Gender Balance in Computing

Evaluation of Teaching Approach: Pair Programming intervention

May 2022

Stefan Kelly, Julia Ryle-Hodges and Clément Bisserbe

THE BEHAVIOURAL INSIGHTS TEAM

Contents

Executive summary

Overview of the project

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 Pair Programming approach is a 12-week intervention which covers a term of computing lessons. Within these lessons, key stage 2 (KS2) pupils are directed to work in pairs, with one pupil taking on the role of 'driver' (who is responsible for the mouse and keyboard) and the other taking on the role of 'navigator' (who is responsible for thinking about solutions to problems and instructing the driver).

Evaluation approach

The intervention was evaluated using a mixed methods approach. The impact evaluation investigated whether there was evidence that the intervention impacted (i) girls' attitudes towards computing and (ii) girls' stated intention to study computing in the future. 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.

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 term (September - December 2021).
2. *Treatment*: Schools in this group received training materials and class plans for a 12-week term of computing lessons to deliver the Pair Programming intervention (September - December 2021).

The primary outcome was measured using the Student Computer Science Attitude Survey (SCSAS), a survey tool for assessing attitudes toward computing for school pupils. The secondary outcome was intention to study computing in the future, measured using a self-report survey question. In total, 116 schools were originally recruited to participate in the trial and 56 completed it (submitted endline survey data).

Implementation and process evaluation

Alongside the impact evaluation, a mixed-methods IPE was conducted 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 are the mechanisms through which the intervention was perceived to bring about change in the anticipated outcome?

BIT researchers spoke with teachers at four schools where the intervention was delivered to understand teachers' experiences of delivering the intervention, alongside the feasibility of delivery, any programme adaptations, and the perceived effects for girls. In two of these schools, we also conducted small group discussions with pupils and lesson observations, to better understand how pupils experienced the intervention and how the intervention was being implemented in practice. It should be noted that the small number of teachers who shared their feedback limits the breadth of experience which can be reported in the IPE. We also explored the responses to a teacher survey giving feedback about the online training.

Key findings

Evidence of impact

Overall we found no conclusive evidence that the intervention increased girls' attitudes towards computing or intention to study computing in the future. While the estimated effect of the intervention on each outcome was positive, it was not statistically significant and small in magnitude. Issues related to the limited sample size, differential attrition and baseline imbalances in both outcome measures introduce some risk of bias in the results, which should thus be interpreted cautiously. However, given the small size of the differences between the two groups and that the potential bias caused by attrition is more likely to result in an over- rather than under-estimation of the impact of the intervention, there is no evidence indicating that the low sample size or potential biases 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 four case study schools:

- **Fidelity**: The COVID-19 context made this a difficult time for schools to implement a 12-week intervention. The Scratch projects and paired work elements of the programme were implemented as intended but teachers cut elements at the start and end of the lesson plan to fit the lesson into their timetabled slot.

- **Feasibility:** Characteristics of the learning environment (such as pupil familiarity with routines and pupil baseline understanding of computing) and teacher characteristics (such as confidence and subject knowledge) affected the ease with which teachers could implement the intervention..

- **Responsiveness:** Pupils took on their roles of 'driver' and 'navigator', fulfilling the responsibilities of the respective roles and communicating about the tasks set.

- **Quality:** Teachers were generally happy with the quality of the programme resources: lesson plans, power-points, half-built Scratch projects (for pupils to complete), unit overviews and online training. Some found these resources time consuming to work through.

- **Mechanisms:** Pupils and teachers felt that the programme facilitated pupils helping each other, which led to greater confidence. They also reported that the programme led to greater collaboration and increased girls' perception that computing is a collaborative subject. Teachers varied in whether they felt the programme specifically addressed girls' barriers to computing engagement, or whether it was equally positive for boys and girls. The school context and pupils' perceptions of the gendered nature of computing could be moderators of the effects of the intervention.

Possible explanations for the lack of evidence of impact despite the case study schools' positive experiences

Despite some teachers reporting that they felt girls' attitudes towards computing improved as a result of the collaboration in pair programming, no evidence of this was found in the impact evaluation based on the pupil surveys.

There are a number of possible reasons for the lack of evidence of impact of the intervention. These reasons are related to the design of the intervention, its delivery, and measurement challenges:

- The COVID-19 challenges schools were facing may have resulted in a lower dosage of the intervention than necessary to have an impact.
- Due to COVID-19 constraints, the training was delivered online (not in person as originally intended) which limited the scope of what could be delivered. This might have led to teachers implementing the 'paired' element of the lessons less effectively than if the training had been received in person.
- Given the outcome measurement challenges inherent to the nature and objectives of the intervention, and the related need to rely on short-term and pupil-reported proxy indicators, the outcome measures may not have been sensitive enough to capture differences in pupils' attitudes and intentions towards computing.
- Due to the characteristics of the sample (schools which contain a teacher interested in taking part in a gender balance in computing project), the sample could have had particularly high baseline engagement with computing, leaving less scope for improvement as a result of the intervention and making it more difficult for the evaluation to detect an impact on these outcomes. The high baseline SCSAS score and stated intention to study computing are consistent with this hypothesis.
- Addressing the hypothesised barrier of computing being taught with an 'individualistic' as opposed to collaborative approach may not be sufficient to change girls' attitudes towards computing.
- It is possible that the intervention may set off a chain of mechanisms which eventually contribute to intended outcomes, but that the changes in these outcomes were not yet apparent when measured immediately after the intervention.

There is not sufficient evidence to confidently determine whether and how these factors may have contributed to the evaluation results.

Recommendations

We make the following recommendations based on the evaluation findings:

1. **Cut down content within each lesson plan**
   The starter and end-of lesson activities which teachers did not always have time to complete could either be removed, or it could be made clear in lesson plans that it is acceptable for teachers to remove content from the plans themselves. The paper-based starters could be replaced by similar activities which can be done as a whole class through the power-point.

2. **Provide additional support for teachers who lack confidence teaching these National Curriculum objectives**
   Training videos could be made to demonstrate how to create the code to complete certain projects, or how to deal with frequently faced challenges. In future, face-to-face training might provide opportunities for support to be more personalised based on teacher confidence.

3. **Offer the lesson resources and training to KS2 teachers**
   Interviewed teachers reported that delivering the Pair Programming lessons had been a positive experience for both themselves and their pupils. It's also likely that the quality of these lesson resources might be higher than computing resources non-specialist teachers can be expected to develop from scratch.

4. **Make non topic-specific Pair Programming resources available to teachers to use outside of these units of lessons**
   Teachers reported that they enjoyed using the paired approach (assigning pupils to pilot and navigator roles) and were keen to try it with other groups of pupils, outside of the units of work given for this trial. RPF could make general Pair Programming resources (non specific to topics within computing) available to teachers, who could then integrate the approach into other lesson plans.

5. **Continue to refine survey tools and support schools to administer them to maximise data reliability and reduce attrition**
   Additional small-scale piloting of survey tools and identifying ways to support schools with data collection (e.g., appointing 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.

6. **Identify strategies to measure outcomes targeted by the intervention further into the future**
   Tracking subject choice multiple years after the intervention, though challenging, would greatly enhance our ability to evaluate the impact of early interventions over a time horizon in line with the mechanisms and barriers hypothesised.

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 Computing 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, though this would be easier to achieve for interventions targeting older pupils closer to their GCSE subject selection. 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[1]. Across England, only 21% of the GCSE Computer Science cohort is made up of female students[2]: many girls are not choosing to continue with computing subjects at the point at which lessons become optional, usually at the start of year 10[3] [4]. The Gender Balance in Computing Project (GBIC) aims to tackle a number of known and well-researched barriers to girls engaging with computing[5], 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 student 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 the Teaching Approach intervention, 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' perceptions of computing.

1.2 Teaching Approach programme

The premise of the 'Teaching Approaches' 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[6] [7]. In particular, the way in which computing is taught may lead to girls perceiving computing as an individualistic rather than a collaborative subject. The intervention evaluated aims to change girls' perceptions of computing by using teaching approaches that encourage collaboration and discussion within computing lessons. This approach is grounded in

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

[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

[3] 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

[4] 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..

[5] 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

[6] 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.

[7] 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.

evidence that incorporating collaborative teaching approaches into STEM subjects improves girls' self-efficacy and achievement.[8],[9]

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 Pair Programming intervention; the evaluation of the Peer Instruction intervention is covered in a separate report.

The Pair Programming intervention

The intervention, developed and rolled out by RPF, is a 12-week programme that incorporates the Pair Programming teaching approach into computing lessons for year 4 or year 6 pupils. Pair Programming is an approach where two pupils work together to write a program or solve a problem while sharing a single computer. This approach has been found to support females' problem solving confidence and persistence across different age groups[10] [11]. In pairs, pupils take turns to be the driver and the navigator. The driver controls the keyboard and mouse and types in the code. The navigator reads the instructions, supports the driver by watching out for errors in the code and thinks strategically about next steps and solutions to problems. The approach is collaborative and discursive.

The intervention aims to encourage interactive engagement in learning and enhance cooperative problem solving and communication between pupils. It includes one introduction to programming lesson, five lessons on programming drawings and six lessons on programming animation.

The logic model (see Figure 1) was developed through discussion between the RPF team and the BIT evaluators. It illustrates the hypothesised mechanisms through which the Pair Programming intervention would affect the intended outcomes of girls' attitudes towards computing and intention to study computing. The key barrier that the intervention was designed to address is a lack of collaboration in computing lessons creating a tension between computing lessons and girls' preferred ways of learning. By making computing lessons more collaborative, this intervention aimed to address this barrier.

Teachers undertook an online training course, which they worked through at their own pace. They were also given the option of attending a webinar or watching a recording of the webinar, if they couldn't attend the live event. During the online training, teachers were directed to the online portals containing all of the resources they needed to deliver the units, and they completed for themselves some of the computing activities that the pupils would be asked to complete during lessons. The original plan had been for teachers to attend face-to-face training in which they would have seen pair programming modelled so that the

[8] 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
[9] 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
[10] McDowell, C., Werner, L., Bullock, H. E., & Fernald, J. (2006). Pair programming improves student retention, confidence, and program quality. Communications of the ACM, 49(8), 90-95.
[11] 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

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.

