Gender Balance in Computing

Evaluation of Teaching Approach: Peer Instruction intervention

June 2022

Julia Ryle-Hodges, Stefan Kelly and Clément Bisserbe

THE BEHAVIOURAL INSIGHTS TEAM

Contents

Contents 1

Executive summary 2

1. Background 7
- 1.1 Gender Balance in Computing 7
- 1.2 Teaching Approach programme 7
- 1.3 GBIC partners 11

2. Methods 13
- 2.1 Impact evaluation 13
- 2.2 Implementation and process evaluation 21

3. Impact evaluation findings 26
- 3.1 Primary analysis: effect of the intervention on girls' intention to study computer science at GCSE 26
- 3.2 Secondary analysis: effect of the intervention on girls' attitudes towards computing 27

4. Implementation and process evaluation findings 30
- 4.1 Implementation 30
- 4.2 Intervention 39

5. Conclusions and recommendations 53
- 5.1 Summary and interpretation of findings 53
- 5.2 Recommendations 54

References 57

Appendices 59

Executive summary

Overview of the project

Although there have been increasing numbers of girls and women pursuing a career in computing, there remains a distinct gap between genders, which can be traced back to secondary school subject choices[1]. In 2020, only 21% of students taking GCSE Computer Science were female[2]. The Gender Balance in Computing (GBIC) programme has been structured around known and well-researched barriers to girls engaging with computing, including a felt lack of belonging, perceived lack of relevance, difficulty linking informal learning about computing (for example Code Clubs) to formal GCSE Computer Science, and a mismatch of teaching approaches to pupil learning preferences, all of which may hinder engagement and may negatively impact girls' perceptions of computing.

The GBIC programme has been funded by the Department for Education (DfE), with the Raspberry Pi Foundation (RPF) serving as the primary delivery organisation and the Behavioural Insights Team (BIT) acting as independent evaluators. This report details the evaluation of an intervention in the Teaching Approach strand of the programme, where the aim was to improve girls' attitudes towards computing by using a collaborative teaching approach. Specifically, the Peer Instruction approach is a 12-week, year 8 intervention which covers two half terms of computing lessons. Within this approach, pupils are set a 'pre-instruction' task to complete before the start of each lesson (known as flipped learning) and go through a structured process of answering multiple choice questions (MCQs) within the lesson, responding first individually to each question before discussing their response with a group. This group discussion of MCQs aims to create a learning environment that fosters collaboration.

Evaluation approach

The intervention was evaluated using a mixed methods approach. The impact evaluation investigated whether there was evidence that the intervention affected i) girls' intention to study computer science at GCSE level and ii) girls' attitudes towards computing. In parallel, a mixed-methods implementation and process evaluation (IPE) was conducted to explain the quantitative findings and explore implementation processes and possible mechanisms of change in targeted outcomes.

[1] BCS: The Chartered Institute for IT. (2021) "Computing is the fastest growing STEM A level, says professional body for IT." [Blog] Available at: https://www.bcs.org/articles-opinion-and-research/computing-is-the-fastest-growing-stem-a-level-says-professional-body-for-it/

[2] Joint Council for Qualifications (2021) "GCSE (Full Course) Results Summer 2021 - Outcomes for key grades for UK, England, Northern Ireland & Wales, including UK age breakdowns". Available at: https://www.jcq.org.uk/wp-content/uploads/2021/08/GCSE-Full-Course-Results-Summer-2021.pdf

Impact evaluation

The impact evaluation design was a two-armed cluster randomised controlled trial (RCT), randomised at the school level with outcomes at the pupil level. The two arms of the trial were:

1. Control: Schools in the control arm taught their usual computing lessons for a 12-week period from October 2021 to February 2022.
2. Treatment: Schools in this group received training materials and class plans to deliver the 12-week period of Peer Instruction computing lessons (also from October 2021 to February 2022).

The primary outcome was girls' intention to study computer science at GCSE, measured using a question in a survey filled out by pupils. The secondary outcome was girls' attitudes towards computing and was measured in the same survey using the Student Computer Science Attitude Survey (SCSAS), a tool for assessing attitudes toward computing for school pupils. In total, 90 secondary schools were originally recruited to participate in the trial and 46 submitted endline survey data.

Implementation and process evaluation

The IPE aimed to answer the following research questions:

1. Which components of the interventions were delivered and in what ways?
2. How did teachers experience delivery of the intervention?
3. In what ways did pupils engage with the intervention?
4. In what ways did the intervention affect girls and what were the mechanisms through which the intervention was perceived to bring about change in the anticipated outcome?

BIT researchers spoke with teachers at five schools where the intervention was delivered to understand teachers' experiences of delivering the intervention, the feasibility of delivery, any programme adaptations teachers made while delivering, and the perceived effects for girls. One teacher from a control school was also interviewed, to understand what 'business as usual' looked like for a year 8 class. In three of the schools where the Peer Instruction lessons were delivered, we conducted additional research activities (small group discussions with pupils and lesson observations) to better understand how pupils experienced and engaged with the intervention and how the intervention was being implemented in practice. We also explored the responses to a teacher survey which collected feedback about the online training.

Key findings

Evidence of impact

Overall we found no evidence that the intervention increased girls' intention to study computer science at GCSE or improved attitudes towards computing. Although the attrition observed limited the evaluation's ability to detect an effect on the intervention, given the very small size of the differences between the two groups, we do not suspect that attrition rates are obscuring a substantive positive impact of the intervention.

Implementation and process evaluation

The findings below emerged from the analysis of the data collected for the five case study schools:

- **Fidelity:** Generally, teachers used the lesson resources (e.g. lesson plans, power-point presentations, worksheets) with little adaptation. However, in their implementation of the Peer Instruction elements of the intervention (the structured, collaborative MCQs and the pre-instruction tasks) there was considerable variation in teacher delivery, with some teachers de-prioritising the collaboration aspect of the MCQs. Shifting the training from in-person to online due to the COVID-19 pandemic may have contributed to this variation as online training might have constrained the exploration of the collaborative nature of the approach. While this was not mentioned by interviewed teachers, it is also possible that some other teachers might have deprioritised group work over the time period of the intervention (October 2021 - February 2022) due to high cases of the Omicron variant of COVID-19.

- **Feasibility:** Teachers reported that the MCQs were straightforward to deliver and that the content of lesson plans was clear. There were some concerns around the amount of time teachers spent preparing for lessons and the challenge of completing all of the lesson activities within the time allocated for the lesson.

- **Quality:** The interviewed teachers felt that the quality of the lesson plans and other resources was high and provided them with everything they needed to teach their lessons. They were particularly enthusiastic about the MCQs: some had even started developing their own banks of questions to be used in other topics.

- **Responsiveness:** Pupils and teachers engaged well with the lessons. Teachers reported higher pupil engagement with Peer Instruction lessons compared to ‘business-as-usual’ lessons and teachers said that they had enjoyed using the resources and being encouraged to try a new approach to teaching.

- **Mechanisms:** There was little evidence that the mechanisms that had originally been hypothesised (increased collaboration leading to girls’ perception that they know as much as their peers; and increased collaboration leading to increased subject knowledge) were at work within the pupils receiving the intervention. However, teachers and pupils suggested two additional mechanisms through which they felt the intervention might contribute to the intended outcomes: greater collaboration increasing the opportunities for pupil help which created a secure learning environment; and increased collaboration leading to more experiences of success. It is possible that these two additional mechanisms interacted with teachers’ fidelity to use of group work: in schools where MCQs were used as intended (as a tool to prompt discussion) there would be more scope for the Peer Instruction intervention to work through the two additional mechanisms identified.

The challenges above, particularly those related to fidelity and feasibility, may have limited the impact of the intervention. Two additional possible explanations for the lack of a measurable effect are:

1. The hypothesised barrier that the intervention was designed to address may not be the most critical barrier to girls choosing GCSE Computer Science. The hypothesised barrier is that the teaching approach used within Computing lessons is currently misaligned with girls' preferences for collaborative work. If this barrier is not the most critical barrier to girls choosing GCSE Computer Science, tackling this barrier is unlikely to dramatically affect girls' choice.
2. The large and structural barriers which prevent girls from choosing GCSE Computer Science, which are reflected in the lower baseline proportion of girls reporting the intention to study GCSE Computer Science (7% of girls vs. 20% of boys) may not be possible to overcome by implementing a single, short-term strategy.

Recommendations

The following steps could help to increase the feasibility of implementing the Peer Instruction programme:

- Simplify lesson resources such that lessons take less long to prepare
- Emphasise the importance of the stages of the Peer Instruction MCQs throughout the intervention materials

As the pupils and teachers responded positively to the Peer Instruction units, we recommend that RPF:

- Share general guidance with teachers on how to apply the Peer Instruction approach to computing lessons: this guidance could be applied to any units of computing - not only the lessons included in this intervention
- Build banks of MCQs for other computing units, or develop guidance for teachers to write questions themselves
- Make available online the Peer Instruction lesson resources and training that were used in this intervention

To increase the potential impact of interventions targeting gender gaps in computing, one possible avenue would be to:

- Consider implementing multiple strategies in parallel to address the barriers to girls' engagement with computing

Finally, possible strategies to address the evaluation challenges encountered could be to:

- Continue to refine survey tools and support schools to administer them to maximise data reliability and reduce attrition
- Measure the outcomes targeted by the intervention further into the future to track actual (rather than intended) GCSE subject choice

In light of the disruptions to the delivery of the intervention associated with the COVID-19 context and the positive experiences of the case study schools, there is reason to believe that implementing the intervention again after addressing the adjustments to its design and delivery suggested in the recommendations above could result in improved effectiveness. In addition, using school administrative data to measure whether girl pupils in the evaluation sample go on to select computer science as a GCSE subject would help to both reduce the need for primary data collection and increase the precision of the results in capturing any impact on the target behavioural outcomes. We thus recommend exploring the possibility of conducting another round of this intervention and evaluation if these suggested adaptations can be made, particularly if the cost of this new round of activities would be low.

1. Background

1.1 Gender Balance in Computing

Computing has a decades-old problem with gender imbalance with limited reliable evidence of what works in closing the gap[3]. Across England, only 21% of the GCSE Computer Science cohort is made up of female students[4]: many girls are not choosing to continue with computing subjects at the point at which lessons become optional, usually at the start of year 9 or 10[5] [6]. The Gender Balance in Computing Project (GBIC) aims to tackle a number of known and well-researched barriers to girls engaging with computing[7], including a disconnect between extra-curricular computing activities and subject choice; a lack of encouragement to studying computing; a lack of familial and other role models in computing; a perceived lack of relevance of computing to students; and a mismatch of teaching approaches to pupil learning preferences. These barriers are addressed in the five intervention strands which comprise GBIC, with the common goal of increasing the number of girls who study GCSE and A Level Computer Science. This report covers one of the Teaching Approach interventions, which uses structured collaborative activities in computing lessons to attempt to address the barrier of the mismatch between teaching approaches and girls' preferred approaches to learning, and therefore improve girls' attitudes towards computing and increase intention to choose GCSE Computer Science.

1.2 Teaching Approach programme

The premise of the 'Teaching Approach' interventions is that the way in which computing is taught may not always match the teaching approaches that girls are most likely to respond positively to[8] [9]. In particular, the way in which computing is taught may lead to girls perceiving computing as an individualistic rather than a collaborative subject. The Peer Instruction intervention aims to change girls' perceptions of computing by using teaching approaches

[3] Royal Society. (2017) After the reboot: computing education in UK schools. https://royalsociety.org/~/media/policy/projects/computing-education/computing-education-report.pdf

[4] Joint Council for Qualifications (2021) "GCSE (Full Course) Results Summer 2021 - Outcomes for key grades for UK, England, Northern Ireland & Wales, including UK age breakdowns". Available at: https://www.jcq.org.uk/wp-content/uploads/2021/08/GCSE-Full-Course-Results-Summer-2021.pdf

[5] Kemp, P.E.J., Berry, M.G. & Wong, B. (2018). The Roehampton Annual Computing Education Report: Data from 2017. *London: University of Roehampton.* Available at: https://cdn.bcs.org/bcs-org-media/3972/tracer-2017.pdf

[6] Pupils start their GCSE chosen subjects in either year 9 (age 13-14) or year 10 (age 14-15) depending on whether the school allocates two or three years to the GCSE curriculum..

[7] Childs, K (2021) Factors that impact gender balance in computing. In Understanding computing education (Vol 1). Proceedings of the Raspberry Pi Foundation Research Seminar series

[8] Goode, J., Estrella, R., & Margolis, J. (2006). Lost in translation: gender and high school computer science. Women and Information Technology: Research on Underrepresentation, eds JM Cohoon & W. Aspray.

[9] Goode, J. (2007). If you build teachers, will students come? The role of teachers in broadening computer science learning for urban youth. Journal of Educational Computing Research, 36(1), 65-88.

that encourage collaboration and discussion within computing lessons. This approach is grounded in evidence that incorporating collaborative teaching approaches into STEM subjects improves girls' self-efficacy and achievement[10] [11].

Within the Teaching Approach programme, two interventions were implemented and evaluated: Pair Programming for Primary age pupils (year 4 and year 6) and Peer Instruction for Secondary age pupils (year 8). This report covers only the Peer Instruction intervention; the evaluation of the Pair Programming intervention is covered in a separate report.

Peer Instruction

The intervention, developed and rolled out by RPF, is a 12-week programme that incorporates the Peer Instruction structured teaching approach into two, half-term units of work for year 8 pupils: 1) Binary representations[12] and 2) Text based programming in Python [13]. Peer Instruction can replace a traditional presentation approach by combining pre-instruction, multiple choice questions (MCQs), and peer discussion, to encourage deeper engagement with the content in question. The content taught within the Peer Instruction lessons is in line with the National Curriculum[14]; the distinctive element of the intervention is the teaching approach that is taken to deliver the content. This Peer Instruction approach has been evaluated within the context of teaching physics in higher education and found to lead to significant learning gains[15]; within computing, trials have found the approach is valued by pupils and that pupils who have experienced Peer Instruction lessons would recommend the approach to others[16].

Peer Instruction follows a number of stages:

1. Pre instruction: the teacher shares the pre-instruction task for pupils to complete in advance of the next lesson. This task introduces a concept which will be covered in the next lesson and often contains a worked example and then questions for pupils to consider.
2. During the lesson, the teacher will present 3-4 MCQs. For each MCQ:
    a. The teacher presents the MCQ to the class - this will have carefully chosen 'distractor' answers to allow the teacher to identify pupils' misconceptions
    b. Pupils think about the question independently and give a solo response (e.g. through sharing their answer on a whiteboard or using a digital voting system)

[10] Werner, L., & Denning, J. (2009). Pair programming in middle school: What does it look like? Journal of Research on Technology in Education, 42(1), 29–49. https://doi.org/10.1080/15391523.2009.10782540
[11] Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. American Journal of Physics, 74(2),118–122. https://doi.org/10.1119/1.2162549
[12] This unit introduces pupils to binary digits and how they can be used to represent text and numbers.
[13] This unit introduces pupils to text-based programming with Python. The lessons start with elementary programs involving input and output, and gradually move on through arithmetic operations, randomness, selection and interaction.
[14] Department for Education (2014) *The national curriculum in England: complete framework for key stages 1 to 4*.
[15] Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, *69*(9), 970-977.
[16] Porter, L., Bouvier, D., Cutts, Q., Grissom, S., Lee, C., McCartney, R., ... & Simon, B. (2016, February). A multi-institutional study of peer instruction in introductory computing. *In Proceedings of the 47th ACM Technical Symposium on Computing Science Education (pp. 358-363).*

   c. The teacher places pupils into informal groups, in which to discuss their answers
   d. Pupils engage in peer discussion in this group and reach a group consensus which they share with the class
   e. The teacher leads a class discussion to address any misconceptions and link to the pre-instruction materials
3. Within the lesson, there are three or four MCQs for the class to work through. The first MCQ is generally linked to the pre-reading, but the subsequent questions might be on slightly different topics.

