. At the set-up meeting, it was proposed that not all schools participating in the pilot would receive a case study visit so a secondary purpose is to collect data from all the schools taking part.

The survey will be completed by the teachers on two occasions January and June and will contain several questions that feature in both and can be examined for change over time. The survey offers teachers the opportunity to disclose information they may not feel comfortable disclosing in a focus group or face-to-face encounter.⁵

The survey will ask teachers to rate a list of activities defined in the pre-intervention phase so we can assess change over time, it will also include some free-text questions to allow teachers to provide

⁴ W Olsen, 'Data Collection: Key Debates and Methods in Social Research', by pages 33-38 (London: SAGE Publications Ltd, 2018), <https://doi.org/10.4135/9781473914230>.

⁵ University of Sheffield, 'Questionnaires', 2018, <https://www.sheffield.ac.uk/lets/strategy/resources/evaluate/general/methods-collection/questionnaire>.

detailed comments. These free-text options will ask for specific examples of a change they have made or a difference they have observed in their students.

In terms of triangulation the initial interviews with the developer and previous pilot participants will inform the January survey design and the some of the June survey, which needs to include some repeat questions. The case study data collection may inform the June survey as it will offer the opportunity to test whether any interesting areas that arose during the visits apply to the wider sample of teachers taking part in the pilot. Likewise, the case study visits represent the opportunity to bring to life some of the examples that teachers may report in the survey allowing for a richer picture to be built up. This will be important given the relative subtlety of some of the indicators.

GCSE science as an outcome measure: capacity of the Deep Thinking approach to improve GCSE grades

An important part of the pilot will be to establish the most appropriate measure for a trial. Therefore, in this pilot, we propose to carry out in-depth simulations to estimate the likelihood that the SOLO /Carmel approach can improve GCSE grades.

We understand EEF's preference is for all initiatives to be demonstrated to show a significant improvement in national tests or public examinations – GCSE grades in this context (rather than marks on a GCSE paper, or scoring on some bespoke assessment). We will test that assumption with EEF at an early stage in the project.

Therefore, we will test if improving deeper thinking (as defined by the SOLO / Carmel initiative) has a substantial effect on GCSE grades. Given the Theory of Change we will establish how many marks on a typical GCSE paper could be improved by the piloted approach. The EEF project summary suggests that deeper thinking could directly improve scoring in respect of practical tasks and extended answer questions (EAQs).

In new science GCSEs, mandatory practical tasks are 'endorsement only'. That is, candidates have to do them, but they do not contribute to grades. However, around 15% of marks in each GCSE paper test insights that candidates have gleaned from practical tasks.

We will investigate whether improving performance on EAQs leads to substantial increases in grades achieved.

We do so in the light of the following:

- New GCSEs are graded 9 – 1, where they were previously graded A* to G. Therefore, one extra grade is available on a paper, and some or all of the grades may be less wide.
- On a Foundation Tier AQA GCSE paper, we found only three out of 53 question parts had a tariff of four marks or more.
- Not all GCSE science candidates could benefit from enhanced thinking on four-mark EAQs. Those who score 4/4, can't improve any more. Those who score 0 or 1 / 4, probably need to learn some basic facts, before they start connecting them. Therefore, probably deeper thinking on EAQs will only benefit candidates who currently score 2 or 3 on a four-mark question.

To address these issues, we will work with the Carmel development team.

- We will scrutinise some representative GCSE science papers (e.g. AQA Biology paper 1, upper tier; OCR combined science paper 2, upper tier, etc.).
- We will choose at least one GCSE science awarding organisation (AO), and one specification that AO offers (single or double award, for instance). We will then go through all the papers for this spec (for both the Foundation and Higher tier routes), and reach agreement about how many marks on the specification are susceptible to being improved by the deeper thinking approach.

- Then, we will simulate a mark distribution for the GCSE specification (and each route within it)⁶, using information from Joint Council for Qualifications (JCQ) results and grade boundary information.
- Then, we will evaluate scenarios for numbers of marks that could be gained given the SOLO approach; for example, we might assume that (best case scenario) SOLO helps candidates to gain 75% of the marks that are affected by SOLO, medium case scenario, 50%, worst case scenario 25%.
- Then we will say 'if trial candidates on average gained 75% of the marks that can be improved by this approach, what proportion of them would be likely to increase their GCSE grade?', and so on.