Schools were able to decide whether to deliver the intervention to their year 4 cohort or year 6 cohort, depending on their pupils' familiarity with Scratch, the programme which the intervention used. Scratch is an online or downloadable coding programme with a simple visual interface which allows pupils to drag and drop 'blocks' of code to programme 'sprites' (or characters).

Teachers were given online access to all of the resources they would need to teach the 12-week unit of lessons. These resources included: a unit overview; individual lesson plans; a power-point presentation for each lesson; some printable resources; teacher demonstration Scratch projects; and Scratch projects for pupils to continue. The lessons were taught over the Autumn term of 2021 (September - December).

In advance of the start of this trial, RPF conducted a pilot of the Pair Programming intervention in January-March 2020 with 27 teachers. Teachers received face-to-face training and gave feedback on both the training and their experience of delivering the Pair Programming lessons. The intervention was adapted based on their feedback before the start of this trial. 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.

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.

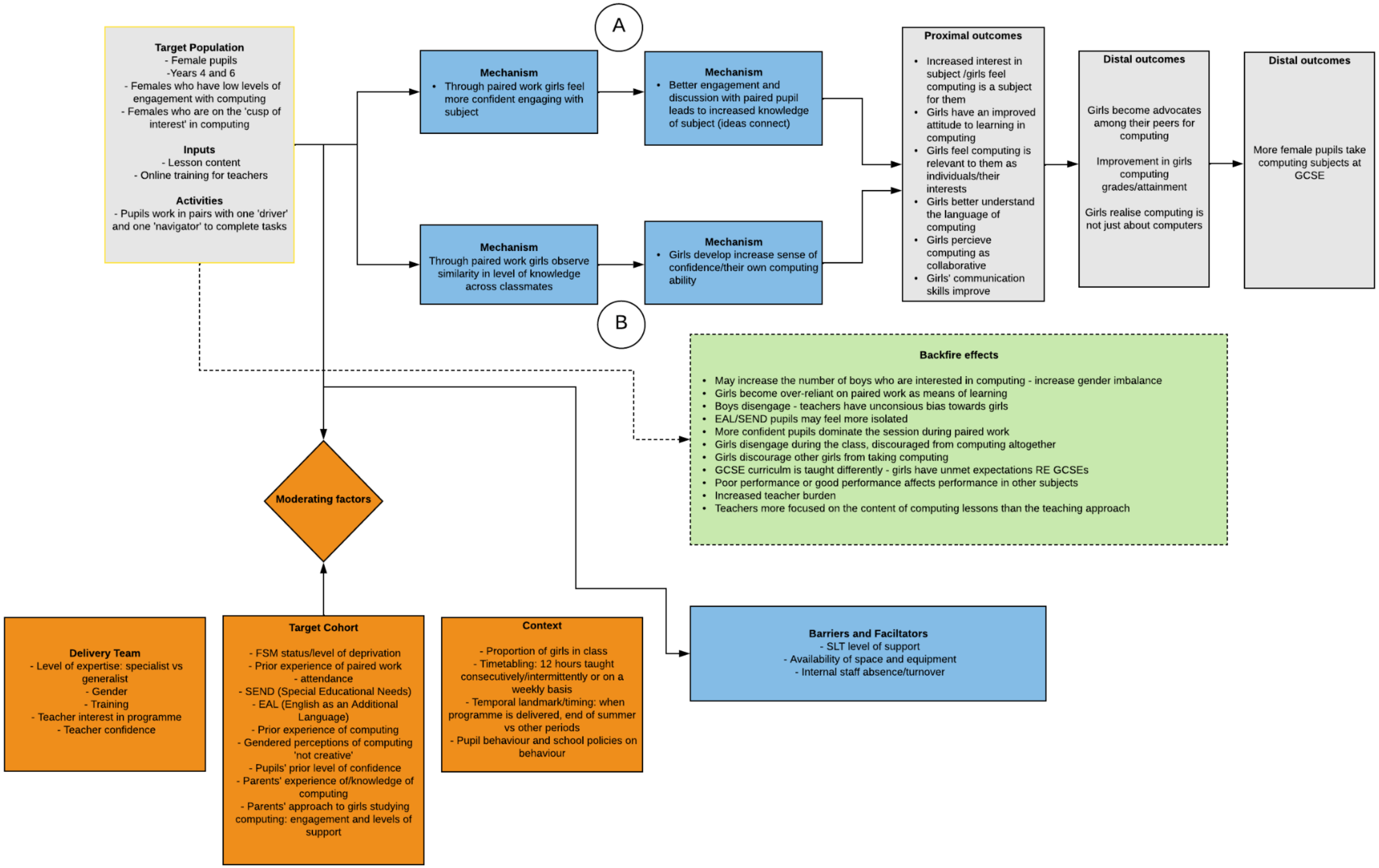

Figure 1: Logic model of the Pair Programming intervention

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 online surveys distributed pre and post intervention in both treatment and control schools to be completed as part of their computing lessons. A mixed-methods 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 describes 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. Girl pupils' attitudes towards computing as measured by the Student Computer Science Attitude Survey (SCSAS)
2. Girl pupils' stated intention to study computing in the future

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

Table 1: Method for collecting quantitative outcome data

Outcome measures	Data to be collected	Point of collection
Primary: General attitudes towards computing	Overall score on the Student Computer Science Attitudes Survey (All 5 constructs equally weighted: Confidence, Interest, Belonging, Usefulness, Encouragement).	- Online surveys, completed on computers in class at baseline (beginning September 2021) & immediately following the culmination of the 12-week programme (beginning December 2021).
Secondary: Intention to study Computing	Single item survey measure of whether the pupil plans to continue studying computing with possible responses "Yes", "No", or "I don't know".	

The SCSAS has been developed to measure attitudes towards computing[12] (see Appendix 1 for the full survey content and adaptations that took place between baseline and endline data collection. It contains 25 questions and has 5 subcategories (5 questions per subcategory): confidence, interest, belonging, usefulness and encouragement. Within each subcategory,

[12] Haynie, K.C. and Packman, S. (2017). *AP CS Principles Phase II: Broadening Participation in Computer Science Final Evaluation Report*. Prepared for The College Board and the National Science Foundation, February 12, 2017. Skillman, NJ.

the 5 items are scored on a four-point Likert scale from 1 (strongly disagree) to 4 (strongly agree), and averaged to create subscores. Thus, each 5-item subscore has a potential range of 1-4. These subscores are averaged for a total score that has a potential range of 1-4, with 4 representing a very positive attitude towards computing. For the secondary outcome measures, pupils self-reported their intention to study computing. This was converted to a binary outcome measure with 1 indicating they had answered “Yes” to whether they intended to study computing and 0 indicating they had answered “No” or “I don’t know”. On behalf of the BIT evaluation team, RPF included the links to the online surveys in the first and last lesson plans of the unit.

2.1.2 Sampling and randomisation

The evaluation was designed as a two-armed cluster randomised controlled trial that aimed to test the impact of the Pair Programming intervention, 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 term in Autumn 2021.
2. Treatment: Schools in this group received training materials and class plans for a 12-week term of computing lessons to deliver the Pair Programming intervention as outlined in section 1.2.

Schools were stratified on the percentage of pupils with free school meal (FSM) status (above or below the median). 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’) and proportion of pupils who are girls. We used school unique reference numbers (URNs) as unique identifiers. BIT conducted the randomisation.

Pupils were blind to allocation during the programme and during outcome data collection, and thus did not know that pupils at other schools received different classes. 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 may have been aware of a control group.

Recruitment of schools was conducted by RPF. All primary schools in England were eligible for this trial. A small number of schools contained pupils taking part in both the Pair Programming (year 4 or year 6) intervention and the Informal Learning: Code Club intervention (years 3-6). This was deemed an acceptable overlap because the extra-curricular Code Clubs generally contained small numbers from each year group therefore any potential contamination (pupils who are both in the Pair Programming trial and the Code Clubs trial) would be small. Only the schools that could offer the full 12-week programme were able to enter 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 this may interact with the treatment effect to compound any effects. However, this is less of a concern if the population schools that this programme may potentially be rolled out to in future also fall into this category.

Data was collected for both boys and girls, but only data from girls was analysed for primary and secondary analyses[13].

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.

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

Outcome	Values	Gender	N (non-missing)	Mean (SD)
Total SCSAS score	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	1,399	2.73 (0.49)
		Boys	1,247	2.83 (0.53)
SCSAS: Confidence subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	1,399	2.82 (0.59)
		Boys	1,247	2.93 (0.64)
SCSAS: Interest subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	1,399	2.69 (0.61)
		Boys	1,247	2.80 (0.66)
SCSAS: Belonging subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	1,363	2.90 (0.58)
		Boys	1,216	2.98 (0.64)
SCSAS: Usefulness subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	1,363	2.81 (0.64)
		Boys	1,216	2.93 (0.69)
SCSAS: Encouragement subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	1,363	2.42 (0.68)
		Boys	1,216	2.52 (0.76)
Intention to study computing in future	1 = “Yes” 0 = “No”, “Don’t know”	Girls	1,416	0.55 (0.49)
		Boys	1,264	0.66 (0.47)

Table 2 shows that at baseline, boys scored higher than girls in attitudes toward computing (Boys mean = 2.83 out of 4, Girls mean = 2.73 out of 4), as well as intention to study computing in future (66% of boys indicating they would like to study computing as a subject in future, compared to 55% of girls). This gender difference is in line with the core issue this intervention is designed to address, and is further evidence of a need to test approaches that may increase girls’ attitudes towards computing.

[13] Boys’ data was also checked for potential backfire effects of the intervention.