Figures 1 and 2 illustrate an example of a pre-instruction task and a MCQ. The pre-instruction tasks were generally two - three pages in length and included worked examples to follow through before presenting a task to be completed. Pupils were not expected to 'hand in' their pre-instruction tasks to be marked: the motivation for completion was designed to come from the link between the pre-instruction task and the MCQ within the lesson. The MCQs were always presented with four different options: the correct answer and three different 'distractor' options based on common misconceptions.

Example 1: Simple assignment

```
days = 365
print(days, "days in a year")
```

Programming terms:

- Assignment (or assign) – where a variable is given a value
- Value – in the above example, the value is '365'
- Expression – refers to the value of a variable

Variable rules:

- Evaluate an expression – to evaluate this expression (days in a year), the days variable must have been assigned a value
- Program execution – when a program is executed (run), a variable must have been assigned a value **before** that value is referred to
- Assignments are not equations, they are instructions to be carried out and they can change

Task

1. Read the Python program below and circle the following terms in the code example below. Circle the value and the expression of the variable 'days in week':

```
daysinweek = 7 * 1
print(days in week, "days in a week")
```

You should have circled '7 * 1' two times as it is the answer to value and the expression!

2. Can a variable be executed without being assigned a value?

3. Statement: Assignments are instructions to be carried out.
   Is this statement true or false? YES or NO (circle the correct answer)

Figure 1: Extract from a pre-instruction task

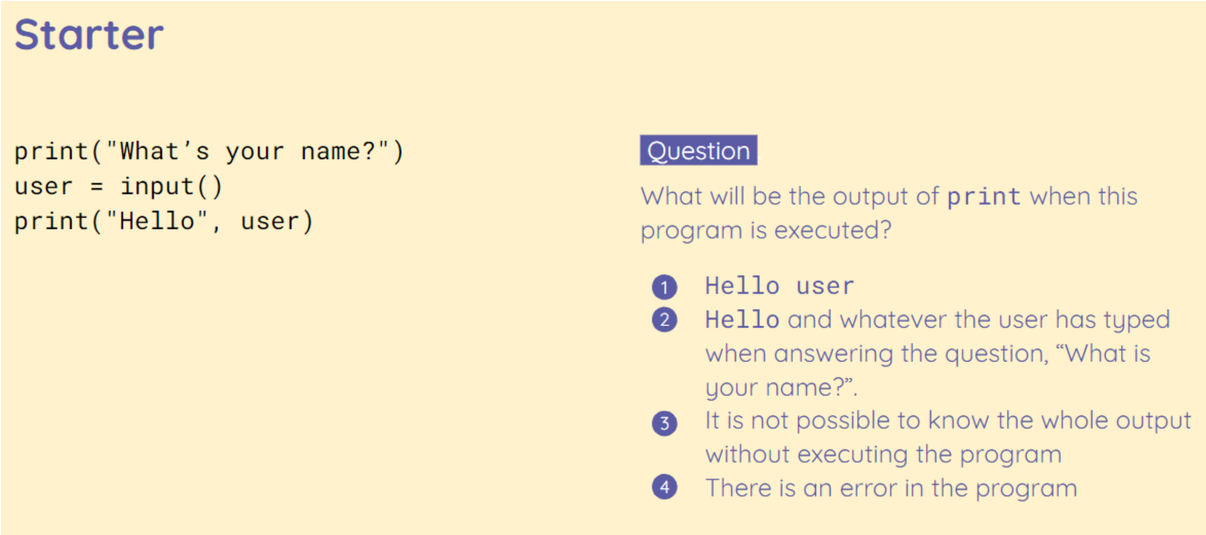

Figure 2: Example MCQ within a teacher power-point

Previously created National Centre for Computing Education (NCCE) lesson plans were used as the basis of the Peer Instruction lessons. These lesson plans had some material removed from them to create time for the MCQs, and the pre-instruction tasks and MCQs themselves were added. These 'base' lessons varied in structure, but the continuing thread through all of the Peer Instruction lessons was a) having a linked pre-instruction task and b) including three or four MCQs which were spread throughout the lesson. For example, a lesson structure might look like:

1. A MCQ that is linked to the pre-instruction task (first MCQ)
2. Teacher live modelling of a task
3. A second MCQ
4. Pupils working on the activities independently, using the activity worksheet for guidance. Pupils moving on to the next activity when they have completed the previous one.
5. A third MCQ
6. Teacher introducing the pre-instruction task for the next lesson

Whilst teachers and pupils might be familiar with MCQs from 'normal' teaching, the Peer Instruction approach to MCQs is different in that it is prescribing a structured and collaborative approach to answering the questions.

The logic model (see Figure 3) was developed through discussion between the RPF team and the BIT evaluators. It illustrates the proposed hypothesised mechanisms through which the Peer Instruction intervention would affect the intended outcomes of girls' intention to study GCSE Computer Science and attitudes towards studying computing. The key barrier that the intervention was designed to address was a lack of perceived collaboration in computer science lessons leading to a tension between computer science lessons and girls' preferred ways of learning. By making computer science lessons more collaborative, the intervention aimed to address this barrier. The Peer Instruction lessons were taught in the second half term of the Autumn term in 2021 to the end of the first half of the Spring term in 2022 (October 2021-February 2022).

Teachers undertook an online training course which they worked through at their own pace.[17] They were also given the option of watching a recorded webinar. During the online training, teachers were directed to online portals containing all of the resources they needed to deliver the units and were given time to explore them. These resources included: a unit overview; individual lesson plans; a power-point presentation for each lesson; a power-point containing all of the MCQs for each unit; pre-instruction tasks; and worksheets for each lesson.

The original plan had been for teachers to attend face-to-face training in which they would have seen the structured process of responding to MCQs modelled so that the nature of collaborative work could be fully explored. Due to COVID-19 constraints, the only option was online and asynchronous delivery which did not allow for the same exploration of the collaborative nature of the approach. This may have diluted the impact of the training.

In advance of the start of this evaluation, RPF conducted a pilot of the Peer Instruction intervention in January-March 2020. Teachers received face-to-face training and gave feedback on both the training and their experience of delivering the Peer Instruction lessons. The intervention was adapted based on their feedback before the start of this evaluation. Whilst the end of the pilot was disrupted due to COVID-19, RPF were able to receive and act on meaningful feedback from the teachers involved.

Interestingly, after the evaluation had ended, Ofsted released guidance on the teaching of Computing[18] which recommended that teachers include formative assessment, multiple choice questions and worked examples within their teaching - all elements which are prevalent within the Peer Instruction approach.

1.3 GBIC partners

This project joins the National Centre for Computing Education, run by a consortium comprised of STEM Learning, the British Computer Society (BCS), and the Raspberry Pi Foundation (RPF) with the Behavioural Insights Team (BIT), combining the extensive experience of organisations who have computing at the core of their mission with expertise in designing and evaluating interventions. The funding body for this programme as a whole is the Department for Education (DfE), and BIT fulfils the role of an independent and external evaluator.

[17] Raspberry Pi Foundation: Peer Instruction online training course.
[18] Ofsted (2022) *Research review series: computing*. Available at https://www.gov.uk/government/publications/research-review-series-computing/research-review-series-computing

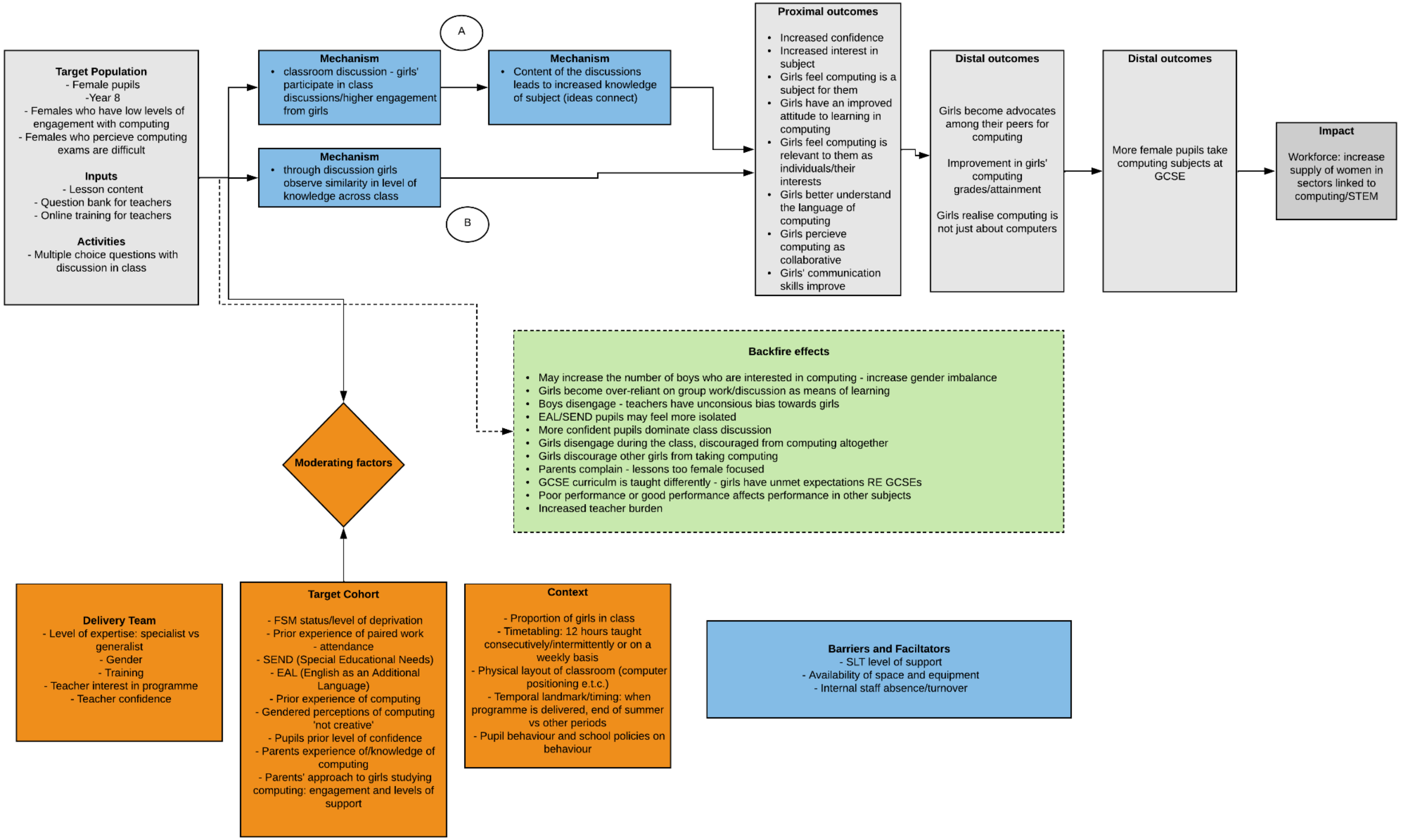

Figure 3: Logic model of the Peer Instruction intervention

2. Methods

The evaluation used a mixed-methods approach. The impact evaluation was designed as a randomised controlled trial, with two arms (one control, one treatment), and was randomised at the school level with outcomes at the pupil level. Quantitative data was collected via computer-based surveys distributed pre and post intervention in both treatment and control schools to be completed as part of computing lessons. An implementation and process evaluation (IPE), which aimed to explore the mechanisms of change and complement the quantitative survey findings, was also carried out. This section addresses the research questions, methods used and the limitations of our approach.

2.1 Impact evaluation

2.1.1 Research questions and outcome measures

The impact evaluation aimed to determine whether the intervention led to a change in:

1. Girls' stated intention to study computer science at GCSE
2. Girls' attitudes towards computing as measured by the Student Computer Science Attitude Survey (SCSAS)

These outcomes were measured through the indicators described in Table 1.

Table 1: Method for collecting quantitative outcome data

Outcome measures	Data to be collected	Point of collection
Primary: Intention to select Computer Science at GCSE	Single item survey measure of whether the pupil plans to continue studying computer science with possible responses "Yes", "No", or "I don't know".	Online surveys, completed on computers in class at baseline (beginning October 2021) & immediately following the culmination of the 12-week programme (beginning February 2022). These surveys were 'built in' to the first and last lesson plans of the 12 week intervention [19].
Secondary: General attitudes towards computing	Overall score on the Student Computer Science Attitudes Survey (All 5 constructs equally weighted: Confidence, Interest, Belonging, Usefulness, Encouragement).	

[19] Not all schools completed the intervention in the 12-week period allotted, as some schools completed the endline survey more than 12 weeks after completing the baseline survey.

2.1.2 Sampling and randomisation

The impact evaluation was designed as a two-armed cluster randomised controlled trial, and was randomised at the school level with outcomes at the pupil level. The two arms of the trial were:

1. Control: Schools in the control arm taught their usual computing lessons for a 12-week period beginning October 2021.
2. Treatment: Schools in this group received training materials and class plans for a 12-week period of computing lessons to deliver the Peer Instruction intervention as outlined in section 1.2.

Randomisation was conducted by BIT, using school unique reference numbers (URNs) as unique identifiers. As not all schools that had registered interest in the trial had confirmed their participation by the point at which schools were due to be randomised, schools were stratified on their confirmation status. Following randomisation, balance checks on other school-level variables were carried out. Groups were found to be balanced in terms of Ofsted ratings (categorised as ‘Outstanding’, ‘Good’ or ‘Inadequate / Requires improvement’), the proportion of schools that were girls only and the percentage of pupils eligible for free school meals.

Pupils were blind to allocation during the programme and during outcome data collection. Teachers were not blind to allocation, as they were responsible for delivering the materials, and, as the schools had registered interest in participating in the trial, the teachers were aware that there was both a control and a treatment group, and of which group their school was in.

Recruitment of schools was conducted by RPF. All secondary schools in England that could offer the full 12-week programme were eligible for the trial. Schools were also required to have female pupils (all-boys schools were excluded).

All schools that entered the sample did so voluntarily, which has implications for the external validity of the findings, as schools that volunteer are likely to be more enthusiastic than the average school, and may thus not be representative of all schools in England.

Data was collected for both boys and girls, but only data from girls was analysed for primary and secondary analyses.

2.1.3 Description of data

Table 2 presents the mean scores and standard deviation (SD) for each SCSAS subscale at baseline, split by gender. The SCSAS is composed of five distinct subscales that represent different facets of studying computing as a subject and the perceived benefits of doing so, as well as the experience of computing lessons. These subscales are confidence, interest, belonging, usefulness and encouragement: the individual questions that make up these subscales can be found in Appendix 1

Table 2: Pupil baseline survey data by gender (outcome indicators emboldened)

Outcome	Values	Gender	N (non-missing)	Mean (SD)
Intention to study computer science at GCSE	1 = "Yes" 0 = "No", "Don't know"	Girls	3,210	0.07 (0.26)
		Boys	3,137	0.20 (0.40)
Total SCSAS score	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,186	2.48 (0.45)
		Boys	3,112	2.72 (0.50)
SCSAS: Confidence subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,186	2.55 (0.53)
		Boys	3,112	2.78 (0.57)
SCSAS: Interest subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,185	2.36 (0.61)
		Boys	3,112	2.70 (0.68)
SCSAS: Belonging subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,140	2.67 (0.51)
		Boys	3,075	2.87 (0.54)
SCSAS: Usefulness subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,140	2.61 (0.56)
		Boys	3,075	2.82 (0.60)
SCSAS: Encouragement subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,139	2.26 (0.61)
		Boys	3,075	2.46 (0.66)

Table 2 shows that at baseline, boys scored higher than girls in intention to study computer science in the future (20% of boys indicating they would like to study computer science as a subject for their GCSEs compared to 7% of girls; $p<0.001$). While the correlation between this measure and actual subject choices is unknown, it is promising that the baseline rate of our analytical sample (7%) is close to an estimate provided by the DfE of 6% of girls selecting computer science.

Table 2 also shows that at baseline, boys scored higher than girls in attitudes toward computing (boys' mean = 2.72 out of 4 compared to girls' mean = 2.48 out of 4; $p<0.001$), and that we observe gender differences in the same direction across all five SCSAS subscales.