By the end of this analysis, we should be able to judge whether a whole GCSE will provide a reasonable outcome measure that could show the potential of Deeper Thinking/ deeper thinking to improve GCSE grades.

⁶ Such detail will be confidential to awarding organisations, and they are unlikely to share such sensitive information. However, grade boundaries on the spec as a whole are in the public domain, as is the percentage of candidates in each grade. Based on such information, and some reasonable assumptions, we should be able to simulate credible mark distributions for the studied papers.

Table 1: 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?	Theory of change workshop part 1. (Oct 2018)	Detailed logic model diagram and accompanying documentation.
	Follow up theory of change consultation post data collection to reflect on results. (July 2019)	Smaller model detailing the intervention on its own. Document to be shared with developers.
Is the intervention theory of sufficient quality (Plausible, Doable, Testable and Meaningful)?		Revisit documentation with developers at the end of the project.
Are there any unintended consequences of the intervention?		
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 Deeper Thinking intervention already. (Nov 2018)	Survey instruments developed with input from previous teachers and developers.
	Teacher survey data (Jan and June 2019)	Survey measures of teacher reports of change in practice.
	Case study observations to corroborate teacher self-report data (April 2019)	Teacher reflections on delivery and impact of the intervention in focus group/interview.
	The face-to-face data collection provides the opportunity to get more in-depth data than that afforded by the survey.	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 to be refined in the ToC workshop but may include the following examples: spreading reviewing over time, increasing peer and self-assessment.	Indicators of interest
	Teacher survey data (June 2019)	Teacher responses to survey measures of perceived changes in pupils.
	Case study visit focus group and interview data (April-May 2019)	Teacher insight into perceived impact on pupils developed in more detail in the case study visit.

Research question	Data collection method	Output
Is the intervention sufficiently innovative relative to the counter-factual? (Given the absence of a control group teachers will be asked to reflect on differences compared to the previous year of teaching.)	Targeted review of similar interventions	Judgment of points of similarity and difference.
	Teacher survey (June 2019)	Teacher reflections relative to the previous year.

Table 2: Summary of the ‘feasibility’ research questions and corresponding methods of data collection.

Research question	Data collection method	Output
Is the implementation happening as intended?	Teacher survey data (June)	Identification of factors that facilitate and hinder teacher delivery.
Can facilitating and hindering factors be identified?	Case study visit focus group and interview data (April-May 2019)	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 (June)	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 in light of their experience.
How does the implementation vary by the 3 areas practical's, analysis and extended answers?	Teacher focus group and interview Teacher survey data (June)	Ratings and insight gathered against the 3 sub-areas.
What is the ideal way of delivering the Deeper Thinking intervention?	Case study interviews with the Lead Practitioner	Insight from individual teacher and school perspective.
What do ‘ideal’ practitioners and schools look like?	Teacher focus groups	

Table 3: Summary of the ‘scalability/ readiness for trial’ research questions and corresponding methods of data collection.

Research question	Data collection method	Output
<p>How suitable is the Deeper Thinking intervention to progress to efficacy trial?</p> <p>Is it clearly defined and scaleable?</p> <p>Will the intervention be practicable and attractive for schools beyond the immediate pilot group?</p>	<p>Given all of the above data a judgement will be made by considering the evidence in relation to the success measures set out in the logic model.</p> <p>These will be reflected on during the second Theory of Change workshop.</p> <p>Targeted review of similar interventions to inform how innovative the programme is.</p>	<p>Final recommendations and report</p>
<p>What might be the main design options for the efficacy trial?</p>	<p>Analyses of statistical power; suggestions as to sample size necessary for trial.</p>	<p>Information to inform the overall judgment of whether to progress to efficacy trial.</p>
<p>Will GCSE Science be a suitably precise and targeted measure to pick up a measurable impact on student attainment?</p>	<p>Exam paper analysis and grade simulations.</p> <p>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. (Dec-Feb 2019)</p>	<p>Detailed discussion of what the outcome measure for an efficacy trial might look like.</p>
<p>Are there secondary outcome measures that can be addressed by an efficacy trial?</p>	<p>Theory of Change workshop</p> <p>Case study visits could also provide insight into unanticipated secondary measures.</p>	<p>Additional measures that the efficacy evaluators may want to incorporate.</p>

LOGIC MODEL UPDATE

The evaluation team have held a face-to-face workshop with the Carmel Education Trust to draw out key assumptions held about the Deeper Thinking product and how it should be delivered.