The SCSAS can be broken down into 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, and the individual questions that make up these subscales can be found in Appendix 1. Table 2 (above) shows the gender difference in attitudes is consistent across the subscales, meaning no individual subscale seems to be the root of the difference in total SCSAS score between boys and girls.

2.1.4 Attrition and final sample

Figure 2 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, RPFattempted 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, high attrition rates were observed between randomisation and completion of baseline surveys, with proportionally more schools dropping out from the control group than the intervention group (50% vs 29%). At least part of this attrition is likely due to disruption caused by the COVID-19 pandemic. In late summer and autumn 2021, a wave of infections likely led to many teachers and pupils self-isolating at home: schools faced serious challenges staying open with the number of staff and students absent and may not have been able to prioritise taking part in research. However, the differential attrition between experimental groups suggests a possible risk of bias in the analysis results. 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 and pupils within schools not completing the endline. The final sample used for analysis also included two schools who completed the endline survey, but did not complete the baseline survey. We chose to keep these schools in the sample, as they appear likely to have completed the intervention - RPF confirmed the teachers were recorded as having completed the online training materials given, and pupil responses were in line with the compliance measure which took up part of the endline survey.

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 99 pupils from the total baseline sample of 3,290 (boys and girls, not necessarily matched to any endline observations) and 13 pupils from the endline sample (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 995 girls - 617 in the treatment and 378 in the control group.

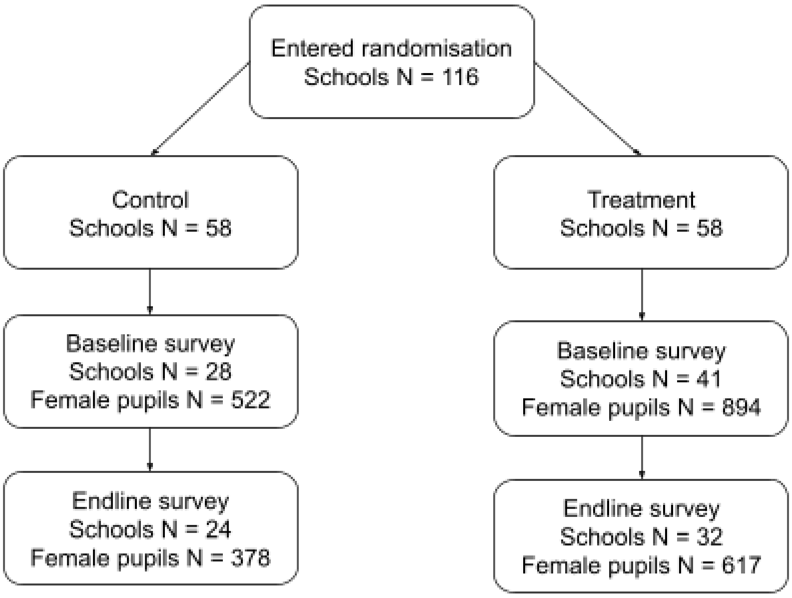

Figure 2: School level attrition

Baseline differences in outcome measures

Table 3 shows that at the point of baseline data collection, the groups were unbalanced in terms of attitudes towards computing and intention to study computing in the future. The difference in SCSAS scores is statistically significant and meaningful in size relative to the differences observed at endline, thus suggesting a risk of bias to the results. We hypothesise that this bias is a result of differential attrition between groups between recruitment and the baseline survey, rather than randomisation failure, as balance checks were conducted on school-level variables post-randomisation. Baseline scores for boys who completed the endline survey are not shown as the final matched sample included girls only.

Baseline data collection was extended such that some schools completed the baseline survey after the intervention had begun, which may have moved the baseline outcomes especially for schools in the treatment group from their ‘true’ baseline pre-intervention. Of the pupils in the final treatment group sample, 32.6% completed the baseline survey over two weeks after the earliest school had begun to submit responses (vs 38.9% of the control group), and 20.1% of the treatment group completed the baseline four weeks or longer after data collection had begun (vs 29.4% of the control group). While those who completed the survey late across both groups could potentially see increased scores in outcome measures due to having received more computing lessons, this may have interacted with a treatment effect leading to this disproportionately affecting those in the intervention group. Girls in the intervention group who responded to the baseline survey two weeks late or longer scored higher in the SCSAS at baseline than control (SCSAS for those completed two weeks later = 2.88, Less than two weeks = 2.76, $p<.05$). No statistically significant difference was seen for the same comparison in the control group. This may partially explain the baseline difference between treatment groups in both primary and secondary outcome measures.

Table 3: Balance checks for all baseline data

Covariate	Percentage (or mean) per arn		p-value	Balanced?
	Control (n = 977)	Intervention (n = 1,780)		
Gender				
Boys	43.6%	47.1%	>0.10	Yes
Girls	53.4%	50.2%		
Non-binary/Other	3.0%	2.7%		
SCSAS				
Baseline SCSAS score (full sample)	2.73	2.90	<0.01	No
Baseline SCSAS score (girls only)	2.69	2.75	<0.05	No
Intention to study computing				
Baseline intention (full sample)	54.2%	63.3%	<0.01	No
Baseline intention (girls only)	49.0%	58.1%	<0.01	No

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 unbalanced in terms of attitudes towards computing and intention to study computing in the future.

Table 4. Balance checks for baseline data of pupils that completed the endline survey

Covariate	Percentage (or mean) per arm		p-value	Balanced?
	Control (n = 378)	Intervention (n = 617)		
SCSAS				
Baseline SCSAS score (girls only)	2.65	2.79	<0.01	No
Intention to study computing				
Baseline intention (girls only)	43.9%	59.8%	<0.01	No

Implications for final analysis

The baseline imbalances and high attrition rates described above carry implications for our analysis.

Firstly, the differential attrition across groups, with proportionally more control schools dropping out of the trial post-randomisation, has the potential to bias our estimates of the impact of this intervention. The direction of this bias will depend on how the schools that dropped out of the control group differed from those that remained:

1. If it were the case that those that dropped out were less enthusiastic about teaching computing at baseline (and thus likely to score lower on attitudes and intentions surrounding studying computing on a pupil-level), this might bias our estimates downwards, as a higher proportion of the control group would be more enthusiastic about computing relative to the treatment group.
2. If it were schools that were higher in factors positively correlated with our outcome measures that dropped out of the control group (e.g. it is possible they would have been more disappointed by not being in the treatment group), it would bias our estimate upwards.

The baseline data suggests that the second explanation is possible, with the treatment group scoring higher at baseline at a statistically significant level in both outcome measures.This has important 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 tutors 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 the impact 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.[14]

Given that the endline data would likely include some pupils who did not complete the baseline dataset, 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 20% of pupils were missing baseline SCSAS scores (above the threshold of 5% for listwise deletion), and multiple imputation[15] was performed, whereby predicted values were substituted where data was missing.

[14] While it would have been possible to perform multiple imputation on missing Ofsted data, this was judged to be inadvisable as not all independent primary 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.
[15] Rubin (2004) *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

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
2. Missingness indicator model[16]
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

Differential attrition across experimental groups can lead to bias in treatment effect estimation. While baseline imbalance in outcome measures between groups can be partially addressed through using baseline SCSAS as a covariate in the analysis, we cannot be confident that there are not unobserved variables driving that baseline difference that we do not control for in analysis. The consequence of this would be that these unobserved variables potentially interact with attitudes and intentions, and could bias the results. The implications for the analysis of the differential attrition observed are described in section 2.1.4.

Another 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. 116 schools were recruited with a 40% attrition buffer to detect an estimated minimum detectable effect size (MDES) of .05 along the 1-4 SCSAS scale, and a 10 percentage point increase in intention to study computing. This meant approximately 35 schools per arm were required to reach this target MDES, which we were short of with 28 schools per arm in the final sample. Overall, this means the final analytical sample was not sufficient to detect an effect at the originally targeted effect size.

Pupil survey outcome measures

Given the nature and objectives of the intervention, defining and measuring outcome indicators were challenges inherent to the evaluation. The intervention aims to reduce gender gaps in school subject choices from year 10 onwards by intervening in earlier years, in year 4 and 6. While this early intervention approach may offer important benefits in terms of reducing barriers that may arise or increase in later years of education, it also creates a need to rely on short-term 'proxy' indicators that can be measured within the evaluation period (in this case directly after the intervention completion), yet could predict school subject choices in year 10. This is particularly challenging as some of the barriers to girls choosing computing as a subject in later years (and that the intervention aims to prevent) may arise

[16] 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.

after year 4 or 6 and before year 10; this would imply a risk that the effect of these barriers are not captured in the data collected while the pupils are in year 4 or 6.

Additionally, the absence of reliable observable proxy indicators requires relying on pupil self-reported data, which may introduce biases related to social desirability bias or limited respondent attention. This risk is particularly high for the indicator capturing self-reported intention to continue studying computing. Given that the year groups this intervention was aimed at are both at primary level, these pupils do not face a choice over studying computing in the near future. This could introduce some measurement error as pupils may select 'Yes' because they know that they will be continuing to study computing by default, resulting in baseline rates of intention much higher than what that rate would be if the pupils in the sample actually did face a choice over studying computing.

To address this dual challenge, the evaluation approach focused on attitudes towards computing as the primary outcome, and hypothesised that these could be measured and predict future subject choice. The survey tool used, the SCSAS, was cognitively tested to increase its reliability in measuring these attitudes with a small group of KS2 pupils from schools outside of this intervention. While these efforts should help, they are unlikely to fully address these challenges, and some issues were raised in testing in terms of different children interpreting some of the questions in different ways.

Some of the feedback shared by teachers during interviews suggest that these challenges may have been present. For the survey questions intended to measure attitudes, some teachers reported that some of their pupils had struggled to understand all of the survey questions, and that some pupils might have been selecting answers without necessarily having a clear sense of what the question was asking them; they flagged this risk for pupils who were lower attaining in English in particular.