It is worth noting that the baseline scores of intention to study computing have a much larger gender gap in this sample of year 8s compared to the sample of year 4s and year 6s who took part in the other trial within the Teaching Approach project: Pair Programming. In the younger cohort of pupils, 55% of girls said that they would like to continue to study computing, compared to 66% of boys. In the older, year 8 sample, only 7% of girls said that they intended to study computer science at GCSE, compared to 20% of boys. Whilst a direct comparison cannot be made, as younger pupils were asked ‘Do you want to study computing in the future?’ and older pupils were asked ‘Do you intend to study GCSE Computer Science?’, it is striking that the gap between genders widens to such an extent over this 2-4 year time period, and that the overall proportion of pupils intending to study GCSE Computer Science (when asked in year 8) is so much lower than the proportion of year 4 or 6 pupils intending to continue to study computing.

It’s possible that the higher intention to study computer science reported in the younger sample is because these pupils have not yet been in a position in which they could choose which subjects they studied: it’s possible that they didn’t fully understand the context of being able to choose which subjects they had lessons in. Despite this, it’s plausible that this increasing gender gap does, to some extent, represent a diverging of attitudes towards computing for boys and girls between key stage 2 and key stage 3.

2.1.4 Attrition and final sample

Figure 4 describes school-level attrition at the different stages between recruitment and the completion of the endline survey in each trial arm. At both baseline and endline points of pupil survey data collection, RPF attempted to minimise attrition (across both treatment and control groups) by extending the window for data collection to account for schools that were delayed in completing surveys, and by sending reminder emails to school that had not completed the surveys by the expected time.

Despite these efforts, attrition was observed between randomisation and completion of baseline surveys, with proportionally more schools dropping out from the treatment group than the intervention group (39% vs 22%). This differential attrition brings risk of bias at the analysis stage, and this risk is explored later in this report through balance checks on the baseline data received from these schools post-attrition.

Attrition was also observed between baseline and endline, in terms of both schools failing to complete the endline survey and pupils within schools not completing the endline survey. In contrast to the attrition between randomisation and completion of baseline surveys this stage saw proportionally more schools from the control group dropping out (34% compared to 16% in the treatment group). One potential reason for schools dropping out at this stage is the disruptions to schools caused by the rise of the Omicron variant of COVID-19 in the UK in winter 2021, which is likely to have caused pupil absences, teacher absences and disruptions to regular lessons, although this would not explain the differential attrition rates of the treatment and control groups.

Overall, this results in very similar attrition rates across experimental groups when viewed from point of randomisation to endline survey completion.

The final sample used for analysis also included one school from the treatment group who completed the endline survey, but did not complete the baseline survey. We chose to keep this school in the sample, and addressed the missing data in line with the analytical strategy.

Once all survey data was collected, data cleaning was conducted to remove any data points deemed potentially unreliable. All data was dropped for pupils who had answered in a straight pattern (e.g., a survey with the answer ‘Strongly disagree’ for every question of the SCSAS). This applied to 23 pupils from the total baseline sample of 7,219 (boys and girls, not necessarily matched to any endline observations) and 76 pupils from the endline sample of 4,114 (boys and girls). In cases where there were duplicate observations (the same pupil entering the survey twice), we kept only the first complete survey from the pupil. If a pupil never fully completed the survey, we retained their first partially complete entry. The final analytical sample consists of 1,959 girls - 918 in the treatment and 1,041 in the control group.

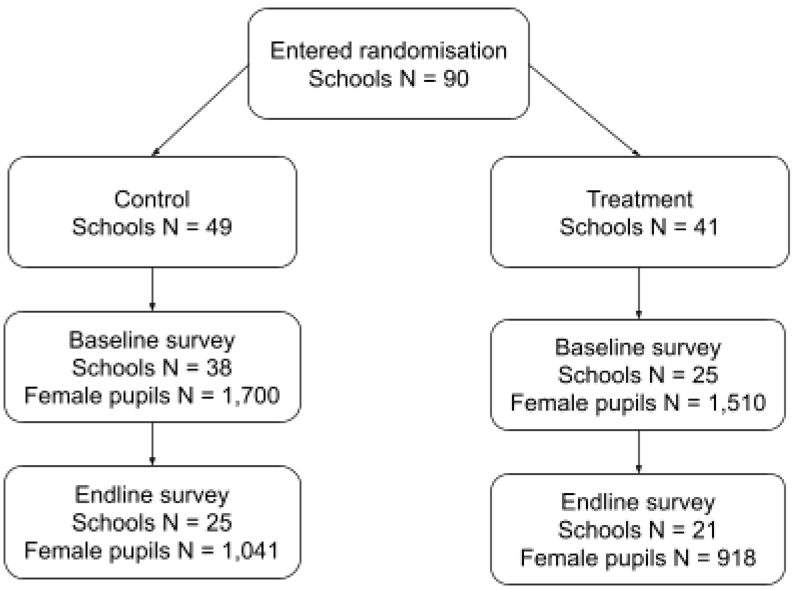

Figure 4: School level attrition

Baseline differences in outcome measures

Table 3 shows that at the point of baseline data collection, the groups were balanced in terms of attitudes towards computing and intention to study computer science in the future when including boys and girls. When restricting the sample to girls only, the control group had a marginally higher proportion of girls that intended to study computer science (8.3% vs 6.4%). There were proportionally more girls in the control group (52%, vs 44% in the treatment group).

Table 3: Balance checks for all baseline data

Covariate	Percentage (or mean) per arm		p-value	Balanced?
	Control (n = 3,281)	Intervention (n = 3,435)		
Gender				
Boys	42.8%	50.5%	<0.01	No
Girls	51.8%	44.0%		
Non-binary/Other	5.4%	5.6%		
Intention to select computer science as a GCSE subject				
Baseline intention (full sample)	13.6%	13.7%	>0.10	Yes
Baseline intention (girls only)	8.3%	6.4%	<0.05	No
SCSAS				
Baseline SCSAS score (full sample)	2.59	2.61	>0.10	Yes
Baseline SCSAS score (girls only)	2.49	2.48	>0.10	Yes

Baseline differences in outcome measures for final analytical sample

Further attrition was observed between baseline and endline data collection. Table 4 also shows that at the point of baseline data collection, the final groups (the composition of which is outlined in section 2.5.1) were balanced in terms of attitudes towards computing and intention to study computer science in the future. The difference in proportion of each group that were girls remained, but the difference in girls' intention to study computer science was no longer statistically significant, which seems to be a function of the reduction in sample size rather than a change in the underlying difference.

Table 4: Balance checks for baseline data of pupils who completed the endline survey

Covariate	Percentage (or mean) per arm		p-value	Balanced?
	Control (n = 1,041)	Intervention (n = 918)		
Intention to select computer science as a GCSE subject				
Baseline intention (girls only)	8.8%	6.8%	>0.10	Yes
SCSAS				
Baseline SCSAS score (girls only)	2.53	2.52	>0.10	Yes

Implications for final analysis

The differences in baseline intention to select computer science as a GCSE subject among outcome levels described above carry implications for our analysis. This outcome is roughly 2 percentage points higher in the Control group compared to the Intervention group among girls who completed the endline survey. The difference is not statistically significant but this may be due to the loss in power resulting from the high attrition: the difference between groups for all girls who answered the baseline is roughly 2 percentage points and this is statistically significant at the 5% level. The lower average mean intention to study GCSE Computer Science in the intervention group at baseline suggests a risk that our treatment estimates may be biassed downwards.

This has implications for interpretation: even though we can control for baseline scores in the outcome measures at the pupil level, there may be unobserved school-level variables that are confounding the treatment effect. If we expect the bias resulting from attrition in the control to be downwards in terms of our outcome measures and that no treatment effect is observed, this reduces the risk that the attrition is 'hiding' a true positive effect of the intervention.

2.1.5 Analytical approach

The full model is presented in Appendix 2. The primary and secondary analyses were both Intention to Treat (ITT) estimates. This means that outcomes were analysed on the basis of the groups that schools and pupils were randomly allocated to, regardless of their compliance with the intervention. The covariates (baseline SCSAS score, school Ofsted rating, school proportion of pupils with free school meal eligibility) were chosen as they could potentially influence the outcomes, thus controlling for these variables could increase the precision of estimates.

All planned covariates were checked for missing data pre-analysis. For some schools in the sample, we were unable to obtain an Ofsted rating due to there not being one publicly available. For these schools, we elected to assign them to an extra value of the categorical variable of Ofsted rating[20].

Given that the endline data would likely include some pupils who did not complete the baseline survey, we specified pre-trial decision rules for dealing with missing data as baseline scores on the SCSAS were to be used as a covariate in the analysis. In the final sample, approximately 21% of pupils were missing baseline SCSAS scores (above the threshold of 5% for listwise deletion), and multiple imputation[21] was performed, whereby predicted values were substituted where data was missing.

In order to fully examine the effect of multiple imputation on our estimate of the intervention's impact, we also present the results of the primary and secondary analysis whereby (i) missingness was instead addressed through missingness indicator and (ii) only complete cases (pupils who completed both baseline and endline surveys) were used. For both the primary and secondary analysis, these specifications are presented in order of:

1. Multiple imputation model

[20] While it would have been possible to perform multiple imputation on missing Ofsted data, this was judged to be inadvisable as not all independent schools are inspected by Ofsted, with schools in our sample likely falling into this category. This would suggest that this data was not missing at random. Thus, using this as an extra category within the Ofsted rating covariate would be more informative than using other school-level variables to predict Ofsted rating.

[21] Rubin (2004) *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

2. Missingness indicator model[22]
3. Complete case analysis

The majority of the pupils in the endline data who could not be matched to any baseline data were from schools that did complete the survey at both time points, meaning that these pupils may have been absent or out of class when baseline survey data was collected.

2.1.6 Limitations

Attrition

An implication of generally high attrition is that the analysis will not be powered to detect a change in outcome measures of the targeted effect size specified pre-trial. 90 schools were recruited with an approximately 30% attrition buffer to detect an estimated minimum detectable effect size (MDES) of a 6 percentage point increase in intention to select computer science at GCSE level. This meant approximately 33 schools per arm were required to reach this target MDES, compared to 21 treatment schools and 25 control schools in the final sample. Overall, this means the final analytical sample was not sufficient to detect an effect at the originally targeted effect size.

In addition, differential attrition across treatment groups may introduce a bias in the results, as discussed in section 2.1.4.

Pupil survey outcome measures

Given that the pupils in this trial were in year 8 for the duration of the intervention, most are over a year away from making their GCSE subject choices. Because of this, we rely on self-reported intention to study GCSE Computer Science as our primary outcome measure, rather than actual GCSE subject choices. While the correlation between this measure and actual subject choices is unknown, it is promising that the baseline rate of our analytical sample (7.8%) is close to an estimate provided by the DfE of 6% of girls selecting computer science. It is possible that the sample rate may also be marginally higher than the national rate given that these pupils are from schools that volunteered to take part in this intervention.

[22] In running this model, we included a binary covariate, coded as 1 if the baseline survey had been completed, and 0 if the baseline survey was incomplete. This allowed us to include all complete endline observations without using multiple imputation.

2.2 Implementation and process evaluation

Alongside the impact evaluation, an implementation and process evaluation (IPE) was conducted. The IPE examined the mechanisms of change and the diversity of implementation and programme delivery.

2.2.1 Research questions

The IPE aimed to address the following research questions:

1. Which components of the interventions were delivered and in what ways?
2. How do teachers experience delivery of the intervention?
3. In what ways do pupils engage with the intervention?
4. In what ways does the intervention affect girls and what are the mechanisms through which the intervention is perceived to bring about change in the anticipated outcome?

2.2.2 Research design

A qualitative design was used as the basis for the IPE as this allowed us to explore the context of each school and the experiences of the teachers and pupils within them. We planned and implemented a case study design, conducting qualitative research with teachers and pupils from the same school, where possible. To add context to the feedback on the online training, we also report on the quantitative findings from the RPF online training feedback survey.

We envisaged conducting multiple research activities in four case study schools, however, possibly due to the challenging COVID-19 context schools were facing, we were only able to recruit three treatment schools in which we could complete our multiple research activities. As we had less detailed case study data from the schools in which we conducted only a teacher interview (and no pupil interviews or observations), we recruited two treatment schools to be ‘teacher interview’ only schools, to try to ensure a breadth of experience was represented. We also spoke to one teacher from a school in the control condition, to hear what ‘business as usual’ looked like for a year 8 class.

2.2.3 Sampling and recruitment

Across the participating schools, case study schools were selected to represent range and diversity, both in terms of the school characteristics as well as the teacher and pupil experience of the intervention. Recruitment criteria were split into primary (characteristics which require representation as we had predefined) and secondary (characteristics which are relevant to the experience of the intervention but which could be used more flexibly in terms of representation).

Case study school sampling

Primary sampling criteria for schools included i) region and ii) proportion of pupils eligible for free school meals (FSM). The secondary criterion was the school’s Ofsted rating. Schools’ proportion of FSM eligible pupils was retrieved from the DfE national information about schools[23]. We recruited two schools from the South of England, one school in the Midlands and one school in the North of England. We did not achieve the diversity we had aimed for in the proportion of FSM eligible pupils in case study schools: three of the schools had slightly lower than national average FSM and two had low proportions of pupils eligible for FSM. The low rate of responses to our initial invitations for schools to participate in the evaluation meant that we were unable to ensure the diversity of FSM percentage. Only approximately one fifth of the schools in our full sample had above the national average percentage of pupils eligible for FSM: this might have contributed to the challenge of recruiting a high FSM school to take part as a case study school.

Staff sampling

For teachers within the case study schools, the primary criterion was gender and the secondary criterion was teaching experience. All teachers sampled were computer science teachers. We recruited seven teachers: three from one case study school and one from each of the remaining four case study schools. Of these seven teachers, four were male and three were female. Of the four teachers who shared their years of experience, there was an even balance between those who had been teaching for more than 10 years and those who had been teaching for fewer than 10 years.

Pupil sampling

We aimed to recruit a sample of two pairs of year 8 pupils from three of the five case study schools. We were aiming to talk to one pair of girls and one pair of boys at each of the three schools. As pupils had to miss either lesson time or lunchtime to complete the interviews, we were only able to recruit one pair/group for each of the three case study schools in which we did multiple research activities. In one school, instead of arranging two paired interviews (one with two boys and one with two girls) we conducted a group interview with two boys and two girls. In another school, we conducted a paired interview with two girls, and in the third school, we conducted a group interview with three girls.

For pupils, primary criteria included gender and the pupil's engagement with computing. The sampling of pupils occurred following a discussion with the teacher who was able to identify pupils with different levels of computing engagement within their class.

The impact of COVID-19 on teacher workload and visiting policies meant that we were not in a position to be rigid with our sampling. Contacted teachers explained that, during the period the programme was running, teachers and schools were dealing with very high levels of staff and pupil absence. This made it challenging for teachers to facilitate additional research activities (such as lesson observations) as part of the evaluation of the intervention. We therefore relaxed our criteria in order to balance the need for range and diversity with the realities of conducting qualitative research in schools during a global pandemic. Table 5 below details the achieved sample of case study schools.

[23] https://www.gov.uk/school-performance-tables

Table 5: Achieved case study sample

School	Profile	Teacher	Pupils	Data collection
S01	- Located in the South East - Slightly below average FSM - Academy - Ofsted rating: Good	- computer science teacher - Fewer than 10 years of experience - Male	- Total n = 4; year 8 pupils (12-13 years old) - 2 female (less engaged in computing) - 2 male (more engaged in computing)	- Teacher interview - Pupil group interview - All activities were conducted virtually
S02	- Located in the South East - Low FSM - LA maintained - Ofsted rating: Good	- Head of computer science department - Fewer than 10 years of experience - Male	- Total n = 3; all female, year 8 pupils, range of engagement in computing	- Teacher interview - Lesson observation - Pupil group interview - All activities were conducted virtually
S03	- Located in the West Midlands - Slightly below average FSM - LA maintained - Ofsted rating: Good	- Head of computer science - More than 10 years of experience - Female	- Total n = 2; both female, year 8 pupils, more engaged in computing	- Teacher interview - Lesson observation - Pupil paired interview - All activities were conducted virtually
S04	- Located in the South East - Low FSM - LA maintained - Ofsted rating: Outstanding	- Head of computer science department - More than 10 years of experience - Female	None	- Teacher interview - Conducted virtually
S05	- Located in the North West - Slightly below average FSM - LA Maintained Ofsted rating: Good	F1 - Head of computer science Department - Female M1 - computer science teacher - Male M2 - computer science teacher - Male	None	- Teacher interview - Conducted virtually

School recruitment

Given the strong relationships RPF had built with participating schools, we asked RPF to reach out to the schools we had identified as meeting our sampling criteria, to ask whether they would like to be involved in the evaluation of the Peer Instruction project. Once teachers had indicated that they would like to participate, BIT staff set up a call to discuss the practicalities of a school visit or whether the research should be conducted online, and scheduled a date for participation.