The evaluation team and delivery team will revisit the logic model at the end of the data collection phase to discuss whether it needs updating.

WHO WILL COLLECT THE DATA?

The survey data will be collected and analysed by the evaluation team and will not be shared with the delivery team beyond the final anonymised report. The delivery team will provide the teacher email addresses to the evaluation team, this will be a complete list of all Year 10 which removes any possibility of bias in teacher selection.

The teacher reflection data which will be shared by the delivery team to avoid duplicating tasks for teachers. The evaluation team will ask the delivery team to provide a selection representing the range of teacher responses. This will be triangulated with the teacher survey which mitigates potential bias.

The case study visits will also be carried out by the evaluation team, the delivery team will not be present to ensure the teachers feel able to talk reflect openly on the intervention.

Data analysis

Data collection method	Data analysis approach
Theory of change (ToC) workshop	<p>Synthesise the paper-based outputs into a logic model.</p> <p>Negotiate the finer details post-meeting</p> <p>Revisit the logic model towards the end of the project</p>
Telephone interviews with teachers and the developer	<p>Audio record and write notes.</p> <p>Use the notes to inform the development of the survey questionnaire and case study research tools</p>
Attend CPD training sessions offered by the developer	Store observation template in NVivo
Notes recorded in an observation template	This output will be used to inform the case study visit templates
School case study visits	The 8 checklists will be analysed in NVivo, the analysis will look for pertinent examples, and similarly and differences
Non-participant observations, observation notes collected on an observation checklist template	
School case study visits	Data to be transcribed
1hr focus group with Year 10 Science teachers	Coded in NVivo to identify themes within and between groups
School case study visits	Data to be transcribed
1hr interview with the Deeper Thinking leadership person	Coded in NVivo to identify themes within and between groups
Capturing ongoing teacher reflections on implementing the Deeper Thinking intervention	Coded in NVivo to identify relevant examples and counter examples
Teacher survey *2	This data will be analysed in SPSS recognising the limits posed by a relatively small number of respondents
The survey will ask teachers to rate a list of activities defined in the pre-intervention phase so we can assess change over time, it will also include some free-text questions to allow teachers to provide detailed comments. These free-text options will ask for specific examples of a change they have made or a difference they have observed in their students.	The qualitative responses will be coded in NVivo

GCSE science as an outcome measure:
capacity of the Deep Thinking approach to
improve GCSE grades

Scrutinise some representative GCSE science
papers

Simulate a mark distribution for the GCSE
specification

Evaluate scenarios for numbers of marks that
could be gained given the SOLO approach

Ethics and registration

This evaluation is subject to the AlphaPlus Ethics Policy. The British Educational Research Association (2018) guidelines are used as a framework for ethical considerations in our research work. The guidelines suggest that all education research is carried out 'within an ethic of respect' for: the person, knowledge, democratic values, the quality of academic research, and academic freedom. A research ethics checklist is required to be completed by the project manager, which refers to sections of the BERA code of practice and is to be used as a guide and aide memoire for project team members with responsibilities for ethical research.

This informs: the need for anonymity of data held; data policy for holding and using data; the researcher role and the interpretations of any findings; any power dynamics and sensitive data held as a result of the study, which could harm an individual or groups of individuals.

We abide by the following principles:

- voluntary participation
- proactivity in avoiding risk of harm
- confidentiality by default
- anonymity by default
- research will be legal

Our research is overseen by our independent Quality Assurance and Ethics Board.

There are three stages of ethical clearance, this project is currently at Stage 2 but we do not expect it to progress to Stage 3 as we are not collecting data from the pupils and the data we are collecting from teachers is not deemed to be sensitive.