The possible implications of these measurement challenges for the results are discussed in section 5.1.

2.2 Implementation and process evaluation

Alongside the impact evaluation, a qualitative 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 mixed methods design was used as this allowed us to explore the context of each school and the experiences of the teachers and pupils within them alongside a broader view of school staffs' experience of the online training. We planned and implemented a case study design, conducting qualitative research with teachers and pupils from the same school, where possible. We also report on the quantitative findings from the feedback survey on the online training.

We envisaged conducting research activities in four case study schools, however, due to the challenging COVID-19 context schools were facing, in two of our 'case study' schools, we were only able to conduct teacher interviews. Throughout this section, we note how the planned activities were adapted in response to COVID-19 restrictions.

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 can 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). Secondary criteria included the school's Ofsted rating. Schools' proportion of FSM eligible pupils was retrieved from the Department for Educational national information about schools[17]. We recruited two schools in the North of England, one in the

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

Midlands and one in the South of England. We were also able to achieve diversity in the proportion of FSM eligible pupils: two schools had above average FSM and two schools were below average.

Staff sampling

For teachers within the case study schools, the primary criterion was gender and secondary criteria included i) teaching experience and ii) teaching role. Teaching role was introduced as some of the teachers delivering the intervention taught primarily computing, across different year groups, whilst some teachers had their own class which they taught all subjects to, including computing. We recruited four teachers: one at each of the Case Study schools.

Pupil sampling

We aimed to recruit a sample of five key stage 2 pupils from each of the four case study schools to take part in pupil focus groups. We aimed to include at least two boys in each of the focus groups to deepen our understanding of any potential backfire effects of the intervention, as well as similarities and differences in their experience of the intervention compared to girls. We were only able to recruit pupils at two of our case study schools, as the other two case study schools did not have capacity to facilitate pupil focus groups as well as a teacher interview. In total, we spoke to 11 year 4 pupils, across two schools.

For pupils, primary criteria included gender and the pupil's confidence with computing, as advised by the teacher prior to inviting pupils to take part in the group discussion. The sampling of pupils occurred following a discussion with the teacher who was able to identify students with different levels of computing confidence 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, alongside making adaptations to facilitate COVID-19 safety precautions. This made it challenging for teachers to facilitate additional research activities (such as pupil focus groups) 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.

Table 5: Achieved case study sample

School	Profile	Teacher	Pupils	Data collection
S01	-Located in the North East -Above average FSM -LA maintained school - Ofsted rating: Good	-Computing specialist teacher -More than 10 years of experience -Male	-6 year 4 pupils (8-9 years old) -4 female (range of confidence in computing), 2 male (range of confidence in computing)	- Teacher interview - Lesson observation - Pupil focus group discussion - All activities were in-person
S02	-Located in the Midlands -LA maintained school -Below average FSM -Ofsted rating: Good	-Computing specialist teacher - More than 10 years of experience -Female	-5 year 4 pupils (8-9 years old) -3 female (range of confidence in computing), 2 male (range of confidence in computing)	- Teacher interview - Lesson observation - Pupil focus group discussion - All activities were in-person
S03	-Located in the North West -LA maintained school -Below average FSM -Ofsted rating: Outstanding	-Classroom teacher (with curriculum responsibilities for computing) -Fewer than 10 years of experience -Female	None	-Teacher interview -Conducted virtually
S04	-Located in London -LA maintained school -Above average FSM -Ofsted rating: Outstanding	-Classroom teacher (with curriculum responsibilities for computing) -Fewer than 10 years of experience -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 Pair Programming 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 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 online 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 incurred by teachers completing the training.

2.2.4 Data collection methods

Staff interviews

Individual, in depth semi-structured interviews were conducted with a teacher who delivered the Pair Programming lessons in each of the case study schools to explore their experiences of the intervention and any factors that influenced their ability to implement the intervention with their pupils. We were able to conduct 30-45 minute interviews with one teacher from each of our case study schools.

Pupil focus groups

Group discussions were held with pupil focus groups at two of the case study schools. These lasted around 20 minutes. 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 ‘finish the sentence’ on a range of sentence starters relating to the Pair Programming lessons and computing lessons more broadly. These activities were designed to create the opportunity for pupils to share their current perceptions of computing, as well as their experience of the Pair Programming 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. In the two schools in which we conducted lesson observations, these took place just before the pupil focus groups, such that in the discussions we could refer back to the Pair Programming lesson.

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) (n=66). 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[18]. 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 focus on providing rich descriptions of participant experiences, whilst looking for explanation and linkages within and across participant groups.

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

---

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

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.
2. The teachers who responded to our invitation to take part in the evaluation might have been the 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. Due to the challenges surrounding COVID-19, we were unable to conduct lesson observations and pupil focus groups at each of our case study schools. Whilst we were able to speak to pupils of different genders and different levels of confidence with computing, the total number of pupils and teachers we spoke to was small, which means we may not have captured the full range of pupil and teacher experience.

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.

3. Impact evaluation findings

3.1 Primary analysis: effect of the intervention on attitudes towards computing

Key findings:

- There was no conclusive evidence of an impact on girls' attitudes towards computing or reported intention to study computing in the future in the intervention group relative to the control group. While the estimated effect of the intervention on each outcome was positive, it was not statistically significant and small in magnitude.
- Differential attrition and baseline imbalances in both outcome measures introduce some risk of bias in the results, which should thus be interpreted cautiously. However, given the small size of the differences between the two groups and the more likely direction of the bias, there is no evidence indicating that these potential biases are obscuring a substantive positive impact of the intervention.

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' 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 2.875 (SD=0.523). For the intervention group it was 2.921 (SD=0.518) and for the control group it was 2.799 (SD=0.524). Compared to the control group, the pre-specified multiple imputation model found a difference of 0.046 points (p=0.331) on a 1-4 scale for the intervention, which is not statistically significant at conventional significance levels. While the difference is positive, it is small in magnitude relative to the 1-4 scale, and thus even if it was statistically significant, may not represent a substantive shift in girls' attitudes towards computing. This finding was consistent across the missingness indicator and complete case analysis model specifications. The baseline SCSAS score was an important predictor of endline SCSAS score across all model specifications.

Unlike other trials that fall 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, we did not conduct analysis relating to these subscales.

Table 6: Impact evaluation results for primary outcome

Outcome: Total SCSAS Score	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group unadjusted mean	2.80		2.77
Treatment group unadjusted mean	2.92		2.94
Estimated treatment effect (standard error)	0.046 (0.047)	0.030 (0.045)	0.059 (0.046)
N	990	990	785

Figure 3 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.

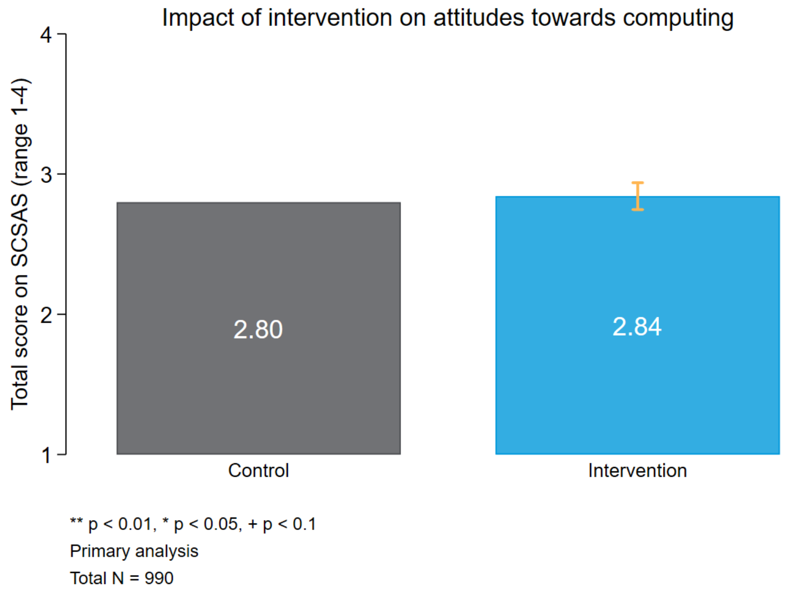

Figure 3: Model-adjusted SCSAS scores by treatment group

For a more detailed description of endline survey responses, Table 7 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 7: 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	278	2.71 (0.61)
	Baseline	Treatment	511	2.87 (0.56)
	Endline	Control	375	2.68 (0.62)
	Endline	Treatment	615	2.84 (0.61)
Interest	Baseline	Control	278	2.65 (0.63)
	Baseline	Treatment	511	2.74 (0.58)
	Endline	Control	375	2.85 (0.72)
	Endline	Treatment	615	2.93 (0.70)
Belonging	Baseline	Control	273	2.82 (0.58)
	Baseline	Treatment	497	2.98 (0.58)
	Endline	Control	371	2.99 (0.57)
	Endline	Treatment	604	3.12 (0.56)
Usefulness	Baseline	Control	273	2.77 (0.65)
	Baseline	Treatment	497	2.86 (0.63)
	Endline	Control	371	2.95 (0.63)
	Endline	Treatment	604	3.02 (0.63)
Encouragement	Baseline	Control	273	2.32 (0.70)
	Baseline	Treatment	497	2.48 (0.66)
	Endline	Control	371	2.53 (0.70)
	Endline	Treatment	604	2.71 (0.73)
Overall SCSAS score	Baseline	Control	278	2.65 (0.49)
	Baseline	Treatment	511	2.79 (0.47)
	Endline	Control	375	2.80 (0.52)
	Endline	Treatment	615	2.92 (0.52)

3.2 Secondary analysis: effect of the intervention on stated intention to study computing in the future

There was no conclusive evidence that the intervention positively impacted girls' intention to study computing in the future relative to the control group. The proportion of girls stating they intended to study computing for the full girls sample was 47.9%. For the intervention group it was 49.8% and for the control group it was 45.0%. Compared to the control group, the pre-specified multiple imputation model found a difference of 4.3 percentage points (p=0.453) for the intervention, which is not statistically significant at conventional significance levels. Compared to the primary analysis, this effect size (while not statistically significant) is larger, but given the high p-values, we cannot be confident that this is indicative of an impact of the intervention rather than random chance. This finding was consistent across the missingness indicator and complete case analysis model specifications.