Online feedback survey teacher sample

All school staff that intended to be involved in teaching the programme lessons (both teachers and teaching assistants) were invited to take part in the online training. All those who completed the online training were asked to complete a short feedback survey as the final step of their training. Completing the feedback survey was required in order for schools to receive the reimbursement of £100 per staff member completing training. This reimbursement was to cover any supply costs, such as paying a cover teacher to deliver a teacher's lesson whilst the computer science teacher completed the Peer Instruction training.

2.2.4 Data collection methods

Teacher interviews

In depth 30-45 minute semi-structured online interviews were conducted with teachers who delivered the Peer Instruction lessons in each of the case study schools. These interviews were designed to explore teachers' experiences of the intervention and any factors that influenced their ability to implement the intervention with their pupils. In one school, we conducted a group (as opposed to an individual) interview with three computer science teachers who had all been teaching the Peer Instruction lessons. In the other schools, the interviews were conducted with an individual teacher. We also had an informal discussion with a teacher from a school in the control group about the curriculum covered in year 8 during 'business as usual'.

Pupil paired/group interviews

Group interviews were held with pupil groups or pairs at three of the case study schools. These lasted around 20 minutes and were conducted through a video call. Pupils were asked to order different skills in terms of how important they are for computing, to discuss whether a series of statements about computing were 'true or false' (e.g., whether boys and girls are equally likely to have computing as their favourite subject) and to reflect on the elements of the Peer Instruction lessons they particularly enjoyed or found challenging. The questions were designed to create the opportunity for pupils to share their current perceptions of computing, as well as their experience of the Peer Instruction lessons, and how this compared to 'normal' computing lessons.

Lesson observations

These were designed to independently assess pupil engagement, lesson fidelity and facilitators and barriers to lesson delivery. The lesson observations were conducted online, with the BIT researcher joining a video call of the lesson. This made it more of a challenge to observe certain elements of the lesson, such as pupils' engagement or to listen in to small group discussions. As a result, the data that was gathered was possibly less rich than would have been gathered from an in-person observation. The lessons themselves were 'in-person' lessons: no pupils were participating virtually. In the two schools in which we conducted both pupil interviews and a lesson observation, we conducted the lesson observation first, so that we were able to refer back to the lesson itself within the pupil interview.

Training feedback survey

RPF invited teachers and teaching assistants who had completed the online training to fill in a short online feedback survey (see Appendix 3). Across the treatment schools, 50 school staff completed the survey. This asked closed-ended questions about school staffs' experience of the training: the amount of time it took, its usefulness and levels of school staff confidence in implementing the intervention following the training.

2.2.5 Analysis

Case study data

Interview transcripts and fieldnotes were managed using the Framework Approach[24]. This involved summarising transcripts and notes into a matrix organised by themes and sub-themes (columns) as well as by individual cases (rows). The managed data was then interpreted with the aim of identifying and categorising the range of phenomena present in each of the sampling groups. We conducted case and theme analysis to provide rich descriptions of participant experiences.

In interpreting the findings from the analysis, important considerations include:

1. The case study approach means that findings should not be generalised across all participants, but rather understood as conveying some of the range and diversity of participant experiences. It is also important to note that the case study schools which provided the data for the IPE do not reflect the breadth of levels of pupil disadvantage in England's schools: we did not hear from any schools which have higher than national average FSM percentage.
2. The schools which responded to our invitation to take part in the evaluation might have been those that had teachers who felt most confident in their teaching practice -therefore the findings may not reflect the breadth of experiences of teachers implementing the intervention.
3. The total number of teachers we spoke to was small compared to the number of teachers that implemented the intervention. As a result, it's possible that we did not capture the full range of teacher experience within the case study schools.

Training feedback survey data

Descriptive statistics were generated based on the quantitative data gathered from the online training feedback survey. For each question, we calculated the percentage of the total sample which chose each of the categorical survey responses.

[24] Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science students and researchers*. sage.

3. Impact evaluation findings

3.1 Primary analysis: effect of the intervention on girls' intention to study computer science at GCSE

Key findings:

- There was no evidence of an impact on girls' intention to study computer science in the future in the intervention group relative to the control group.
- There was no evidence of an impact on girls' attitudes towards computing as measured by SCSAS.

The results of the primary and secondary analysis are presented in Tables 6 and 7. Primary and secondary model specifications, along with full regression tables, can be found in Appendix 2.

There was no evidence that the intervention positively impacted girls' intention to study computer science in the future relative to the control group. The proportion of girls stating they intended to study computer science at GCSE for the full sample of girls was 9.1%. For the intervention group this proportion was 8.8% and for the control group it was 9.5%. Compared to the control group, the pre-specified multiple imputation model found a difference of 0.3 percentage points (p=0.899) for the intervention, which is not statistically significant at conventional significance levels. This finding was consistent across the missingness indicator and complete case analysis model specifications.

Table 6: Impact evaluation results for primary outcome

Outcome: Intention to study computing	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group mean	9.5%		9.7%
Treatment group mean	8.8%		8.9%
Estimated treatment effect (in percentage points)	0.3pp	0.0pp	-0.1pp
N	1,959	1,959	1,542

Figure 5 shows the raw control mean and estimated treatment effect of the intervention using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention. The 95% confidence interval of this treatment effect is also shown on the bar of the intervention group.

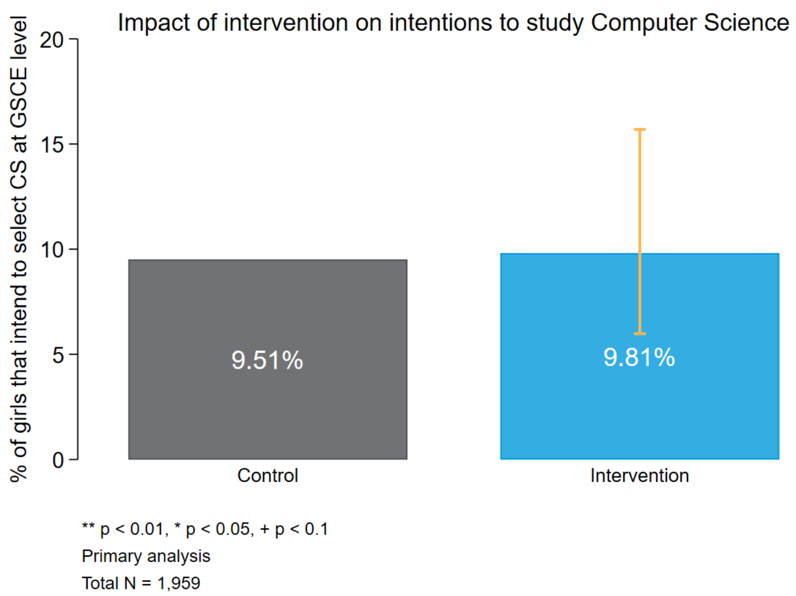

Figure 5: Model-adjusted intention scores by treatment group

3.2 Secondary analysis: effect of the intervention on girls' attitudes towards computing

There was no evidence that the intervention positively impacted girls' attitudes towards computing relative to the control group, as measured by scores on the SCSAS. The mean score on the SCSAS scale (range 1-4) for the full analytical sample was 2.42 (SD=0.528). For the intervention group it was 2.39 (SD=0.538) and for the control group it was 2.45 (SD=0.516). Compared to the control group, the pre-specified multiple imputation model found a difference of -0.04 points (p=0.391) on a 1-4 scale for the intervention, which is not statistically significant at conventional significance levels. This finding was also consistent across the missingness indicator and complete case analysis model specifications.

Unlike other trials within the GBIC programme, this intervention could not be clearly tied to any of the five subscales of the SCSAS (confidence, interest, belonging, usefulness and encouragement). Because of this, and concerns over performing a large number of significance tests[25], we did not conduct analysis relating to these subscales.

[25] Estimating the impact of the programme on multiple outcomes increases the risk of a false discovery (finding a significant result by chance). To address this risk, corrections are applied in the analysis (such as the Benjamini-Hochberg step-up procedure), which raises the threshold at which a finding is considered statistically significant for all tests conducted. Therefore, we only conducted tests related to the outcomes most closely tied to the intervention goals, to maximise our ability to detect statistically significant results for these outcomes.

Table 7: Impact evaluation results for secondary outcome

Outcome: Total SCSAS Score	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group unadjusted mean	2.45		2.47
Treatment group unadjusted mean	2.39		2.40
Estimated treatment effect (standard error)	-0.04 (0.050)	-0.05 (0.048)	-0.06 (0.043)
N	1,914	1,914	1,510

Figure 6 shows the raw control mean and estimated treatment effect of the intervention using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention. The 95% confidence interval of this treatment effect is also shown on the bar of the intervention group.

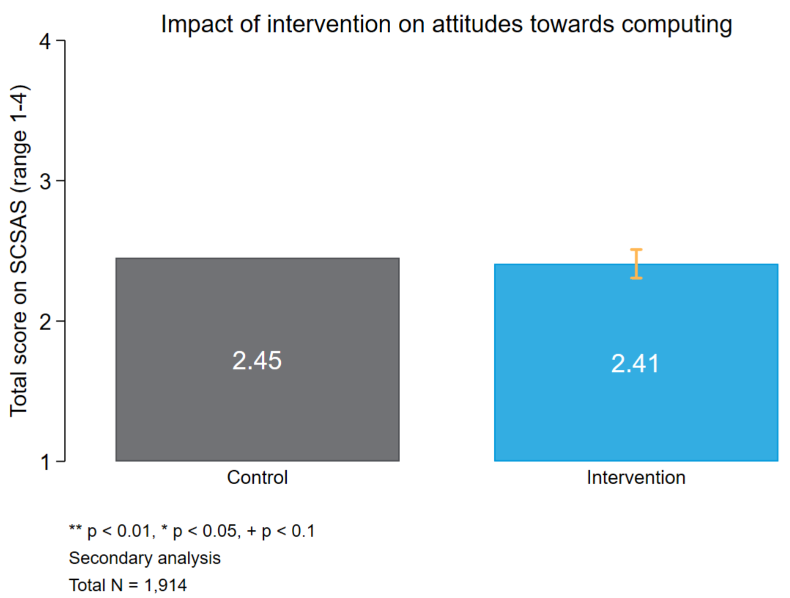

Figure 6: Model-adjusted SCSAS scores by treatment group

For a more detailed overview of endline survey responses, Table 8 describes the baseline and endline mean score of each SCSAS subscale by treatment group, for the girls who completed both the baseline and endline survey, and whose data was thus used in the complete case analysis model specification.

Table 8: Baseline and endline SCSAS subscale and overall scores by treatment group for girls who completed both surveys

Subscale	Survey	Group	N (non-missing)	Mean (SD)
Confidence	Baseline	Control	777	2.63 (0.52)
	Baseline	Treatment	765	2.62 (0.51)
	Endline	Control	1,009	2.55 (0.59)
	Endline	Treatment	905	2.47 (0.61)
Interest	Baseline	Control	776	2.42 (0.60)
	Baseline	Treatment	765	2.37 (0.59)
	Endline	Control	1,000	2.25 (0.66)
	Endline	Treatment	900	2.19 (0.71)
Belonging	Baseline	Control	760	2.70 (0.52)
	Baseline	Treatment	754	2.72 (0.45)
	Endline	Control	989	2.63 (0.54)
	Endline	Treatment	894	2.59 (0.55)
Usefulness	Baseline	Control	760	2.65 (0.54)
	Baseline	Treatment	754	2.64 (0.51)
	Endline	Control	972	2.54 (0.63)
	Endline	Treatment	889	2.47 (0.65)
Encouragement	Baseline	Control	760	2.31 (0.60)
	Baseline	Treatment	754	2.28 (0.59)
	Endline	Control	969	2.27 (0.66)
	Endline	Treatment	884	2.21 (0.68)
Overall SCSAS score	Baseline	Control	777	2.53 (0.45)
	Baseline	Treatment	765	2.52 (0.42)
	Endline	Control	1,009	2.45 (0.52)
	Endline	Treatment	905	2.39 (0.54)

4. Implementation and process evaluation findings

The IPE findings are split into two main sections: Implementation and Intervention. The Implementation section explores the extent to which the different aspects of the intervention were delivered as intended, focusing on 1) Fidelity and 2) Feasibility. The Intervention section explores 1) Quality of the intervention resources, 2) Responsiveness (or engagement) of pupils and teachers with the intervention elements and 3) Mechanisms through which the intervention could have affected the target outcomes.

4.1 Implementation

4.1.1 Fidelity

Key findings:

- Teachers reported that the COVID-19 context had made it a challenge to teach the 12-week unit within the designated time frame.
- Schools struggled to set the pre-instruction task as homework: some combined sometimes setting the task as homework with sometimes using it as a lesson starter, whilst others stopped setting it altogether.
- There was variation in the implementation of the MCQs: in some schools, there was little focus on the group discussion element of the questioning, which could mean the intended mechanism of Peer Instruction leading to increased collaboration was not facilitated. The need to shift from face-to-face training to online training due to the COVID-19 context, as well as health and safety measures (such as encouraging social distancing) may have contributed to this variation.
- Teachers used the power-points, lesson plans and worksheets as intended, with few adaptations.

This section describes how closely the implementation of the programme within the school context matched the intended implementation, as described in the training and materials from RPF.

Lessons taught

All of the interviewed teachers reported that they had struggled to finish teaching the units of lessons within the timeframe of the programme. Generally, this was due to the COVID-19 context: a lot of teachers had been off school isolating, and in some cases, whole year-groups were closed, which meant that pupils missed lessons. Where teachers got 'behind', they generally finished the lessons after the end of the intended implementation period[26].

Pre-instruction tasks

Teachers generally adapted how they set the pre-instruction tasks for the pupils, rather than setting them as home learning each week. These tasks were generally documents which included some text for pupils to read, and then questions to answer. None of the schools that we talked to fully implemented the pre-instruction tasks by setting them as homework before each of the Peer Instruction lessons[27]. Some teachers used a combination of setting the tasks as homework and using them as 'starters' at the beginning of lessons. Other teachers found that it was a challenge to set these tasks alongside the other demands of teaching the Peer Instruction lessons.

Teachers reported that it wasn't always possible to set the pre-instruction tasks before each lesson because of conflicts with their homework policies. These policies included only setting computing homework once a fortnight and having to leave a certain amount of time between setting homework and the next lesson. Teachers appreciated having the option of setting the pre-instruction task either to be done at home or at the start of the next lesson.

"According to our policy, they have to have so much notice. I can't give them homework on the Monday for the next day [when there are two computing lessons a week]. I've done a mixture of having it at the beginning of the lesson and a mixture of homework." (S03)

Where teachers adapted the pre-instruction task to use it as a starter, they reported that this worked well. Teachers reported that they were able to read the text out loud to any pupils who preferred to hear it and that it fitted with their school policies to start a lesson with a reading 'starter'.

"We were flipping that [pre-reading] *around and making it a starter, so we were sending it out to all the screens before they come in and straightaway they had it there, ready to go." (S04)*

"I looked at it and thought this is actually a really good starter activity so they did it in the first five minutes of the lesson." (S05)

Some teachers did not consistently set the pre-instruction tasks throughout the programme. These teachers felt they didn't have the time required to set the pre-instruction tasks. These tasks would generally be set on an online platform which meant teachers needed to upload the document in the correct format and allocate it to the relevant pupils.