- **Stage 1** involves the project manager completing an Ethics Review Checklist. This is the first stage of three. It will help the project manager (and others) decide to what extent the project manager needs to become involved in the second and third stages. Once completed the project manager (and the QA and Ethics Board) will be in a position to make this judgement.
- **Stage 2** will involve the project manager discussing any ethical dimensions to the research in some depth with another 'knowledgeable person of standing'; this is a very likely outcome of completing the checklist.
- **Stage 3** will involve the project manager obtaining formal 'ethical clearance' through the Company's procedures; some projects will need to proceed to this stage.

AlphaPlus employees must all hold a DBS standard check, which should be registered with the DBS service for annual updates to maintain currency.

Data protection

SURVEY DATA

The survey distribution list of teacher emails will be stored in a password protected file in a restricted access server folder. It will only be available to the evaluation team.

The survey will be hosted in Qualtrics, a professional survey package. When the data is extracted teacher email address will be replaced with a numeric UID.

The survey analysis will be transparent, the analysis will be carried out by Kathy Seymour and the SPSS files will be available to Hayley Limmer who will carry out appropriate quality assurance checks. A selection of data checks will be made against the original data to quality assure the results.

Teachers will be assured that their data will not be shared with their school.

No individual teachers or schools will be named in any of the published reports. Hayley Limmer is an ONS Accredited Researcher and has been trained to ensure she does not present data in a disclosive manner.

QUALITATIVE DATA

Audio recordings will be transferred from recording devices to the secure SharePoint folder at the first opportunity.

The transcription documents will be edited to remove any teacher or schools' names and replace them with pseudonyms. This will take place prior to entering the files into NVivo for coding and analysis.

Analysis will be carried out in NVivo which will provide a clear rationale for quality assurance checks of the analysis.

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.

Alpha Plus 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?

School name and exam board for GCSE science will be collected at the school recruitment stage. This will also include publicly available information.

Personal teacher information that will be collected includes:

- Teacher names
- Teacher email addresses and phone numbers
- School name

We will only use the data to contact teachers in advance of evaluation activities and to send online survey requests. 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.

The evaluation data provided by teachers will cover their experiences of using the Deeper Thinking Intervention. This data will be collected through a combination of focus group discussions, interviews, classroom observations and surveys.

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 (within 28 days of completion of the project) 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

GDPR COMPLIANCE

AlphaPlus complies with GDPR for all its work. Our data protection policy has been thoroughly updated in the light of the new regulations. At project initiation, we share our privacy policy with all relevant stakeholders. In our work with EEF, we are experienced at negotiating data sharing agreements.

Throughout projects, data are held in secure locations within the EU with access limited on grounds of need. Project managers are obliged to delete data once they are no longer necessary for projects.

Where research participants' identities were known at some point (e.g. in completing questionnaires), subsequent anonymisation protects personal data as per article 9.

Regarding article six, lawful processing will be 'necessary for the performance of a task that is of 'legitimate interest'.

Although we of course note that qualitative data collected from teachers, and opinion data from surveys, etc. constitute sensitive data under GDPR, our pilot design does not use NPD data; which we suggest is proportionate and reduces risks of data breaches.

Personnel

Personnel	Team	Role
Andrew Boyle	AlphaPlus	Is director of research and will direct this project.
Hayley Limmer	AlphaPlus	Will be the day-to-day lead on all activities in this project. She will co-ordinate with colleagues, and lead on the conception and implementation of process evaluation.
Ben Smith	AlphaPlus	Is a senior statistician, he will carry out the GCSE simulations
Clare Dowland	AlphaPlus Associate	Is a long standing senior associate. Clare will lead project start-up and legal activities (e.g. protocol and MoU development, EEF and sub-contractor agreements, etc.).
Kathy Seymour	AlphaPlus Associate	Kathy is a senior associate, she will host, distribute and analyse the teacher survey.
David Bailey	Carmel Education Trust	Director of Research and Development
Dorothy Warren	Carmel Education Trust	Science Consultant
Sarah McGee	Carmel Education Trust	R&D project officer
Joe Collin	EEF	EEF Carmel relationship manager
Camilla Nevill	EEF	EEF Evaluation Manager