Table 8: Impact evaluation results for secondary outcome

Outcome: Intention to study computing	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group mean	45.0%		42.8%
Treatment group mean	49.8%		49.5%
Estimated treatment effect (in percentage points)	4.3pp	3.3pp	5.4pp
N	994	995	789

Figure 4 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.

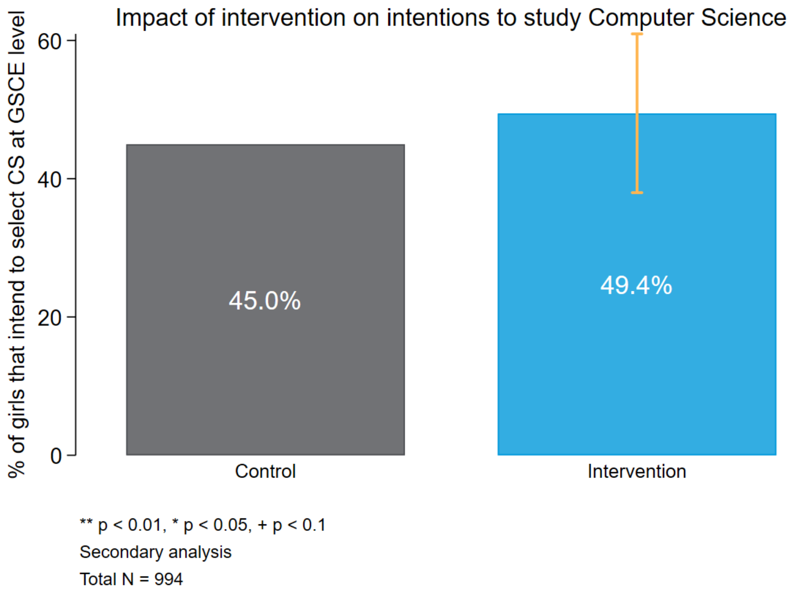

Figure 4: Model-adjusted intention scores by treatment group

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

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.

Key findings:

- Teachers found it challenging to complete the 12 lessons within the timeframe of the intervention.
- Teachers often shortened the lessons by skipping the starters or final sections.
- Fidelity to working in pairs and using the half-built Scratch projects was high.
- Teachers all made small adjustments based on their individual school contexts.

Number of lessons delivered

Generally, teachers reported that it was a challenge to complete the 12-lesson unit within the timeframe of the intervention. Both COVID-19 related (teacher self-isolation) and non-COVID-19 related causes (having ‘Coding week’ taking place in school causing teachers to ‘get behind’) were cited.

Teachers differed in how they responded to the challenge of completing the lessons within the timeframe: one taught two computing lessons a week for a number of weeks, one continued the lessons after the intervention time frame had finished, one did not teach the final lesson in the unit and one managed to complete all lessons within the time-frame. As the teachers who spoke to us for the IPE are likely to have been teachers who were particularly committed to the intervention, it’s plausible that these teachers completed more of the intervention lessons than other teachers in the treatment group.

” For three or four weeks, I had to do two lessons in a week - which was difficult to fit in. Obviously, there's a lot of timetabling issues in school - maths and English and the core subjects will take preference over everything else.” (S04)

In addition to some lessons not being taught, there were high rates of pupil absence from lessons. In the first week of the Autumn term in 2021, nationally primary attendance was around 90%, or on average three children absent from each lesson[19]. In one of the lessons we observed, 12 pupils were absent with COVID-19. A further cause of pupils missing computing lessons was small groups being taken out for interventions (such as reading or maths catch-up groups) during lesson time: we saw this in two of the case study schools.

The combination of teachers not being able to teach the full range of lessons and pupils being absent from lessons means that individual pupils might not have accessed the full range of lessons that was originally intended. The challenge of completing lessons might have contributed to the attrition in the trial described in section 2.2.3 as teachers might have felt that they couldn't fully participate in the project if they saw that it would be a struggle to complete the unit of lessons within the allocated time frame.

Structure of lessons

Teachers consistently reported that they cut sections from the lesson plans, in order to fit the lessons into their timetables. On the whole, they skipped (or passed through very quickly) the starter activities at the starts of lessons, or the plenary activities at the ends. This was partly because teachers either did not have a full hour timetable for the session, or they found that time was taken up with other tasks such as children arriving and logging in.

"I never do the 2 slides at the end: how did you find it, all the success criteria and did we enjoy it. I never get around to those. Whether someone with an hour would better do that, I don't know." (S02)

As well as shortening the starters, teachers also adapted them so that they could be done as a whole class in front of the whiteboard, as opposed to children needing separate pieces of paper (for example, students creating a flip book using post-it notes). This saved preparation time for teachers and implementation time within the lessons.

"Some of the bits, like with the A3 and writing stuff down. That's good and in an ideal world, I think you could do that, but with time restraints, I couldn't do it all the time" (S04)

"We adapted that and did it on our whiteboard. We still did them, but just not as beautifully as the plan said." (S03)

Teachers also reported that they skipped some of the slides at the start and end of the lessons, for example, those which encouraged pupils to reflect on the paired nature of pair programming, because they felt they weren't necessary once the students were familiar with the content of the sessions.

[19] Primary Absence: see 2021 Week 36 (2021/07/09) https://explore-education-statistics.service.gov.uk/find-statistics/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak#dataBlock-f827fefa-e90b-4cd3-37de-08d9fdd68e7c-tables

"*I skip a number of the slides or keep the main activity content, mainly because we are 11 weeks into it…those kinds of conversations about how we feel about this and that..thumbsup to give feedback...I don't need the slide for that" (S01)*

Some teachers reported that they simplified the structure of the lesson to ensure that pupils would have enough time on the paired activities. They adapted the structure so that there was one section in which the teacher delivered the main input, and then the students did the Pair Programming, as opposed to pupils moving between the main input, Pair Programming, a second main input and a second Pair Programming.

Use of power-points

The power-points were provided for the teachers as the main presentation which they would work through with their pupils. These contained the demonstrations and instructions for the start-of-lesson and end-of-lesson activities, as well as for the Pair Programming activities.

Interviewed teachers all used the RPF produced power-points within the Pair Programming lessons but they varied in the extent to which they adapted the power-points. Given different teachers' perceptions of what was appropriate for their classes, it was helpful that they were able to easily adapt the slides.

"I changed one or two bits, but I just had a look, took bits out, added bits in - but the basis was there, so it made it a lot easier. You can work with that frame and add to it or change it as you need to." (S04)

Some teachers reduced the amount of text on the slides, as they felt the current amount of text was too much, whilst others thought the amount of text was appropriate for their class. The extent to which teachers edited the slides was linked to how long teachers spent preparing for the lessons: those who edited less reported that the lessons were quick to prepare for.

Whilst using the slides in lessons, teachers tended to turn the statements into questions, as opposed to reading aloud the statements to the class. For example, instead of reading through the blocks (sections of code) for the 'Benna's Ball' animation, the teacher asked the pupils to predict which blocks would have been used. This is in line with guidance that questioning within lessons is an important strategy for building pupil engagement[20].

Paired work

Generally, fidelity to working in pairs during the Pair Programming lessons was high. In 30 out of the 32 treatment schools which completed the endline survey, pupils reported usually working in pairs in computing lessons (94%). In the control schools, the proportion was 6 out of 24 schools (25%), implying that paired work of some sort was fairly common as business as usual. However, the way in which paired work is set up might differ between those pupils who were working in pairs in the control group, and those working in pairs in the

[20] Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American educator*, *36*(1), 12.

treatment group. In the control group, pairs might not have been assigned separate responsibilities, and it's unlikely that the same amount of lesson time would have been dedicated to discussion of the success of paired work.

The interviewed teachers varied in whether they set up their pairs of drivers and navigators as 'mixed ability' or 'set ability': in some classes they were matched on computing attainment, and in other classes, they were set up as mixed attaining pairs (with one higher attaining and one lower attaining student). The instructions for teachers on how to set up these pairings were relatively open: "Potential pairings could be between pupils with relatively more or less advanced skills, or between pupils who have similar attitudes towards classwork or a 'social affinity'[21]."

Two of the interviewed teachers based the pairings on what they used for maths lessons: this meant that the pupils were already familiar with working with their partner and expectations about paired work were already set.

"There was just a bit of tweaking work on those pairs [maths pairs] that weren't really working, and change one or two about, just for specific knowledge in computing and abilities in computing." (S04)

Teachers generally set up the pairs so that they were mixed gender, although that was not always possible given the gender balance of the class.The online training did not specify whether teachers should set up mixed or same gender pairs. The teachers who set the pairs up mixed gender reported that this was because they 'normally' use mixed gender pairs.

It was clear that time and effort had gone into setting up the pairings.

"I thought very carefully about pairing the children up from my knowledge of, how they are, their current abilities, where they progressed to in terms of computing." (S01)

Within the four case study schools, fidelity to having a driver and navigator with clearly defined roles was also high: pupils understood the different responsibilities which went with the different roles and responded promptly to the indication that it was time to swap roles (normally by changing seats). The ways in which teachers indicated it was time to swap differed: from a triangle to an online timer, to teacher announcement. One teacher said this could be a challenge to keep track of when it was time to switch.

"Sometimes I have the time on a board; then I switch over. Sometimes I forget, so I just make up five minutes in my head…sometimes it gets a bit difficult, managing that, but I like the idea of switching over." (S04)

Use of Scratch projects

For the Pair Programming activities, children were given 'half-built' Scratch projects which they continued working on. This meant that some of the code (blocks) and characters (spites) were already in the project, so the pupils didn't need to work from scratch. This

[21] Retrieved from online training materials.

enabled the pupils to focus more on the specific objective of the lesson, for example sequencing.