"As it progressed I kind of forgot to do them [Teacher M2 agrees]. It was taking me such a long time to prepare a lesson, I couldn't manage it." (S05: M1)

[26] The endline pupil survey was left open to capture the responses of these pupils who finished the lessons after half term.
[27] RPF instructions had been for teachers to set pre-instruction tasks as work to be completed at home where possible. For any remaining pre-instruction tasks, the instruction was to ask pupils to complete them alone and in silence at the beginning of the relevant lesson.

Multiple Choice Questions (MCQs)

Stages of the activity

The interviewed teachers reported using the MCQs in their lessons but there was wide variation in whether the stages of the MCQs (as instructed within the Peer Instruction Approach) were followed. These stages of answering MCQs are specific to using MCQs within Peer Instruction, and do not necessarily correspond to how teachers would 'normally' use MCQs. The stages of a MCQ within the Peer Instruction approach are: question is displayed to class; solo response; peer discussion; peer response; share results; class discussion. Outside of the Peer Instruction approach, teachers might use MCQs as an assessment tool or to generate group discussion, but are unlikely to go through the 6 stages described above.

Within the training, teachers were told: "*The peer discussion in Step 3 is particularly important and should run for two to four minutes. This can feel like a long time, but it is time well spent!*". As most lessons contained four MCQs, in total, pupils could spend around a quarter of the total lesson time (if lessons are 1 hour) in the peer discussion stage of the questions.

In practice, teachers seemed to have condensed the stages of the MCQs, in order to get through the questions more quickly. Sometimes teachers just shared the question and asked for a solo response before the class discussion, whilst others skipped the solo response and went straight to peer discussion.

For example, in one school, the teacher reported that they normally adhered to the following structure: the teacher displayed the question, the pupils talked for two minutes in their groups, then the whole group decided their response which they wrote on a whiteboard and the teacher facilitated a class discussion. In another school, in a lesson we observed, the teacher displayed the question, asked for silence as the pupils had thinking time, then asked for solo responses which led into the whole-class discussion. In this school, there was no peer discussion time for any of the three MCQs asked in the lesson observed. From lesson observations, where teachers skipped (or de-emphasised) the peer discussion time for the MCQs, there was then little opportunity for peer discussion during the lesson, and the scope for increased collaboration within Peer Instruction lessons might not have been realised.

Teachers noted that they tended to skip the peer discussion if the solo responses had indicated pupils' good understanding of the concept. Whilst the online training stated that: "*Teachers should shift the focus away from getting the correct answers, and instead promote the participation and discussion aspects of the approach*", it seems that both more and less experienced teachers identified that they had sometimes missed or dramatically shortened the peer discussion sections, perhaps reflecting a lack of salience of this message in the online training resources or teaching materials.

> *"If there was any misconceptions then we would discuss it. If they all got it right we moved on, so it was just only if there was any misconceptions there really." (S05)*

In this example, the teacher's comment implies that they might have been thinking of the MCQs primarily as a formative assessment tool as opposed to a discussion generating tool.

It's possible that fidelity to the collaborative element of the multiple choice questions was lower than anticipated across the treatment schools. Pupils in both treatment and control schools were asked the question 'do you normally solve problems in groups?' in the endline survey. Despite the emphasis on group problem solving in the Peer Instruction approach, a similar proportion of pupils in the endline sample in both the treatment and control groups (around 20%) reported that they '*mostly solve problems in groups*'. As we would expect, more pupils in the control group than the treatment group reported that they '*mostly solve problems on their own*' but the size of this difference was smaller than expected (36% control group, 28% treatment group).

Having 28% of the treatment group report that they '*mostly solve problems on their own*' suggests that implementation in some treatment schools may not have led to increased group work, though this survey question provides an imperfect measure of fidelity to the intervention for a few reasons: a) pupils might interpret 'mostly' in different ways; b) the Peer Instruction lessons did still contain tasks/problems which pupils were expected to solve independently, so it's possible that pupils were completing the collaborative elements of the MCQs as intended but still felt that they were 'mostly' working independently; and c) a fairly large proportion of pupils in both treatment and control groups (just under 20%) responded 'don't know' to this question.

It's possible that the in-person training being replaced by online training (because of the COVID-19 context) contributed to lower fidelity to the collaborative elements of the Peer Instruction intervention. Another possible contributor to lower fidelity in the use of group discussion is that teachers may have actively decided against using group work during the intervention period because of high COVID-19 cases. Whilst none of the interviewed teachers reported avoiding group discussion as a health and safety measure, it's possible that other teachers in the treatment group were less likely to encourage group discussion for this reason.

Given that the use of MCQs to prompt group discussion was a key element of the logic model for the intervention (see Figure 3) and the starting point for the hypothesised mechanisms, this variation in teacher delivery of MCQs might have limited the impact of the intervention.

Group set up

The teachers were instructed not to set up groups in advance, and for these to be 'informal' groupings. A target group size was not given.

Teachers varied in whether they set up small groups or pairs for the 'group' discussions. In one school, the teacher changed the pairs (or threes) based on pupils' solo responses to the question. In the other schools, teachers asked pupils to talk to their peers who were sitting closest to them, often having thought carefully about the most effective seating plan.

Some teachers found that grouping pupils with friends led to good quality discussion. These teachers arranged the seating plan so that students were sitting near their friends.

“Let them collaborate with the ones they felt comfortable collaborating with. I think that did work better.” (S02)

“I’ve tailored the groups to encourage social connection. Students that I know are friends and that work well together.” (S01)

Pupil MCQ responses

The RPF resources recommended that teachers use a voting system that was either manual (e.g. pieces of coloured card, mini whiteboards) or digital (e.g. clickers, web-based quizzes such as Socrative). **Teachers tended to use mini-whiteboards or ‘hands up’, although some teachers attempted to use digital voting systems some of the time.**

When teachers used a digital way of responding, they noted that it worked well but was not feasible to implement all of the time, as it was time-consuming to set up.

“It’s like a QR code type thing they just hold up, and I could scan it, and it would tell me exactly what answers they were all holding up, and nobody would know, so it was really quite anonymous except to me, which was quite nice.” (S02)

Teachers accepted the disadvantages of using ‘hands up’, particularly pupils having the immediate comparison of how other pupils had responded, but often found this was the most practical to implement, as it didn’t require the preparation of any resources.

“"From a practicality point of view, that's [Plicker[28]] another thing for prep, for them to put all that stuff out. [With hands up] people get a bit tentative. I don't want them being led." (S02)

Worksheets

For each lesson, there was a worksheet which detailed the pupil tasks and (when appropriate) gave the starter code for the pupils to work from.

The interviewed teachers reported that they set the worksheets digitally on the schools’ online learning platforms, as well as printing paper copies when necessary, for example, if pupils needed to do ‘working out’ on the sheets. Teachers did not generally make changes to these worksheets before setting them for students.

Power-points

The power-points were used as intended by all of the interviewed teachers, with little adaptation. Some teachers reported that they felt they should not adapt the power-points, because they were taking part in a research project. This research context might have therefore contributed to the high fidelity in the use of the presentations.

"Normally when I get something like that [power-point slides] *I will change it to my teaching style, but I didn't for the purpose of this* [research project]*." (S04)*

“ [I was] *going through materials and trying to honour what was there as closely as I could, and not go off and do completely my own spin on the thing.” (S02)*

[28] A digital way of gathering pupil responses by scanning the classroom for bar codes which the pupils hold up.

As instructed in the lesson materials, the teachers in the observed lessons switched between the power-points and modelling live-coding in Python. The teachers did this confidently and with minimal time spent changing between the two.

4.1.2 Feasibility

Key findings:

- Teachers reported that the amount of content in the lesson plans was difficult to fit in to their timetabled lesson time.
- The MCQs worked well where pupils had well established routines for responding; these questions would have been easier to deliver had they been on the same power-point as the main lesson presentation.
- Preparation of the lesson resources was a fairly time-consuming task for teachers.
- While teachers were generally positive about the online training, there is scope to adapt the training so that it builds teacher confidence to a greater extent.

This section explores the factors which affected how easy or difficult it was for teachers in the case study schools to implement the intervention as intended.

Amount of content in the lesson plans

From discussions from teachers and observations of lessons, it emerged that the amount of content in the lesson plans was challenging to fit into the allocated lesson time of 50 to 60 minutes. In some schools, teachers responded by cutting the amount of time for discussion of the MCQs, whilst in others, they preserved the discussion time, but cut time for pupils to work on their coding tasks.

"I've had occasions where they've all got it [MCQ] right, but then I've had other groups who are really struggling with someone so most of the lesson has been a discussion and not on the main tasks." (S03)

"*I knew I wasn't giving them enough time to do it [the coding task]...It tends to be the middle I'll tend to squeeze by the way I've worked, but obviously it depends on where the key content is coming." (S02)*

Teachers reported that timing was particularly a challenge with Peer Instruction lessons, because the amount of time spent discussing MCQs depended on pupils' understanding, which would vary from lesson to lesson and class to class.

"*I ended up flowing into the next lesson sometimes. It's hard to predict because I could do one lesson with one class and it can go completely different with the next class because of how much time it has taken to do the multiple choice questions." (S03)*

Some schools had lessons which were less than an hour long, and even in those which did have an hour allocated to computing, the teachers were not able to spend the full 60 minutes on the lesson, as transition time and other administrative tasks (e.g. the register) also ate into the lesson. Interviewed teachers reported feeling rushed during the lessons: this was consistent with the observed lessons in which teachers were trying to start their final activity (for example the final MCQ) one minute before the lesson ended.

"I probably missed half of what the lesson was telling me to do." (S03)

Factors affecting the feasibility of the MCQs

In schools where MCQs were already frequently used, teachers reported the Peer Instruction MCQs working well, as pupils were familiar with the format of questioning. It's worth noting that, whilst pupils may have been familiar with responding to MCQs, it's unlikely that they had previously been used by teachers as part of a Peer Instruction approach, with the multi-stage process and a focus on peer discussion.

"*That's [MCQs] been a big drive at my school for two or three years, so there wasn't that much change into what we were doing, but it was nice to have those hinge questions ready to use. The questions worked great for us because we use them all the time." (S04)*

Teachers reported that the MCQs worked particularly well when pupils were in an established routine of how they responded to questions. This meant that little time was spent on the practicalities of answering so more time could be given to focused thinking and discussion. In some cases, teachers spent time establishing these routines at the beginning of the unit.

"We spent a lesson just teeing them up...we just got into the habit…it's like right, all absolutely silent, it's [the MCQ] gone up." (S01)

"When they come into the rooms, they automatically bring their whiteboards and pens out…normally you have to say 'on your whiteboards, let's go', but you didn't even have to say that. They saw the 'Q' [on the slide] and straightaway it was: get the whiteboard out. They knew what was coming. (S04)

One challenge which all teachers described was having the MCQs on a separate power-point to the main lesson power-point and either having to switch between presentations during the lesson or copy and paste text from one into the other in advance of the lesson. The resources were prepared in this way to give teachers flexibility in which questions they used in their lesson plans but the teachers we spoke with felt that it would have been more helpful for the MCQs to be in the main lesson power-points.

"It would be easier if the multiple choice questions were within the main lesson power-point" (S01)

Teacher subject knowledge and confidence

The interviewed teachers reported that they were confident in teaching the content of the peer instruction lessons themselves but did not feel they could set lessons for cover teachers or non-specialist teachers. Whilst schools would hope for all KS3 computing lessons to be taught by computer science teachers, some schools did not have enough specialist teachers to cover all of the classes, which meant that the classes taught by non specialists were taught a different unit of lessons which were easier for non-specialist teachers to use.

"That's me worried that there's going to be a non-specialist [teaching computing]. *Even though they* [lesson plans] *were so detailed, I don't think they* [a non-specialist teacher] *could do it properly…if you haven't got that knowledge." (S03)*

"It was difficult to set this when I was off- you can't get a supply teacher to do it." (S05: M2)

The teachers who were delivering the lessons demonstrated that they could use their subject knowledge to facilitate discussions around the MCQs. In observations, these teachers used skillful questioning to uncover pupil misconceptions and then supported pupils to work out for themselves what the correct answer was. This required a level of clarity of explanation which would not be possible without a secure understanding of the underlying subject.

Lesson preparation

Teachers generally reported that preparing the lessons was time consuming, although one teacher said that preparation took a comparable amount of time to her experience of working from other units of lesson resources.

Teachers found that collating the separate resources for one lesson took a long time. Some spent time before the lessons pulling all of the materials into the one teacher presentation so that everything was in one place.

"The amount of resources that need to be pulled together for one lesson…there's seven different documents for one teacher to look at for one lesson. It's a bit overwhelming." (S05: M1)

The 'Representations' unit tended to be more time consuming to prepare for than the 'Python' unit, as it involved more paper based-preparation.

"There was an awful lot of prep work for those [Representations] *lessons…it involved paper, guillotines…stuff that I wouldn't normally do. That was a lot of effort to suddenly realise I've got to schedule this in amongst all my other lessons that we're doing." (S02)*

Interviewed teachers reported that some of the lessons (particularly those in the 'Representation' unit which required resources such as instruments and torches) required collaboration with other departments. Where colleagues from other departments were keen to support, this collaboration worked well:

"We even had the music teacher come up and speak to them about how they could do Morse Code with their bodies and triangles." (S03)

In some cases, this need for collaboration was a barrier where the other departments weren't able to support the computer science teachers. For example, teachers emailed colleagues asking for resources but received no response. Teachers responded to this by adapting the lessons so that the called-for resources were not required.

Online training

Reach

19 out of the 21 treatment schools which completed the endline pupil surveys had at least one teacher who completed the online training feedback survey. Of the two schools that didn't complete the online training feedback survey, it's possible that they had teachers who completed the training but not the survey. In most schools, more than one teacher completed the training: this is because more than one computing teacher will generally teach across year 8. In some schools, teaching assistants also completed the training, which teachers reported was helpful.

"The LSA [Learning Support Assistant] *I have for one of the classes did it* [the training] *as well, and she understood why it was important…that was really good that I didn't have to explain it to her as well." (S03)*

Time commitment

There was variation in how long school staff reported spending on the online training: from 1-2 hours (23%), 2-3 hours (25%) to 3-4 hours (44%). The final section of the training, 'Explore the resources' was a fairly open task, with school staff directed to 'take some time to read through the Units overview and lesson plans'. It's possible that this flexibility within the training contributed to the variation of time spent on the training.

Effectiveness

A similar proportion of the teachers who reported they were either 'quite confident' or 'very confident' in teaching computing to KS3 pupils (92%) also reported they were 'quite confident' or 'very confident' in delivering the Peer Instruction resources (93%). More teachers reported they were 'very confident' in teaching KS3 pupils in general (66%) than in delivering the Peer Instruction resources (25%) but it's possible that, at least in part, this is because the teachers had not delivered the Peer Instruction resources before.

Teachers generally reported that they were happy with the training they had been provided and felt that it had given them a good understanding of the aims of the intervention as well as the lessons to be taught.

"It worked, it prepared me pretty well…It was clear. Pedagogy-wise, it all made sense." (S02)

Some teachers reported that they found it a challenge to get to grips with the amount of content and how to access the different resources. It's possible that the fairly open task of 'taking some time to read through the Units overview and lesson plans' could be made more structured within the training to support teachers to navigate the resources.

"Took me a while to get my head round it… to sift through all of the instructions. Something was on the website and something was on google drive. It took a long time to figure out what it is that I'm actually supposed to be doing in the first lessons." (S05: M2)

One teacher suggested that it would have been useful to have an online forum through which to discuss their experience of delivering the lessons with other computing teachers.

"I was thinking that there's a lot of prep and cutting up here. Is that what everyone else is going through? Any smart ideas?" (S02)

It's worth noting that the online format of the training was in response to the COVID-19 context, and was not what RPF had originally intended. Some of the points related to the reach, time commitment and effectiveness of the online training may be relevant if RPF were to deliver training online in the future.

4.2 Intervention

4.2.1 Quality

Key findings:

- Teachers felt that the quality of the lesson plans and other resources was high and provided them with everything they needed to teach the lessons.
- They were particularly enthusiastic about the MCQs; some had even started developing their own banks of questions to be used in other topics

This section explores teachers' perceptions of the quality of the Peer Instruction resources.