Risks

Risk	Mitigation
<p>Some teacher CPD sessions are delayed resulting in some schools using the Deeper Thinking Intervention more than others</p>	<p>The evaluation team have no control over the implementation of the intervention but nevertheless any delays in implementation will have implications for the quality of data that the evaluation team are required to analyse.</p> <p>The evaluation team will account for variation between schools during the qualitative write up of results.</p>
<p>Teachers do not implement the Deeper Thinking intervention straight away following the CPD session</p>	<p>As above.</p>
<p>Schools struggle to arrange a focus group with Year 10 Science teachers on the case study visit day.</p>	<p>We will be as flexible as possible in terms of the date we visit to accommodate the needs of the school.</p> <p>We may carry out 1-1 interviews or paired interviews with teachers instead (within the constraints of the 1 day visit).</p>
<p>Teachers are reluctant to participate in the case study observations</p>	<p>Provide clear communication materials that highlight that the intervention is being evaluated and not the teacher. Any delivery issues relate to the intervention not a shortcoming in teaching.</p>
<p>The length of the pilot trial is too short to build up an evidence base appropriate to evaluate the Deeper Thinking Intervention.</p>	<p>The research questions include scope for indicators of promise.</p>

Timeline

Dates mm/yy	Activity	Staff responsible/leading
08/18-11/18	PHASE 1: PROJECT START-UP	
Nov 18	Agree contract, research protocol, etc.	Clare Dowland & Hayley Limmer
Oct 18	Theory of change workshop	Hayley Limmer
Oct 18-Nov 18	Logic model developed and agreed	Hayley Limmer
July 19	Revisit logic model in view of data collected	Hayley Limmer
Nov 18	Ethical approval	Hayley Limmer & Andrew Boyle
Nov 18	MoU signed off	Clare Dowland
Nov 18	Contract signed off	Clare Dowland
Aug 18- Dec18	Carmel Education Trust to confirm the schools who are taking part	Carmel Education Trust
Dec 18	Carmel Education Trust to transfer list of Year 10 science teacher emails to AlphaPlus for evaluation activities	Carmel Education Trust
	PHASE 2: PROCESS EVALUATION ACTIVITIES	
Nov 18	Carmel to provide contact details of 3 teachers who will be willing to take part in a telephone interview to discuss the previous pilot. Nov 2018.	Carmel Education Trust to set up contact Hayley Limmer to carry out interview
Dec 18	Design Teacher Survey and circulate for sign off	Hayley Limmer
Jan 19	Carmel Education Trust to provide details of 4 training sessions to AlphaPlus to attend one of each of the 4 different CPD sessions.	Carmel Education Trust
Jan 19	Collect Teacher Survey (baseline Jan 2019)	Hayley Limmer & Kathy Seymour
June 19	Collect Teacher Survey (June 2019)	Hayley Limmer & Kathy Seymour
July 19	Analyse survey data	Hayley Limmer & Kathy Seymour
Feb 19	School observation visits instrument development	Hayley Limmer
April- May 19	School observation visits (April-May 2019)	Hayley Limmer
May 19	Carmel Education Trust to share teacher reflection data with AlphaPlus	Carmel Education Trust
May-June 19	Analysis of case study data	Hayley Limmer
	PHASE 3: HOW CAN SOLO IMPROVE GRADES	
Dec 18- Jan 19	Scrutinise GCSEs, simulate grade and marks Analyses of statistical power; suggestions as to sample size necessary for trial	Ben Smith
	PHASE 4: JUDGEMENT AS TO WHETHER SOLO IS SUITABLE TO GO TO TRIAL	

Dates mm/yy	Activity	Staff responsible/ leading
June-July 19	Analysis to come to a view regarding suitability to go to trial	Hayley Limmer & Andrew Boyle
July-Aug 19	Evaluation of coherence and credibility of theory of change	Hayley Limmer & Andrew Boyle Ben Smith
	PHASE 5: REPORTING	Hayley Limmer
Aug 19	Report writing	Hayley Limmer
Aug 19	Presentation of findings and recommendations regarding trial to EEF	Andrew Boyle