Fidelity to the use of the half-made Scratch projects was high. Interviewed teachers described pupils completing these projects as a central part of the lesson and did not tend to edit the projects before sharing them with pupils.

There was wide variation in how teachers shared these Scratch projects with pupils: accessing from an online portal; saving the school server; or pupils typing the url directly into a browser. Generally, the teachers set up the sharing of the Scratch programs in the way that was in line with what they would normally do in computing lessons. It seems that this flexibility of teachers to share the resources in different ways was a strength of the project as different teachers were able to use the method which worked best within their school context.

Additional adaptations to the school context

In order to make the lessons fit within their individual school contexts, the teachers interviewed had each made additional minor adaptations to how the lessons were delivered. For example, in one school, the teacher brought in a question which referred to the schools' values; in another, the teacher added an image of a storyboard that the pupils had created in English the previous year, for the lesson which referred to storyboards.

4.1.2 Feasibility

Key findings:

- Characteristics of the learning environment such as pupil behaviour, pupil familiarity with routines and pupil baseline understanding of computing affected the ease with which teachers could implement the intervention.
- Teacher confidence and subject knowledge were also important moderators - some were concerned that teachers who did not have experience in teaching the national curriculum objectives covered in the lessons might find delivery more challenging.
- Some practical considerations such as length of lessons and the resources available in the classroom affected feasibility.
- Online training generally prepared the teachers well for delivery, but could have been shorter and placed at a different time of the school year. Face-to-face training might have been more memorable than the online asynchronous delivery.

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.

Learning environment

Pupil behaviour

Where pupil behaviour was generally good, teachers found it straightforward to facilitate the paired work within Pair Programming. In the lessons we observed, pupils took on the roles of driver and navigator, with the drivers generally listening well to the navigators' instructions, and the navigators letting the drivers control the mouse and keyboard.

"I've got a really good class; really well behaved, so I don't have any issues there [with working in pairs]" (S04)

As the lesson observations fell close to the end of the 12 week unit of lessons, it's possible that the behaviour we saw reflected the improvement pupils had made in working together throughout the lessons: teachers did comment that it took a few lessons to set up behavioural expectations.

"*There were a few times…where they weren't collaborating properly…At the beginning [of the unit] we had to have at least two or three [lessons]....just reminding them what the role of the navigator and what the driver was." (S04)*

Some teachers noted that, if they had a class with more challenging behaviour, it would be more difficult to manage the paired work.

"If you have a class that's quite noisy and behaviour is an issue you might keep them on their own because they're quieter. If you're trying to keep a lid on the energy of the room, making them work on their own is better even if educationally it's not as sound." (S02)

Pupil routines

Teachers reported that embedded pupil routines enabled the lessons to flow smoothly. For example, in the observed lessons, pupils understood that the expectation was for them to enter the room, log in to the computer at their station, and then come and sit on the carpet by the teacher ready for the main input. Pupils knew where they should be sitting, who their partners were, which partner was starting in which role, where to find the Scratch project and how to switch places. This minimised the amount of time spent talking about these practical considerations during lesson times.

"Now they've got used to the system, this is the lesson and this is what we're doing. That's been good." (S02)

Pupil confidence in computing

According to teachers, pupil confidence in computing made it easier for lessons to be delivered. In one school with a computing specialist teacher, the pupils were confidently using the Scratch vocabulary required for the lesson (e.g. glide, repeat, motion). This meant that the teacher was able to focus on the lesson objective, as opposed to spending time introducing the other concepts. Another teacher felt that pupils' previous experience with Purple Mash (a programme similar to Scratch) enabled pupils to thrive within the Scratch lessons.

"*Then it was just teaching them about Scratch and just the block based code, and how it's similar to Purple Mash code and how they've got transferable skills from that to this one" (S04)*

Teacher characteristics

Teacher subject specialism

All of the teachers interviewed were either class teachers with additional responsibilities for computing, or taught mainly computing. During their feedback, these teachers made an effort to convey that their experiences, as specialists delivering the intervention, might not necessarily have been in line with classroom teachers who do not have any particular specialism in computing.

”[About the training] *I've got an idea of what we're doing. Because I'm in the privileged role of a specialist teacher” (S01)*

Teachers noted that it would have been easier for specialists to deliver the lessons than non-specialists. One concern was that the subject knowledge required for non-specialists could have been overwhelming.

“Obviously it says you don't need any experience, but I think if you were coming to it new and you hadn't used it, it would be quite daunting, because there's a lot of information, there's a lot of content to get through" (S04)

Specialist teachers reported that non-specialist teachers did enjoy delivering the lessons, although, in one school (reported by the interviewed specialist), the non-specialist teacher tended to get less far through the lesson materials than the specialist.

"*Every week, afterwards, she* [the non-specialist] *was like 'Oh it was really good actually. It was dead simple to follow.'..We got to the extension bits quite quickly every week and then she would come and say 'oh we haven't managed to get that far, but we did do this”. (S03)*

It's worth noting that the subject knowledge required of teachers was within the expectations of what all primary teachers would need to be familiar with in order to teach the National Curriculum. If there are teachers who find it a challenge to deliver this content, this is a reflection of a lower than expected level of confidence in teaching computing, as opposed to a criticism of the design of the intervention itself.

Learning environment expectations

One teacher noted that she had initially felt tension between the amount of noise that the class were making and the fact that some noise would be an inevitable consequence of pupils working in pairs. She felt that teachers need to be comfortable with a certain level of noise for the Pair Programming to work well.

“As soon as we started, I was saying 'shhh', because we're an open corridor, I don't like the noise. You have to fight your desire to get a quieter class.” (S02)

Time and practical challenges

The challenge of including all of the lesson content within the time allocated for the lesson and alongside the other time-consuming practicalities of teaching emerged as a strong theme in both teacher interviews and lesson observations. Teachers commented that even if a computing lesson had an hour allocated to it within the school timetable, that didn't mean that they could spend a full hour on lesson content. Time consuming practicalities included handwashing, wiping down laptops, collecting laptops, passing laptops onto another classroom, arriving in the classroom, logging on, taking the register, and saving work.

"That's the thing in that plan, there's no time to log on, no time to save the work. I was looking at the timings, thinking why does this look fine on paper? I know full well that won't happen in real life because there is no five minutes to just log on and come and sit on the carpet." (S02)

Teachers also reported that there were some set-up costs (in terms of teacher time) to preparing the Pair Programming lessons for their pupils, for example creating the Scratch accounts for the children.

Whilst the four case study schools did have enough hardware to run the Pair Programming lessons, a common theme which emerged was that their hardware now better suited the needs of their pupils, as COVID-19 catch-up funding had been allocated to IT and the Department for Education had sent out laptops to schools.

"If we didn't have catch up funding, we wouldn't be able to pay for the hardware we need." (S01)

Practical challenges varied depending on where the pupils completed the lessons. If pupils were in their 'normal' classroom, the teachers noted that they had to spend time transporting the laptops. If classes took place in a computer suite, it was difficult for the pupils to complete the paper based tasks, as they didn't have pencils and books easily accessible. Teachers did not present these challenges as insurmountable: they had each adapted how they delivered the lessons (normally by cutting content at the start or end) to make sure they could deliver the main content of the lesson and pupils could have a reasonable amount of time working on the Pair Programming tasks.

Lesson preparation

There was striking variation between teachers in how much time they felt they needed to spend preparing lessons. Some of the teachers we spoke to (which were also the teachers who made the most significant adaptations to the slides) felt they spent quite a long time preparing.

"It was quite dense, in regards to the content and detail in the lessons, you need a bit of time to go through it all and have a look." (S04)

Other interviewed teachers felt that the lessons were quick and easy to prepare and comparable with the amount of time they would 'normally' spend preparing for computing lessons.

A potential source of this divergence could be whether teachers are comparing lesson preparation time with other commercially produced units of lessons (which they might expect to spend some time going through) with lessons which they have planned themselves previously (which might take very little preparation time). It seemed that teachers generally thought it was reasonable that they should make some adjustments to the provided planning.

“ You can work with that frame and add to it or change it as you need to” (S04)

School support

The support which teachers received from senior management affected how easily teachers felt they could deliver the intervention sessions. When the senior leadership team were on-board with the aims of the programme, the teachers cited this as a source of support. In one school, both the headteacher and computing lead were particularly interested in supporting girls’ engagement with computing:

“It’s a testament to the senior leadership team, particularly the head and the previous head as well, their vision for computing to be an enabler as it should be.” (S01)

Another teacher noted that the school being unable to arrange additional release time for the teacher to prepare for the course had been a challenge. This is despite the fact that RPF offered financial reimbursement to schools to cover the cost of teachers and teaching assistants completing the training.

"*It would be good if the teacher had time out [of class] to go through information, so they've got a clear, logical structured learning journey in their mind about how they're going to do the lesson or unit." (S04)*

It’s worth noting that these case study schools might be schools which have a high level of support from the senior management team; teachers in schools where computing is less supported might not have been granted permission from the leadership team to take part in these additional research activities.

Online training

Changes due to the COVID-19 context

The original plan had been for teachers to attend face-to-face training during which the Pair Programming approach could be modelled in pairs with teachers. This would have allowed the nature of the collaborative aspect of Pair Programming to be fully explored. Due to COVID-19 constraints, RPF were not able to deliver in person training, and instead delivered the training online. This was not RPF’s preferred delivery model and it is worth noting that, when the theory of change was jointly constructed between RPF and the evaluations, it was assumed that the teacher training would take place in person.

Reach

Of the 58 schools allocated to the treatment group, 43 completed the online training. All of the 32 treatment schools which completed the end-of-project survey had also completed the online training.