Lesson plans

Generally, interviewed teachers felt the quality of the lesson plans was good, and that they included all of the detail required. In some cases, teachers felt like more detail was included than was necessary, and that this detail meant that it took some time to understand what should happen in each lesson.

"They [the lesson plans] *have been pretty good. I don't need all the material that's there, but knowing I've got a very detailed lesson plan and an overview one, I found that really quite useful for the first time to go through." (S02)*

"For lesson 2 of the first unit, I had to go away and really read it and think what are the…because I couldn't get what they were trying to grasp at. Once I'd started teaching it, I was like: ah, that works." (S04)

Teachers felt that the level of challenge was pitched appropriately for non-set classes (classes which contained all pupils, i.e. were not streamed based on computing attainment). For top sets (highest attaining computing students), teachers used the same material but completed the lessons in a shorter amount of time. Data from lesson observations also suggested that the level of challenge was appropriate: pupils were able to make progress on the specific lesson objective of the lesson they were on, and had clearly developed their understanding from the beginning of the lesson to the end.

"On the whole, the content was pretty spot on [in terms of pitch]*." (S95:F1)*

"The top set, they were getting it, she [the top set teacher] *often had to do a lesson and a bit within one lesson because they were getting through it quicker than my* [non-set] *group were." (S03)*

Interviewed teachers reported that the progression of the lesson objectives worked well, gradually building pupils' confidence.

"The lesson resources were very good and they did progress well. It kind of eased them into it. When they finally did the Python Programming, they were like 'oh right - I understand why we're doing this now.' In the past, we've just started off with Python Programming and they've already made up their minds that they can't do it." (S05: M1)

MCQs

Teachers' appreciation of the MCQs came through strongly in the teacher interviews. They reported that the questions were clear and helped to pick up on pupil misconceptions. Teachers found it particularly helpful to have the questions prepared for them, as they had found that creating their own questions was very time consuming.

"They're a good set of multiple choice questions. I'm going to look at what I've got in other modules and see whether I can enhance them with some better questions sets that go along those lines. They're very simple and they're not intimidating for pupils."

Power-point slides

Teachers generally felt that the power-point presentations were of good quality and well-structured.

"*The power-points are great; they're really easy for the kids to understand, and they're sequenced in a logical way to walk the students through each topic, which I really like." (S01)*

Some teachers felt that the power-point presentations were too 'text heavy' which was unengaging for the pupils. However, this feedback wasn't apparent in all of the teacher interviews and teachers were able to adapt the power-point slides as they saw fit.

"They're very busy. They're very click, click, click all the way through." (S04)

In the lesson observations, the teachers did not read aloud all of the text that was on the slides; they skipped some sections or rephrased text as questions. In this way, the teachers ensured that they covered the content necessary, but also that this was done in an engaging way.

"I ask them questions that aren't on the power-point, but that's every teachers' style normally. You'll find something you think, oh, I can question them over that - it doesn't need to be on the power-point." (S01)

4.2.2 Responsiveness

Key findings:

- Pupil engagement with the MCQs was high for all pupils - some teachers felt it was particularly high for girls.
- The text content of the pre-instruction tasks meant that some pupils struggled to access them - teachers had various ways of dealing with this challenge so the pupils could engage with the tasks.
- Teachers generally reported higher levels of engagement with the Python unit (which involved more time at computers) than the Representations unit.
- Teachers are intending to continue using the resources with future cohorts and some plan to (or have already) use MCQs with other year groups.

In this section we explore the extent to which pupils and teachers engaged with the various elements of the Peer Instruction lessons.

Pupil responsiveness

MCQs

Data from teacher interviews, pupil focus groups and observations indicated that the pupils enjoyed responding to the MCQs and that within each class, the majority of the pupils generally engaged well with them. Teachers reported that they enjoyed hearing the discussions between peers and that groups were often very keen to share their feedback. The lesson observation in which the pupils completed the group stage of MCQs supported this finding: pupils were having detailed discussions in their groups about the differences between the options, with pupils taking it in turns to talk and building on one another's responses.

"Multiple choice questions were fantastic- the kids love them."(S05: F1)

"Everyone wanted to get their opinions across in groups…I quite enjoyed that because you've got real, clear, positive engagement." (S01)

In the observation lessons, teachers skillfully managed their classes, to ensure that every child gave their solo response. For example:

"[Girl name] just waiting on you." (S03)

Some teachers felt that pupils responded particularly well when working in groups with their friends, as the pre-existing social relationships made collaboration easier.

"Not only do they feel like they're able to talk to their friends and give their opinions, but because they've already got those social connections, they're actually able to listen to each other as well, without overriding each other." (S02)

The interviewed girls similarly reported that they preferred working in a group (as opposed to answering questions individually) as they were able to hear from people who had different ideas to them and check their answers.

“I prefer talking in a group because then you get the other side of other people’s thoughts” (S02: female pupil)

The interviewed boys reported that they valued group discussions within the MCQs, but focused more on the importance of having time for individuals to consider their answers first, before talking as a group.

“I like when the question is first asked you do it by yourself, then when people have got their answers, then you discuss it” (S02: male pupil)

“Yeh like think pair share, you can have a bit of time to think for yourself then you can bounce ideas off other people” (S02: male pupil)

Pre-instruction tasks

Teachers and pupils reflected that the reading expectations for some of the pre-instruction tasks were high for some of their year 8 pupils who were less strong readers, and that some pupils struggled to access them independently. Teachers generally found ways around this, such as putting the pre-reading task into Microsoft Teams so that pupils could use ‘immersive reader’ which reads aloud the written text, teachers reading the pre-instruction task aloud to their class at the start of the lesson or teachers recording themselves reading the text and sharing that with pupils.

“I got a lot of emails from students saying ‘I just don’t understand it. It’s too long. I can’t read it.’ “ (S02)

“In [Microsoft] Teams, they’ve got immersive reader support, so they actually listen to it being read….It makes it a lot easier for those that may struggle a little bit” (S01)

Some teachers reported that, even where pupils were able to tackle the level of challenge of the pre-instruction tasks, they did not necessarily complete the tasks outside of lesson times. These teachers tried to motivate the pupils to complete the tasks, telling them that it would help them in their next lesson.

“What I had to do for it to work was….put the date of when it was due and reminding them actually, you need to know this to be able to be more confident.”(S03)

In line with this, in some schools, teachers reported that the fact that pupils knew that the first MCQ in the lesson was likely to be linked to the pre-instruction task made them more likely to engage well with the task itself.

“I’d say there’s been an increase in engagement because students know that they’re going to be tested on it in the next lesson.” (S02)

Main teacher input

In the lessons that were observed, pupils engaged well with the teacher input: they responded to their teachers’ questions, listened closely to the teacher and engaged in focused discussion. To increase engagement further, some teachers went ‘off-script’ with explanations or tasks they thought their pupils would enjoy. For example, in the sequencing lesson, one teacher ‘cheated’ by getting the Python code to show him the ‘lucky number’ so that he could correctly guess the ‘lucky number’ each time it changed. Pupils responded to this by laughing and making other suggestions.

Some teachers felt the amount of content on the power-point made the lessons less engaging for pupils. Teachers acknowledged the sensitive balance of how much content to include in the power-points: more content is supportive of teachers with less subject knowledge but could also make the lessons more ‘power-point led’ with less flexibility for the teachers.

“There was lots and lots of content that actually I think switched the students off, because it was very much power-point led rather than teaching led, but I get that, because of the non-specialists so I totally do.” (S04)

Responsiveness differed across lessons

Teachers reported that pupils’ engagement varied between lessons. In general, pupils tended to engage more with the Python lessons (as they had the opportunity to work on computers themselves), than the Representations lessons.

“The first unit was a bit dry, but the second one, they really got into it, yes” “I think [in the first unit] we had one or two lessons where there was no computer interaction at all…that was quite difficult then, I guess because they’re just used to coming in and they know they’ll be computer based.” (S03)

Whilst engagement was generally higher for the Python unit of lessons (compared to the Representations unit of lessons) the individual lessons teachers and pupils identified as particularly engaging tended to be within the Representations unit. These were the lessons that included practical activities, such as the ‘Across Time and Space’ lesson in which teachers set up a ‘museum’ within their classrooms and the ‘Lights and Drums’ lesson in which pupils used torches and instruments to send messages in Morse Code.

[On the museum lesson]*“ You just got that nice classroom buzz, you can hear the kids talking to each other ‘ohh is that what that means?’...Those little bits of classroom gold.” (S05: F1)*

“Another group did Morse Code, but they did it with blinking. Most of them did it with torches and things, but they really enjoyed that and they remember it really well because it was something different.” (S03)

One teacher felt the level of pupils engagement with this unit of Python lessons was greater than pupil engagement when they had taught this topic previously.

“I’ve taught Python lessons before and they were definitely more engaged this time. The class I’ve got for this - there are a handful of silly boys who just want to mess around and I could tell they were really engaged with the programming whereas in previous lessons they were saying ‘I can’t do it’.” (S05: M1)

Gender

Some teachers reported that girls engaged particularly positively with the intervention lessons. In one case, the teacher reported that they felt the girls engaged better with the multiple choice questions than the boys because they were more likely to read the questions carefully.

"The girls have generally responded better, not necessarily because they're brighter, but because they're reading the question carefully. Whereas the boys are rushing it more…that's a trend I'm seeing. They [the boys] *think they know the answer but they've not read* [all the options]. *(S03)*

Teachers reported that they were pleased with the level of girls' engagement with the lessons, particularly in terms of their contributions during group or paired discussions.

"There were very few girls who weren't engaged. The one or two individuals that were sometimes coasting it out, they were all boys. The girls were all getting involved, even if some weren't a huge amount." (S02)

"I was very pleased that a lot of the girls were doing a lot of the talking. It was one of the binary lessons, doing the binary conversion" (S05: F1)

Whilst girls' engagement in discussions was also noted in the lesson observations, the girls tended to be less confident speaking in front of the class and, across both schools, tended to speak more quietly so the teacher needed to repeat what they had said to the rest of the class. Where there were vocal interjections in these lessons, they were generally boys.

Another teacher reported that, in their school, girls were more likely than boys to complete the pre-instruction tasks, thus setting them up well for the MCQs in the lesson.

"The flip learning was a positive because it was more the girls that would do that than the boys" (S04)

Teacher responsiveness

One theme which emerged during the case studies was teachers' enjoyment of using the Peer Instruction resources, and the value for experienced teachers to be encouraged to try out a new teaching approach.

"It makes you change your thought processes - you get out of the old habits you've now taught it in for ten years in a row." (S04)

"It took me out of my comfort zone - it made me teach differently to be honest." (S03)

Teachers were also positive about the MCQs in their lessons: they felt it was a useful tool for assessment as well as for supporting pupils' engagement.

"It's getting away from the old-fashioned 'hands up, who knows the answer?' to making sure everyone is responsible for their own answer." (S05: F1)

Whilst teachers did voice some challenges they had faced with setting the pre-instruction tasks, and concerns around the level of accessibility, they generally liked the idea of pre-instruction tasks, and some suggested that this would prepare pupils well for other subjects or future study.

"I really like the fact we've got pre-instruction in, because I'm already doing pre-instruction when we get to Years 10 and 11, and I want to see it happening earlier." (S02)

Teachers' intention to continue with Peer Instruction

The interviewed teachers generally intended to re-use at least some of the Peer Instruction resources, but varied between whether they would be using the approach more broadly (for example using some of the activities and creating multiple choice questions for different units) or whether they would use the exact lesson resources again with year 8s in future years.

"We'll do it again because it covers some complicated topics in a friendly way." (S01)

"I think we'll be doing it next year. Next year we'll be better at it as teachers." (S05: F1)

"Why wouldn't I take some elements of what we've done here? I like the collaboration aspect" (S02)

Some teachers reported that they had already taken elements from the Peer Instruction resources and applied them to other year groups.

"[in Year 11] *with loops and for-loops they weren't getting it. So we made our own questions, showed them some code, and they seem to be more comfortable with multiple choice and discussing. Miss X did it with year 11 and it worked really well." (S05: F1)*

4.2.3 Mechanisms

Key findings

- There was little evidence for the two pathways hypothesised in the logic model ('participation in discussion leading to increased subject knowledge'; and 'through discussion, girls observing similarity in level of understanding across the class')
- Two alternative pathways were suggested by teachers and pupils in the case study schools: 'pupil help' and 'experience of success'.
- Teachers felt that this intervention alone was unlikely to meaningfully shift girls' intention to study computing, but that it could be a contributing factor.
- Barriers to the working of intended mechanisms include school options systems, high baseline levels of engagement with computing, and the social influence that computer science is seen as a 'male' subject.

At its broadest level, this intervention aimed to use a Peer Instruction approach to increase collaboration in computing lessons which would, in turn, improve girls' attitudes towards computing and their intention to continue to study the subject. The section of the logic model in Figure 7 sets out the hypothesised mechanisms through which the intervention was designed to affect the intended outcomes. This section explores the extent to which the data from the IPE support the hypothesised mechanisms within the logic model.

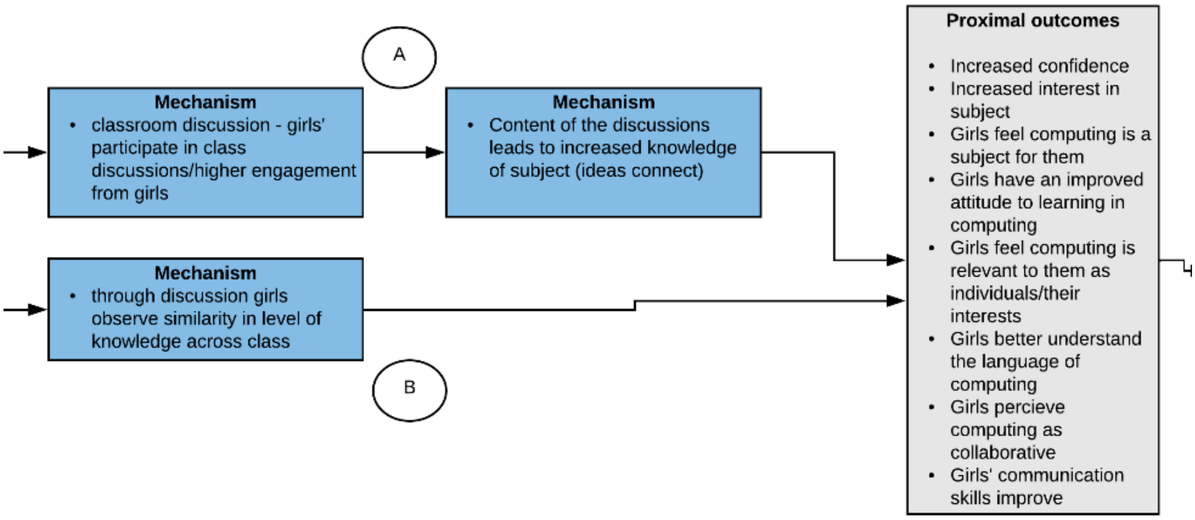

Figure 7: Peer Instruction Logic model mechanisms and proximal outcomes

Mechanisms observed

Pathways hypothesised by the logic model

Within the case study schools, teachers and pupils provided some tentative evidence for pathways A and B of the logic model although the links between the stages of the pathways were not always clear.

Pathway A of the logic model

Through classroom discussion girls' participation in class discussion/engagement increases → Content of the discussions leads to increased knowledge of the subject (ideas connect)

Evidence from the four case study schools did support the idea that girls participated more in class discussion as a result of the Peer Instruction lessons. The peer discussion during the MCQs seemed to be the part of the lesson in which teachers noticed an increase in girls' discussion.

"I was very pleased that a lot of the girls were doing a lot of the talking." (S05: F1)

Some teachers also commented that they felt girls were improving in their answering of MCQs. However, none of the teachers directly attributed this improved understanding to pupils' discussion and, as the Peer Instruction lessons might have differed from 'business as usual' in a wide range of ways, it's not possible to say with confidence that it was the engagement with group discussion which caused the improved knowledge.

Pathway B of the logic model

Through discussion, girls observe similarity in level of knowledge across the class

From the evidence gathered from the case study schools, there were no indications that discussion led to girls observing similarity in level of knowledge across the class, which then in turn affected proximal outcomes. Whilst teachers did comment that they felt girls' confidence had improved, they did not attribute this increase in confidence to an observation of level in similarity, but to being part of the Peer Instruction lessons more generally.

"When we initially started it [the Peer Instruction lessons] *you had a lot more of the girls and the ones that lack confidence putting a question mark, because they didn't feel…It's happening less* [girls putting question marks] *now we're further through." (S04)*

Additional pathways observed

Whilst the implementation and process evaluation did not generate strong support for the two pathways proposed in the logic model, the evidence did suggest two additional potential pathways through which the Peer Instruction approach could affect intended outcomes. The two pathways are: Pupil help (Mechanism C); and Providing an experience of success (Mechanism D).

Pupil help (Mechanism C)

Pupils felt positive about the scope for peer help during Peer Instruction lessons and identified being able to ask classmates for support as one of their favourite elements of the units of lessons. Girls reported that they valued being able to 'make sure' they had answered the question correctly and were generally considering this advantage from the perspective of someone who would be receiving help, rather than offering it.

"I think they [Peer Instruction lessons] *are actually better than the* [normal ones] *because you can ask other people around you to help more." (W02: female pupi)*

"One of the kids (who struggles) said to me, 'Miss, you didn't teach us at all today,' he said. 'Girl X taught me everything today.' He was paired with Girl X, and I was like, 'Well, that's what it's about.' " (S03)

Data from the teacher interviews and lesson observations also identified peer 'help' as a central aspect of these Peer Instruction lessons. Within the observed lesson, the teacher was directing particular pupils to help other pupils who needed support. Both the pupils offering support and those receiving support responded well to this: talking on task about the specific challenge they had 'got stuck' on.

"*Because of the way this unit was done, there was an encouragement to help each other out, which they might not normally have done so much" (S05: M2)*

Teachers and pupils reflected that the help from other pupils might be particularly valuable because it means less confident pupils can ask for help in a more secure (and less public) context than suggesting an answer in front of the whole class and possibly getting it wrong. This was linked to the proximal outcome of increasing girls' confidence in computing.

"Because you might be scared to tell the teacher that you don't get it - you can tell your friends and then they can help you to understand [when you're doing the task by yourself]*" (S01: female pupil)*

"It [girls talking more in the lesson] *might have been something to do with the fact they felt it was OK to discuss it with each other. I don't know if that's a bit different to the more individual work we'd usually do….I was very pleased to note that a lot of girls felt confident and were talking about it."" (S05: F1)*

It is worth noting that teachers from different school contexts identified that the group discussions created a more 'secure' environment for girls.

"[in the past] I didn't do a lot of group work and they enjoyed [in the Peer Instruction lessons] *the fact that they worked together on the questions. They really like that. They felt more secure, I think." (S03)*

The other proximal outcome which teachers and pupils associated with this pathway was girls' sense of computing as a collaborative subject. Pupils reported that they would not necessarily have identified collaboration as an important skill for computing if they had been asked before they had done the units of Peer Instruction lessons.

"I think teamwork was quite good [high in the ranked order of skills necessary for Peer Instruction lessons] because if you didn't understand it then you can ask your friends and they can help you" (S01: female pupil)

"I think I would have had different answers [if asked before we did Peer Instruction] *because we didn't work together so much. Before I would have said perseverance because most people just did it on their own before." (S03: female pupil)*

Experience of success (Mechanism D)

A further pathway between Peer Instruction lessons and intended proximal outcomes (particularly pupil confidence in computing) was pupils' experience of success. This was not explicitly referred to in the logic model but could be thought of as close in concept to pathway B: greater discussion leading to girls' observing similarity in level of knowledge across the class.

Both pupils and teachers noted that the structure of the lessons, combined with the MCQs, provided many opportunities for girls to experience success, which in turn, boosted their confidence. In particular, teachers thought the structure of the unit and the way in which activities build on previous activities set up pupils to do well.

"Even the ones that are your middle ability, they've had those lightbulb moments, which have been really nice. Because it's been short, sharp, go-go-go, they've had that achievement." (S04)

Pupils reported that receiving immediate positive feedback during the MCQs could build confidence as it 'proves' that you understand the concept.

"I think it may have persuaded some of them [some of my friends to take GCSE Computer Science] because if they get the questions at the start or the end correct, it gives them a confidence boost." (S03: female pupil)

Distal outcomes

Interviewed teachers generally felt that the Peer Instruction lessons were unlikely to have a dramatic impact on whether girls chose computer science as a GCSE suject, but that it's possible that it contributed in a positive direction. The caution in estimating impact seems to be more due to teachers' perceptions of the scale of the challenge of persuading pupils to take GCSE Computer Science, as opposed to because of concerns they have about the Peer Instruction units themselves.

"To say 'Oh right, it changes the game a bit for the girls', I'm not sensing that. I'm just sensing that the lessons became more accessible for some pupils there. I'm very positive about what's going on. I like the collaboration. I like the intent behind it. I think it's very difficult to say there's a clear link for girls opting. There's going to be lots of things that helped contribute to that, but it's definitely worth considering the approach in some parts of that." (S02)

"I'm praying that it makes a difference to how many girls choose GCSE but it's hard to say off the top of your head. I think it's a piece of the puzzle and we will definitely carry the idea of the collaborative Peer Instruction forward" (S05: F1)

The interviewed pupils similarly reported that they thought it was unlikely the units of lessons would 'change someone's mind' about whether to take GCSE Computer Science, but they did think the lessons might change some pupils' perceptions of computing. For example, one girl described changing her perception of what computer science was about as a subject, as a result of the Peer Instruction lessons. When considering whether the Peer Instruction lessons would affect his friends' (both boys' and girls') GCSE decision making, one boy responded:

"I think maybe not change their minds, what they're going to do for the GCSE, but more change how they think about computing as a lesson." (S01, male pupil)

Specificity of the mechanism to girls

Mechanisms C and D above (help provided and experiences of success) were generally suggested by teachers as specific to girls' engagement with computing. Whilst one teacher felt that the positive engagement with lessons as a result of the Peer Instruction intervention was equally apparent for girls and boys, others reflected that elements of the lesson specifically supported girls' engagement. This is possibly due to the fact that girls' confidence was identified as starting at a level lower than boys'.

"The girls who wouldn't really have liked computer science before were very shy, they weren't keen to put an answer down." *(S03)*

Girls having lower baseline confidence than boys in computer science was also reflected in the gender difference in baseline subscale scores (see table 2): on average, girls responded less positively to questions such as "I am confident that I can solve problems by using computing" and "I am good at learning computing skills on my own."

Barriers and facilitators of the mechanisms

School level factors

Teacher skill in managing questioning

High levels of teaching skill enabled the teachers to successfully deliver the Peer Instruction lessons. The observed teachers demonstrated substantial skill in managing the MCQ sections of the lesson: this involved making rapid decisions about when to move on and when to spend longer on a point; using their subject knowledge to give clear explanations of the four possible answers; and consistently and throughout the lesson praising pupils for their positive engagement and effort, particularly ensuring that they praise their female pupils.

“Excellent answers, [Girl X] is using her knowledge of history - fantastic!...Well done girls, you’ve not done it already?! Amazing!” (S03)

It’s possible that the teachers who were happy to be part of the evaluation were particularly confident and experienced teachers who manage the challenging aspects of teaching the Peer Instruction lessons well (e.g. time management). These teacher characteristics may not fully apply to the ‘average’ teacher.

Baseline levels of school computing encouragement for girls

Participating schools seemed to be already using multiple strategies to engage girls in computing. As the schools which took part in the GBIC intervention had volunteered to be part of a research project investigating gender balance in computing, it’s unsurprising that these schools also tended to have other initiatives which aimed to increase girls’ engagement with computing, such as code clubs, female external speakers and trips.

The teachers themselves said that they felt strongly about supporting girls’ engagement with computing.

"We [computing teachers] *give up a lot of time, after school, lunch times" (S02)*

"It's been really good to teach it, and especially the push for girls, for me, that's really important.” (S04)

It is possible that, in these environments where the teachers were already actively trying to promote girls’ engagement with computing, there was less scope for the intervention to change girls’ perception of the subject, compared to a school where teachers had not been actively trying to improve girls’ computing engagement.

School options system

Some interviewed teachers thought that doing the Peer Instruction units would have led to a greater number of girls taking computer science than in previous years, had it not been for their schools’ GCSE options system. Where girls are forced to choose between computer science and other subjects (which tend more towards a female cohort, such as dance), teachers felt that made it less likely that girls would choose GCSE Computer Science.

"Based on this project, I think we would have had a fourth class of computer science, but I just think now we're going to have two or three…My problem is if you're a bright kid, do triple science, and then they've got very little options left. I believe 100% our numbers would have gone up again, with more girls, but now they [the school] *have made French compulsory, so it's giving them less options."(S03)*

"If we're in the same [options block] *group as dance, we just get wiped out by that." (S05:F1)*

At a different school, where the pupils make their GCSE options in year 9, the teacher felt the intervention might have made a difference to choices made, had the lessons taught immediately preceding the decision making process.

"I think if we had been at that point [options] now, I think we would've got a lot more of the girls take-up" (S03)

Perceived social norms

Pupils' perceptions of computing as a 'male' subject

Where pupils have an open perspective on computer science as a gender-neutral subject, it could be that there is more scope for the mechanisms of the intervention to lead to meaningful change. There was broad variation amongst pupils in whether they thought gender was linked to confidence or skill in computing. The interviewed girls who were highly engaged with computing did not think there was a gender difference.

"You don't need to be a certain gender to like computing because you don't need to be a certain aspect like strong or anything, you just have to know what you're doing." (S03: female pupil)

Other girls reported that boys enjoyed computing more and linked this to the amount of time spent playing computer games.

"In our computing lessons the boys definitely understand more than the girls and enjoy it better I'm pretty sure" (S02: female pupil 1)

"Maybe they do more video games than girls, and so they understand how to use computers maybe better than girls."(S02: female pupil 3)

Where pupils are holding these more gendered views, this could act as a barrier to the Peer Instruction lessons leading to the intended proximal outcomes, even if the lessons themselves can increase girls' perception of computing as a collaborative subject.

The scale of the challenge

A final barrier which emerged strongly through discussions with teachers was the power of social influences on pupils' subject choice and the (perceived) relative lack of 'levers' within school to combat the social influence on girls to not take computing. Teachers felt that these social pressures restricted the scope of their efforts in school to increase the gender balance in computing.

"There's this really - not nasty - but there's this trend at the moment that I've definitely seen and witnessed, where boys are expected to know this sort of stuff [computer science skills], in a sense. It's a social sort of, for boys 'Oh, you're technology, you can do this', and the girls …some people seem to

think you shouldn't be able to know how to do this. It's not like that; it should be the other way around. Not even the other way around; it should just be equal." (S01)

"It [experience of computing lessons] isn't the significant factor that's going to be driving choice." (S02)

Teachers talked about the lengths they had gone to to increase the number of girls taking GCSE Computer Science by a small number, only for this progress to disappear the following year because COVID-19 restrictions meant all of the additional activities they had put on for girls couldn't take place.

"We actually had a really nice year last year where we managed to get six girls in the room. That was because we were talking about it so much to them, and talking about women in computing, women in tech. We had a woman come in who was currently working in the cybersecurity sector, she did a talk about it and what her experience has been like and our cohort was brilliant. Sadly, we didn't do that last year [because of the COVID-19 restrictions], so we only got two girls." (S01)

It's worth noting that the baseline differences between boys' and girls' intention to study GCSE Computer Science is large (20% of the boys' cohort vs 7% of the girls' cohort) and girls' average scores were lower than boys' average scores on each of the five subscales of the computing attitudes survey (see table 2). Within this broader context, an intervention designed to increase girls taking GCSE Computer Science will need to motivate girls to choose the subject to such an extent that this motivation overcomes the broader barriers to choosing computing.

5. Conclusions and recommendations

5.1 Summary and interpretation of findings

We observe no meaningful change in attitudes measured or intentions

Overall we found no evidence that the intervention increased girls' intention to study computer science at GCSE or improved attitudes towards computing. Although the attrition observed limited the evaluation's ability to detect an effect on the intervention, given the very small size of the differences between the two groups, we do not suspect that attrition rates are obscuring a substantive positive impact of the intervention.

Teachers and pupils reported positive experiences of the Peer Instruction lessons and the intervention was generally implemented well

Teachers reported that they enjoyed teaching the Peer Instruction lessons, often because it encouraged them to try a new approach. They generally reported that pupils engaged well with the lessons, particularly when there were hands-on activities. Teachers generally included the MCQs within their lessons, although in some schools, there was little focus on the group discussion element of the questioning. Teachers also found that it often worked better for pupils to complete the pre-instruction tasks as a lesson starter, as opposed to a home learning activity. In the schools we visited, teachers used the power-points, lesson plans and worksheets as intended, with few adaptations. It is worth noting that the teachers who were happy to give feedback on the intervention may have been towards the higher end of the implementation quality range of the participating schools. Within their feedback, teachers and pupils identified some delivery challenges (listed below). It is possible that some of these challenges may have limited the impact of the intervention.

Evidence from the case study schools highlighted some implementation challenges

The main implementation challenges identified were:

- **The amount of preparation necessary for teachers to be able to deliver a lesson:** some teachers found the process of collating the resources for the lessons time consuming, and that finding their way around the online resources could have been easier.
- **The amount of content to fit into one lesson:** some teachers struggled to cover all of the content that was set for each lesson within the timetabled time.

Delivery deviated from the intended approach in the following way:

- **Some teachers did not prioritise the group discussion stage of MCQs**: in some case study schools, the fidelity to the group/pair discussion aspect of the MCQs was low; teachers focused more on receiving solo responses as opposed to encouraging pair/group discussion before receiving collective responses. Shifting the training from in-person to online due to the COVID-19 pandemic may have contributed to the variation in the extent to which teachers prioritised group discussion as online training might have constrained the exploration of the collaborative nature of the approach.

While this was not mentioned by the interviewed teachers, it is also possible that some other teachers might have deprioritised group work over the time period of the intervention (October 2021-February 2022) due to high cases of the Omicron variant of COVID-19.

The challenges above, particularly related to fidelity and feasibility, may have limited the impact of the intervention. Two additional possible explanations for the absence of evidence of impact are:

- **The hypothesised barrier is not most critical to the intended outcome:** it is possible that the intervention did indeed increase pupils' perception of computing as a collaborative subject, but that lifting this barrier was insufficient to meaningfully improve girls' attitudes towards computing or intention to study it in future. The barriers discussed by teachers seem to be more linked to girls not seeing computing as 'for them' as opposed to girls specifically struggling with seeing computing as a more individualistic subject.
- **The large and structural barriers preventing girls from choosing GCSE Computer Science can't be overcome by a single strategy:** this intervention might need to take place alongside other efforts by the school (or possibly even broader social changes) in order to provide sufficient motivation for girls to overcome the barriers currently preventing them from choosing GCSE Computer Science. These barriers are reflected in the gender difference in the baseline rates of pupils choosing GCSE Computer Science with 20% of boys and 7% of girls intending to choose it as a GCSE option. Teachers suggested that they felt the intervention would have had a more meaningful effect, had other barriers (e.g. subject choice process or timing) not been in the way.

5.2 Recommendations

Recommendations to support implementation

The following steps could help to increase the feasibility of implementing the Peer Instruction intervention:

1. **Simplify the lesson resources such that lessons take less long to prepare**
   The MCQs could be included within the lesson power-points and the amount of printing and cutting that needs to be done for the first unit could be reduced..
2. **Emphasise the importance of the stages of the Peer Instruction MCQs throughout the intervention materials**
   Teachers could be reminded of all of the stages of a Peer Instruction MCQ each time a question is presented on the power-point, for example, through a diagram representing the stages. The importance of the pair and group discussion stage of the MCQs could be emphasised to a greater extent in the training (which may be easier to do in an in-person setting) and in lesson plans/unit overviews. It might also be helpful to further emphasise to teachers how MCQs within the Peer Instruction

approach differ (both in purpose and in delivery) from MCQs used for formative assessment.