The number of members of staff completing the training varied between schools, from 1 to 5 completions from each school which completed the training. Where more than one member of staff completed training, this was due to multiple class teachers or teaching assistants also completing the training. At 30 out of the 43 schools which completed the training, only the class teacher did the online training. This might have been because of the challenge of arranging release time for teaching assistants to complete training, despite the fact that RPF offered schools reimbursement to cover the cost of staff cover for training time. One of our interviewed teachers reported that he would have liked for the class teaching assistant to complete the training, but that it hadn't been possible to organise release time.

"I didn't do it with my TA, because she was sometimes in the lesson/out of the lesson, and it would mean having to take time out, and I don't think the school would allow that." (S04)

The members of staff delivering the intervention did not all complete all of the training available: in addition to the online course, there was a recording of a webinar, which 70% of those completing the training reported that they had watched. Had the training taken place in person, it would have been more straightforward to ensure that the teachers and teaching assistants completed all of the training available.

Content

Teachers were generally positive about the content of the online training sessions, but felt that the subject knowledge demands might have been high for non-specialist computing teachers. In answering how they thought the training would seem to a non-specialist, one teacher responded:

"I imagine it would have been really hard because I started so far back, and I remember even the simplest things…. even a description you can't follow if you didn't know what you're talking about." (S02)

Part of the teachers' positive assessment of the content was as a result of its similarity to what they would have been teaching, had they not been taking part in the research project. Teachers did not have any suggestions for changes to the content of the training, although it was suggested that the online course itself could have been shorter. Of the 66 teachers and teaching assistants who completed the online training, 33% reported that it took longer than three hours to complete.

"As short an amount as possible...that's when you're going to get a better response from a non-specialist" (S01)

Accessibility

The interviewed teachers struggled to remember the training, when asked about it at the end of the unit of teaching. They all commented that it had taken place a long time ago and some would have preferred to have had the option of completing it in September as opposed to before the summer holiday. It's possible that the training would have been more memorable, had the teaching staff had the in-person experiences of modelling the Pair Programming approach that RPF had planned.

"You get back in September and you're going: 'What was in that training?' " (S02)
"Teaching staff are knackered....it was the last thing I wanted to do at the end of the term." (S01)

Teachers spoke of minor challenges with following the links to the various parts of the training and ensuring they had found all of the different resources needed for each lesson.

> *"I did get a bit confused with all of the links. I couldn't find some bits of training" (S03)*

An experienced teacher suggested that video demonstrations (as opposed to written instructions) might have made the material more accessible for non-specialist teachers, as a training video could demonstrate exactly where the teachers need to click whilst doing their own demonstrations for the class. Alternatively, if the training had been able to go ahead in person, teaching staff might have had the opportunity to get advice from RPF staff on their use of Scratch, whilst they were practising building projects.

Effectiveness

Interviewed teachers generally felt that the training had prepared them well for delivery and could not think of anything that was 'missing'. Of the 66 school staff who completed the online training survey, all but one found the online training either quite useful or very useful in familiarising themselves with the resources of the project. 61 out of the 66 school staff reported that they felt either quite confident or very confident in delivering the resources following the online training. It's possible that the variation in school staffs' confidence in delivering the resources (2% not at all confident; 6% not very confident; 70% quite confident; 22% very confident) reflects the variation in computing subject knowledge of the different school staff attending the training.

4.2 Intervention

4.2.1 Quality

Key findings:

- Teachers thought that paired work was particularly well suited to computing.
- They thought the quality of the prepared resources was high and that the resources were sufficient.
- Teachers suggested there is some scope for improvement around the clarity of pupil task instructions.

This section explores teachers' and pupils' perceptions of the quality of the Pair Programming resources.

Paired work

The idea that paired work is particularly well suited to computing came through strongly in the case studies: teachers implied that the Pair Programming approach has been an appropriate approach to trial in their computing lessons because the clearly defined roles support collaboration. Some reported that they had either already started using the Pair

Programming approach with other cohorts, or that they would continue to use it, after the trial had finished.

“I don't know why I've never thought to do computing like that, actually because it's a really good vehicle for the fact that there are two roles, clearly defined. There's all your conversation and knowledge comes through that, and then they're both equally having a turn." (S04)

One teacher was concerned that working in pairs could lead to the workload being unbalanced between pupils, with one pupil taking on more responsibility and the other ‘coasting’. However, another teacher suggested that, because of the roles and swapping over, Pair Programming did ensure that paired work was balanced.

"*We've always had collaboration, but what's helped [in this project] is that [in the past] we've had a dominant one. Whereas this approach, obviously, you get to see that they are both equally [contributing]” (S01)*

Quality of the written resources

Overall, teachers were positive about the quality of the written resources (e.g. unit overview, lesson plans, power-points) and felt that they had everything they needed - there was nothing missing. This is perhaps the ‘flipside’ of other comments teachers made about there being a lot of material for teachers to get through when preparing lessons, which sometimes made it time consuming. Teachers emphasised that they found the half-built Scratch projects particularly useful as a source of pupil support.

“The part-made programmes were really helpful - because I think children need that with the algorithms and coding” (S04)

Clarity

Interviewed teachers felt the clarity of the resources was generally good. Having the code on the power-point and having the overview of the learning journey was noted as being particularly helpful. There were some minor concerns around the clarity of pupils’ Scratch task instructions. For example:

“It wasn't always exactly clear what the children had to do…like the animated one today, Benna's Ball..I thought, at what point does it need to stay squashed and then return to round? How long does it stay squashed? Some of the times when I’ve looked at them I’ve thought: All right. I’m not exactly sure what the children are supposed to do there, but this is what I think it is - we’ll tweak it” (S01)

Level of challenge

All of the treatment schools were able to decide whether to run the intervention with a year 4 or year 6 cohort - teachers were advised to consider how much previous coding experience the children had before making the decision. If the year 6s would have already covered much of the Pair Programming material within other computing lessons, it was recommended that teachers deliver to a year 4 cohort. Of the 32 schools which completed the end-of-project survey, four chose to deliver the Pair Programming lessons to year 6 pupils and 28 delivered to year 4 pupils. The four case study schools were all delivering to year 4 pupils.

Teachers' perception was that the level of challenge was appropriate for their year 4 cohorts, acknowledging that the children in their schools had a good background in coding. One teacher had thought that the maths requirements (for example understanding that there were 180 degrees in the internal angles of a triangle) was quite high for year 4s, but had dealt with this by giving a brief explanation and creating a poster which the children could refer back to during the lesson.

The pupils themselves were generally content with the level of challenge. Whilst one girl who had been identified as 'highly engaged with computing' claimed that she would prefer if the lessons were 'a bit less simple', another less engaged pupil said Scratch was 'too hard', suggesting that the pitch of the lessons is somewhere in the middle of the distribution. During the lesson observations, the pupils started by being unsure of how to complete their Scratch activities, but by the end of the session, the majority of pairs had successfully created the code. This implied that the level of challenge was appropriate for these classes.

4.2.2 Responsiveness

Key findings:

- Responsiveness to the intervention from both pupils and teachers was generally high.
- Pupils took on the roles of driver and navigator, as intended.
- Pupils engaged well with the lessons and enjoyed completing the half-built Scratch projects with their partner.
- Teachers reported enjoying teaching the Pair Programming lessons.

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

Working in pairs

Observations of lessons, alongside discussions with pupils and teachers all point to pupils having taken on the roles of driver and navigator well. Within lessons, it was clear which pupil was fulfilling which role, with the driver generally sitting directly in front of the mouse and keyboard, and the navigator sitting to one side, and pointing at the screen. The discussion between the pupils also reflected the roles which they were holding. These interactions tended to be an on-going two-way discussion, as opposed to a one-sided process in which only the navigator gave instructions.

As well as discussing which blocks should be used, pupils talked about the specifics of inserting the blocks (e.g. how to spell words they were inputting). A lot of the talk was around planning what pupils were going to do next, for example:

"*Stop - we haven't done the second one yet…Maybe put 'ball round' then put it there, that might work*".
(S01: female pupil)

Some pupils expressed a preference for being either a driver or a navigator, but as they swapped regularly, all pupils seemed content with the setup. The interviewed teachers all expressed that the pupils enjoyed the paired element of the lessons.

"They really liked working together as driver and navigator. They liked the aspect of swapping around." (S04)

In some lessons, teachers noticed that some pupils had slipped into the navigator taking a passive role in the activity. If this happened, the teacher recapped the Pair Programming expectations.

"There were a few times when we had to recap at the beginning of the lesson or at the end of the lesson, where they weren't collaborating properly, or they weren't taking on their roles properly - they were just letting the person who was on the computer get on with it, and then wait their turn" (S04)

Teachers noted a tension between some pupils saying that they would prefer to work independently, and the teachers' perception that pupils were engaging well with the Pair Programming approach. Pupils reported that they sometimes like working independently in part because they can concentrate more easily, and in part because they don't need to deal with the potential conflict that comes with working with one of their peers.

"Even those who are maybe a little bit more reluctant…those who put their hands up today and said they still prefer to work independently, they are still all engaging quite clearly in that with their pair and doing it really, really well. However much they say they prefer working independently, I think they clearly showed how much they enjoy it, engage with it. And you know they're achieving with it - so we should be doing this." (S01)

The pupils themselves were able to articulate that they felt there were elements of paired work that they enjoyed and other elements they found more challenging. Elements which they enjoyed included working with friends, being able to help each other, and the opportunity to develop friendships.

"*I like working with both [both in a partner and by yourself] because when you do Pair Programming you're collaborating with your partner, making links and you have to tell them what to do. But if you have a really good idea and then they put the wrong thing in the wrong place, it's quite annoying." (S01: female pupil)*

Pupils also noted that who their partner was made an important difference to how easily they felt they could work with them.

"Sometimes when you're in your pairs...it's trickier in different pairs because you don't pick your pairs - they get picked for you. Sometimes it might be easier working in one pair than another" (S02: female pupil)

Engaging with the main input

Pupils' engagement with the main input (when the teacher was talking through the slides and demonstrating the activity) was high: the lesson observations, pupil discussions and teacher interviews all pointed to high levels of pupil enjoyment and attention during lessons. In the observed lessons, more than half of the pupils were putting their hands up to answer questions; pupils had their bodies turned towards the teacher and tracked where the teacher was pointing with their eyes.