These changes could make the Peer Instruction intervention easier for teachers to implement as intended. Whilst they may not fully tackle the barriers to girls selecting computer science as a GCSE subject, they could increase the potential for the Peer Instruction lessons to lead to greater collaboration within computing lessons which could, in turn, increase girls' intention to study computer science at GCSE.

Recommendations for future use of intervention resources

Despite the challenges described in section 5.1, given the positive response of teachers and pupils to the Peer Instruction lessons, we recommend that RPF:

1. **Share general guidance with teachers on how to apply the Peer Instruction approach to computing lessons (guidance applicable to all computing units - not only those units covered in the current Peer Instruction lesson plans)**
   RPF could publish online guidance on how to incorporate pre-instruction tasks and structured, collaborative MCQs into computing lessons. RPF could also build banks of MCQs which could be used with other units. Given that the teachers themselves reported that they had started trying to build banks of MCQs for other units, it would be helpful for RPF to develop more of these sets of questions, which could be used alongside pre-existing publicly available lesson plans, such as those on the NCCE website. Alongside this, RFP could publish guidance on how teachers could create their own MCQs, supported by evidence from other trials of the effectiveness of the Peer Instruction approach[29] [30].

Following the recommendations to support implementation a further recommendation would be to:

2. **Make available online the Peer Instruction lesson resources and training that were used in this trial**
   As the Python and Representations units cover content that many schools would normally be teaching in year 8, we do not foresee any significant opportunity costs to teachers using these lessons, as opposed to the lessons they would otherwise teach to cover the same content. If the training continues online, it could be valuable to emphasise the collaborative nature of Peer Instruction, to try to reduce variation in teachers' use of pupil discussion during the MCQs.

[29] Watkins, J., & Mazur, E. (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. *Journal of College Science Teaching*, *42*(5), 36-41.

[30] Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics,* 74(2),118–122. https://doi.org/10.1119/1.2162549

To increase the potential impact of interventions targeting gender gaps in computing, one possible avenue would be to:

3. **Consider implementing multiple strategies in parallel to address the barriers to girls' engagement with computing**
   Recognising the teachers' feeling that girls face a large number of significant and structural barriers to choosing computer science at GCSE, it's possible that any intervention should be conceptualised as a 'part of the puzzle': this intervention (or elements of it) could be implemented alongside other interventions designed to improve girls' engagement with computing. It might be that a combination of different efforts, targeting girls at a range of points in their school careers, is necessary to help girls overcome the barriers that they currently face.

Recommendations for future evaluations

Finally, possible strategies to address the evaluation challenges encountered could be to:

1. **Continue to refine survey tools and support schools to administer them to maximise data reliability and reduce attrition**
   The implementation and evaluation of the intervention examined in this report was particularly challenging given the COVID-19 context, in addition to the challenges often associated with evaluating school-based interventions, including attrition. While possible improvements in the COVID-19 context in schools should facilitate future evaluations, doing additional small-scale piloting of survey tools and identifying ways to support schools with data collection (e.g., appointing research staff to visit schools to help administer the survey), while resource-intensive, could be a cost-effective way to reduce attrition and increase data quality, thereby enabling a more precise diagnosis of the effects of the interventions and how to maximise them.

2. **Measure the outcomes targeted by the intervention further into the future**
   Tracking relevant behavioural outcomes (in this case, actual GCSE subject choice) after the end of the intervention would require planning, greater collaboration with schools and longer evaluation timelines. However, it would also enable closer measurement of the behavioural outcome targeted by this intervention (girls selecting computer science as a GCSE subject), in addition to the short-term proxy indicators used in this evaluation.

In light of the disruptions to the delivery of the intervention associated with the COVID-19 context and the positive experiences of the case study schools, there is reason to believe that implementing the intervention again after addressing the adjustments to its design and delivery suggested in the recommendations above could result in improved effectiveness. In addition, using school administrative data to measure whether girl pupils in the evaluation sample go on to select computer science as a GCSE subject would help to both reduce the need for primary data collection and increase the precision of the results in capturing any impact on the target behavioural outcomes. We thus recommend exploring the possibility of conducting another round of this intervention and evaluation if these suggested adaptations can be made, particularly if the cost of this new round of activities would be low.

References

BCS: The Chartered Institute for IT. (2021) "*Computing is the fastest growing STEM A level, says professional body for IT*." [Blog] Available at: https://www.bcs.org/articles-opinion-and-research/computing-is-the-fastest-growing-stem-a-level-says-professional-body-for-it/

Childs, K (2021) *Factors that impact gender balance in computing. In Understanding computing education* (Vol 1). Proceedings of the Raspberry Pi Foundation Research Seminar series.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics,* 69(9), 970-977.

Department for Education (2014) *The national curriculum in England: complete framework for key stages 1 to 4.* Available at: https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum

Goode, J. (2007). If you build teachers, will students come? The role of teachers in broadening computer science learning for urban youth. *Journal of Educational Computing Research*, 36(1), 65-88.

Goode, J., Estrella, R., & Margolis, J. (2006). *Lost in translation: gender and high school computer science*. Women and Information Technology: Research on Underrepresentation, eds JM Cohoon & W. Aspray.

Joint Council for Qualifications (2021) "GCSE (Full Course) Results Summer 2021 - Outcomes for key grades for UK, England, Northern Ireland & Wales, including UK age breakdowns". Available at: https://www.jcq.org.uk/wp-content/uploads/2021/08/GCSE-Full-Course-Results-Summer-2021.pdf

Kemp, P.E.J., Berry, M.G. & Wong, B. (2018). The Roehampton Annual Computing Education Report: Data from 2017. *London: University of Roehampton.* Available at: https://cdn.bcs.org/bcs-org-media/3972/tracer-2017.pdf

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics,* 74(2),118–122. https://doi.org/10.1119/1.2162549

Ofsted (2022) *Research review series: computing*. Available at https://www.gov.uk/government/publications/research-review-series-computing/research-review-series-computing

Porter, L., Bouvier, D., Cutts, Q., Grissom, S., Lee, C., McCartney, R., ... & Simon, B. (2016, February). A multi-institutional study of peer instruction in introductory computing. *In Proceedings of the 47th ACM Technical Symposium on Computing Science Education* (pp. 358-363).

Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science students and researchers*. Sage.

Royal Society. (2017) After the reboot: computing education in UK schools. https://royalsociety.org/~/media/policy/projects/computing-education/computing-education-report.pdf

Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys (Vol. 81). John Wiley & Sons.

Watkins, J., & Mazur, E. (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. *Journal of College Science Teaching*, *42*(5), 36-41.

Werner, L., & Denning, J. (2009). Pair programming in middle school: What does it look like? *Journal of Research on Technology in Education*, 42(1), 29–49. https://doi.org/10.1080/15391523.2009.10782540

Appendices

Appendix 1: Pupil baseline and endline survey measures

Hello! It's time to do the survey.
Please read each question carefully and take your time to answer.
Please don't worry about people you know seeing your answers - that won't happen.

1.1	Please type your first name	Text entry		
1.2	Please type your last name	Text entry		
1.3	Please select the gender you identify with	Female	Male	Non-binary/ Other
1.4	Please select the day you were born/month you were born/year you were born	Drag downs		
1.5	Please pick the name of your school from the list below	Drag down		
1.7	Do you want to study any of these subjects in future?			
	Computer Science	Yes	No	Don't know
	Science	Yes	No	Don't know
	Technology	Yes	No	Don't know
	Engineering	Yes	No	Don't know
	Maths	Yes	No	Don't know

Page 2
Thanks! Now it's time for the rest of the questions.
You can answer by selecting the button next to the answer you want to give.
[Not shown to students: **Subscales -** 1-5 Confidence, 6-10 Interest, 11-15 Belonging, 16-20 Usefulness, 21-25 Encouragement]

Please rate these sentences based on how much you agree with them.

2.1	I am confident that I can do computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.2	I am confident that I can solve problems by using computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.3	I can learn computing skills without much help	Strongly disagree	Disagree	Agree	Strongly Agree
2.4	I am good at solving hard questions in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.5	I think I will do well in computing	Strongly disagree	Disagree	Agree	Strongly Agree

Page 3
Please rate these sentences based on how much you agree with them.

2.6	I would choose more computing lessons if I could	Strongly disagree	Disagree	Agree	Strongly Agree
2.7	In the future I'd like to do more computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.8	I like to use computing to solve problems	Strongly disagree	Disagree	Agree	Strongly Agree
2.9	Solving questions in computing lessons makes me feel happy	Strongly disagree	Disagree	Agree	Strongly Agree
2.10	I like computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree

Page 4

Please rate these sentences based on how much you agree with them.

2.11	I feel happy in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.12	I feel like I belong in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.13	I have lots of friends in my computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.14	I know someone who uses computing in their job	Strongly disagree	Disagree	Agree	Strongly Agree
2.15	I have friends who think computing is interesting	Strongly disagree	Disagree	Agree	Strongly Agree

Page 5

Please rate these sentences based on how much you agree with them.

2.16	Knowing about computing will help me get a job	Strongly disagree	Disagree	Agree	Strongly Agree
2.17	To get the job I want I will need computing skills	Strongly disagree	Disagree	Agree	Strongly Agree
2.18	I can use things I learn in computing lessons in other lessons too	Strongly disagree	Disagree	Agree	Strongly Agree
2.19	I'll need to be good at computing skills for my lessons as I get older	Strongly disagree	Disagree	Agree	Strongly Agree
2.20	Computing is an important subject	Strongly disagree	Disagree	Agree	Strongly Agree

Page 6

How much do you agree or disagree with the following statements? There are no right or wrong answers.

2.21	A friend, or someone I know said I should do computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.22	Someone I know has made me feel interested in computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.23	Someone I know has said my work in computing is good	Strongly disagree	Disagree	Agree	Strongly Agree
2.24	I have been taught about how computing is used outside of lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.25	Someone in my family has made me want to learn computing	Strongly disagree	Disagree	Agree	Strongly Agree

Page 7 **(endline survey only)**

3.1	When you are solving problems in computing lessons, do you usually work on your own or in a group?	Always on my own	Mostly on my own	Not sure	Mostly with a group	Always with a group

Appendix 2: Model specification

Primary outcome: Intention to study computer science at GCSE level

The primary outcome is binary, and therefore we used a logistic regression to assess the Intention-To-Treat (ITT) effect of our treatment on this outcome. Owing to the clustered nature of the data, we used cluster-robust standard errors in analysis, clustering at the school level.

$$Y_{is} \overset{indep}{\sim} bernoulli(p_{is}); logit(p_{is}) = \alpha + \beta_1 T_s + \beta_2 BL_i + B_3 propFSM_s + B_4 Ofsted_s + \epsilon_{is}$$

Where

- Y_{is} is a binary indicator for pupil *i* reflecting intention to study computer science at GCSE level in school *s*
- p_{is} is the probability of a positive intention for pupil *i* in school *s*
- α is the constant
- T_s is a binary indicator of treatment assignment for pupil *i in school s*
- BL_i is the baseline SCSAS score for pupil *i* in school *s* collected before the intervention
- $propFSM_s$ is the proportion of pupils eligible for Free School Meals in school *s*
- $Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) “Outstanding”; (ii) “Good”; and (iii) “Below good” (the combination of “Requires improvement” and “Inadequate”)
- ϵ_{is} is the error term for pupil *i* in school *s*

Table 9 below provides the full results for the primary analysis using multiple imputation (column 1), missingness indicator (column 2) and complete case analysis (column 3).

Table 9: Logistic regression coefficients for primary outcome (standard errors in parentheses)

Outcome: Intention to study computing	*(1) MI*	*(2) Miss. Ind.*	*(3) CCA*
Treatment group (reference category is control)			
Intervention	0.035 (0.274)	0.001 (0.266)	-0.013 (0.260)
Baseline SCSAS score	2.561** (0.844)	3.189** (0.292)	3.190** (0.295)
Missing Baseline SCSAS	-		-
Ofsted rating *(reference category is Outstanding)*			
Good	-0.386 (0.378)	-0.599+ (0.308)	-0.416 (0.324)
Below Good	-1.333* (0.568)	-1.345** (0.404)	-1.787** (0.634)
Missing	-0.114 (0.487)	-0.298 (0.421)	-0.304 (0.441)
Percentage FSM	0.009 (0.014)	0.010 (0.013)	0.006 (0.014)
Proportion of girls			
Constant	-9.128** (2.169)	-10.745** (0.714)	-10.760** (0.718)
Control group mean	0.095	0.095	0.097
Observations	1,959	1,959	1,542
R^2	-	0.171	0.212

Note: Standard errors clustered at the school level
*+ p<0.1; * p<0.05; *** p<0.01

Secondary outcome: stated intention to study computing

The secondary outcome is continuous, and therefore we used a linear regression to assess the Intention-To-Treat (ITT) effect of our treatment on this outcome. Owing to the clustered nature of the data, we used cluster-robust standard errors in analysis, clustering at the school level.

$$Y_{is} = \alpha + \beta_1 T_s + \beta_2 BL_i + B_3 propFSM_s + B_4 Ofsted_s + \epsilon_{is}$$

Where:

- Y_{is} is the Total SCSAS survey mean score for pupil *i* in school *s*
- α is the constant
- T_s is a binary indicator of treatment assignment for pupil *i in school s*
- BL_i is the baseline SCSAS score for pupil *i* in school *s* collected before the intervention

- $propFSM_s$ is the proportion of pupils eligible for Free School Meals in school *s*
- $Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) “Outstanding”; (ii) “Good”; and (iii) “Below good” (the combination of “Requires improvement” and “Inadequate”)
- ϵ_{is} is the error term for pupil *i* in school *s*

Table 10 provides the full results for the secondary analysis using multiple imputation (column 1), missingness indicator (column 2) and complete case analysis (column 3)

Table 10: OLS regression coefficients for secondary outcome (standard errors in parentheses)

Outcome: SCSAS score	*(1) MI*	*(2) Miss. Ind.*	*(3) CCA*
Treatment group (reference category is control)			
Intervention	-0.044 (0.050)	-0.053 (0.048)	-0.057 (0.043)
Baseline SCSAS score	0.741** (0.133)	0.839** (0.028)	0.840** (0.028)
Missing Baseline SCSAS	-	2.039** (0.065)	-
Ofsted rating *(reference category is Outstanding)*			
Good	0.021 (0.058)	-0.006 (0.045)	0.003 (0.036)
Below Good	0.025 (0.089)	0.056 (0.100)	-0.058 (0.031)
Missing	-0.016 (0.069)	-0.044 (0.051)	-0.038 (0.053)
Percentage FSM	-0.003 (0.003)	-0.003 (0.003)	-0.002 (0.002)
Constant	0.610+ (0.309)	0.403** (0.112)	0.388** (0.104)
Control group mean	2.45	2.45	2.47
Observations	1,914	1,914	1,510
R^2	-	0.382	0.484

Note: Standard errors clustered at the school level
*+ p<0.1; * p<0.05; *** p<0.01

Appendix 3: Secondary Training survey

Gender Balance in Computing Teaching Approach (i1a Stage 2) Secondary training survey

Section 1: About you

What is your full name? [Free text answer]

Please select the name of your school from the menu [Menu of all intervention schools]

How confident are you in teaching computing to Key Stage 3 pupils?

◯ Not at all confident
◯ Not very confident
◯ Quite confident
◯ Very confident

Please confirm that you have completed the online training for the Peer Instruction Teaching Approach project

▢ Yes, I have completed the training
▢ No, I have not completed the training

Section 2: About the training

Roughly how long did the training take you to complete?

◯ Less than 1 hour
◯ 1-2 hours
◯ 2-3 hours
◯ 3-4 hours
◯ 4-5 hours
◯ More than 5 hours

Did you watch the webinar recording at the beginning of the training?

◯ Yes

◯ No

If you watched the webinar, how useful did you find it in preparing you for the project?

◯ Not at all useful

◯ Not very useful

◯ Quite useful

◯ Very useful

How useful did you find the online training in familiarising you with the resources for the project?

◯ Not at all useful

◯ Not very useful

◯ Quite useful

◯ Very useful

How confident do you feel in delivering the resources for the project after the online training?

◯ Not at all confident

◯ Not very confident

◯ Quite confident

◯ Very confident