"The children have enjoyed it...we're up to week 11 and the kids are still just as enthused…I think they've engaged really well, I mean, I can't think of anybody who is not engaging with it." (S01)

In one observation lesson, when the teacher announced the learning objective, the pupils shouted 'Yes!'. This is consistent with the children in the focus groups struggling to think of a part of the lesson which they didn't enjoy.

During class discussions, girls in particular tended to build on one another's responses. Teachers skillfully drew less engaged girls into the discussion by directing questions at them.

Girl A: "I think it's a forever loop"
Girl B: "I agree with [Girl A] - I think it's either a forever loop or it could be a repeat loop" (S01: female pupils)

Engagement with the Scratch activities

Pupils and teachers reported the part of the lesson which the pupils most enjoyed was working in pairs to complete the Scratch activities. Pupils tended to be on task and paired discussion was focused around what the pairs were trying to achieve within their Scratch projects, whether programming a ball so that it 'squashed' when it hit a wall, or sequencing the movement of sprites (characters) on a space background.

Boy: "What do you want him to say? Shall we try..let's set off?"
Girl: "Yeh but we need a start block first" (S02: male and female pupils)

Teachers suggested that the pupils particularly engaged with the second unit of activities: animation, compared to the first unit: drawing, because the pupils had more freedom in the Scratch projects they were creating.

"Overall, they all preferred the animation unit to the drawing one, because they had more ownership of that, and they all loved that." (S03)

One teacher reported that girls were more likely than boys to focus on the aesthetics of the Scratch project, by using the drawing tool and editing colours. This sometimes directed girls' attention away from the main objective of the lesson, for example, sequencing. The two lesson observations reflected this pattern, with girls tending to spend more time making changes to colours and patterns of sprites and backgrounds than the boys. This was particularly apparent where there were two girls in a pair.

"The girls engage more with the aesthetics: 'Let's change the colour of this, the sprite, the background.' " (S02)

Teachers' preference for the lessons and intention to continue

The teachers interviewed reported that they had enjoyed delivering the lessons to their pupils and all shared that they are planning to continue with using the paired approach in computing lessons, even when teaching other classes. The teachers considered their own school contexts when planning their continued use of the approach, for example, one teacher planned to use Pair Programming on alternate weeks, to respond to some pupils' stated preference for working independently.

"I absolutely loved it, and as computing lead, I'm thinking that we might teach all of our lessons like that." (S03)

"Yes, the resources were really good; that would be helpful in the future - maybe for next year, if I decide to do Scratch, or even when I start a code club again; that's really helpful for that." (S04)

4.2.3 Mechanisms

Key findings

- According to teachers, working in pairs increased girls' confidence in computing.
- Working in pairs was also linked to girls perceiving computing as a more collaborative subject.
- Teachers suggested Pair Programming could have affected girls' outcomes through two mechanisms not hypothesised in the logic model: giving a sense of ownership and focusing on content which is appealing to girls.
- Some teachers felt the Pair Programming lessons supported girls' particularly, whilst others felt the approach was equally positive for girls and boys.
- If pupils are skilled in working in pairs and have open ideas about gender and computing, there might be more scope for the mechanism to take place.
- Conflict between pairs or a school context in which much has already been done to address the barriers to girls' engagement with computing might restrict the scope for the collaboration mechanism to impact outcomes.

At its broadest level, this intervention aimed to use a Pair Programming 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 4 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.

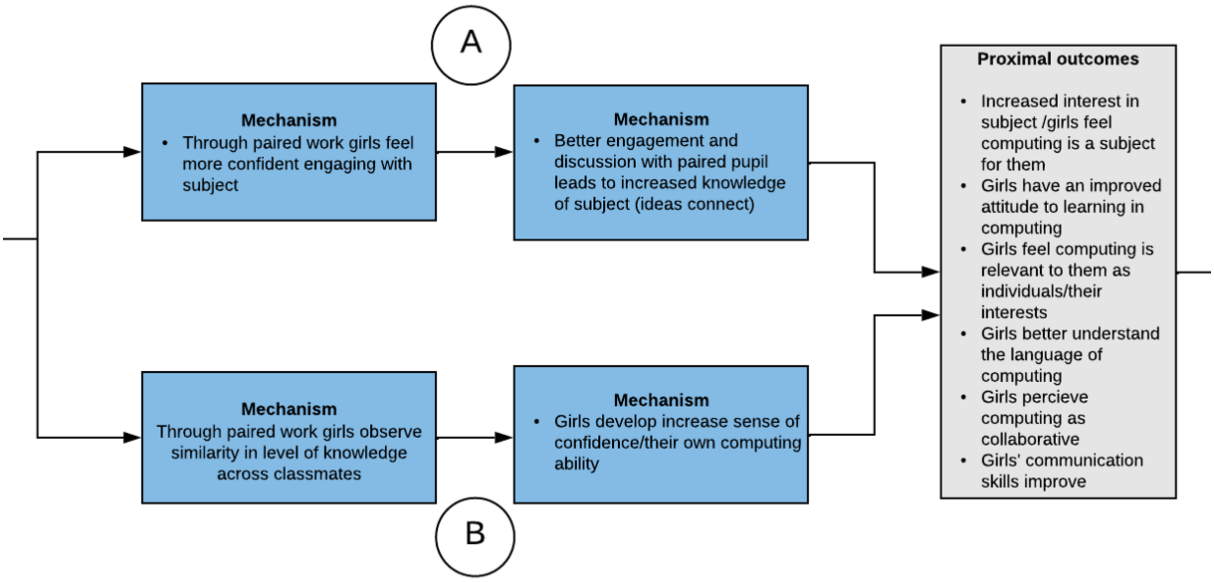

Figure 5: Pair Programming logic model mechanisms and proximal outcomes

Mechanisms observed

a) Pathways to collaboration observed

Within the mechanism of paired work leading to increased collaboration (one of the proximal outcomes), teachers identified two pathways which mapped onto Pathway A and Pathway B in the logic model.

Pathway A of the logic model

Through paired work girls feel more confident engaging with the subject → Better engagement and discussion with paired pupil leads to increased knowledge of subject

Teachers and pupils felt that having the support of a partner boosted pupils' confidence. This seemed to work through the pupils using one another as resources for when they were finding the work challenging. Pupils reported that they liked having a partner there to support them. Girls referred to this being useful if they '*needed help*' or '*were stuck*'.

"I got to work with someone else so they could help me if I needed help. I prefer working with a pair (girl)." (S02: female pupil)

"It might be easier to do Pair Programming [compared to 'normal' lessons] because if you're stuck your partner can be helpful." (S01: female pupil)

The logic model suggested that the confidence engaging with the subject would lead to increased knowledge of the subject. Whilst some teachers did suggest that they thought pupils learnt more in Pair Programming lessons than in 'normal lessons', they did not explicitly attribute this improved understanding to the pupils' confidence:

"The pair programming definitely helps. I think it boosted their confidence. They had a partner to work with so immediately that makes it more interesting for them. I don't know, I feel like they just acquired more knowledge." (S03)

One teacher suggested that there might have been a cycle of pupils working in pairs and, as a result gaining confidence, which then led them to want to work independently.

"Towards the end, they were getting frustrated [about working in pairs]...maybe that's a sign that they're more confident." (S02)

Pathway B of the logic model

Through paired work girls observe similarity in level of knowledge across classmates → Girls developed increased sense of confidence in their own computing ability

The theme of observed similarity of knowledge was less apparent in the evaluation than Pathway A, although one teacher whose girls were generally less confident in computing than the boys in their class did report that the paired work built confidence and helped the girls to see that they were able to do tasks which they previously may not have believed they could do.

"I do think that having that equal time to have a go at both, thinking of the girls I've got, will have helped my girls, because they lack a bit of confidence. They were learning very quickly that actually 'Yes, we are sure. We can do this.' " (S03)

b) Evidence of collaboration leading to proximal outcomes

In line with the logic model, there was a consistent message from both pupils and teachers that using the Pair Programming approach encouraged girls to perceive computing as more collaborative. When pupils were asked to rank the skills necessary to do well in computing lessons, they referred specifically to communication and teamwork being important for Pair Programming lessons, and contrasted this with 'other' computing lessons. They also referred to the roles of 'navigator' and 'driver' when explaining why teamwork and collaboration were important. Pupils ranked teamwork and communication over creativity, focus, problem solving and resilience. This implies that the paired approach did enable a sense of computing as a collaborative subject.

"We moved up communication [in the order of importance ranking] because in computing if there's a navigator and a driver, you need to communicate." (S01: female pupil)

"Is it [this question] for Pair Programming? if it is, it would be teamwork". (S01: female pupil)

"Teamwork [is the most important skill for computing] because if you were doing Pair Programming and you didn't do teamwork, then maybe you would not agree with stuff and then you'd get in a massive argument" (S01: female pupil)

In addition to increasing girls' sense of computing as a collaborative subject, some teachers also felt that Pair Programming had improved their pupils' collaboration skills: a second proximal outcome in the logic model. Teachers thought this was in part due to the structured roles within the Pair Programming approach.

"*I was quite explicit obviously of what those roles were. I think it has improved collaboration, definitely" (S01)*

Girls tended to be positive about the collaborative elements of the Pair Programming approach - often choosing it as their 'highlight' of the lesson. In a few cases, girls shared that their favourite part of the lesson was working with other people, and then gave a specific example, such as:

"My favourite thing about the computing lesson today was working with other people. Me and Oliver were setting the code to like 1000 and being silly with it to see what happens. It's really funny" (S01: female pupil)

A third proximal outcome from the logic model which was noted by teachers, was girls having an improved attitude to learning in computing, as a result of the collaboration enabled by the paired approach. This was highlighted for girls who had a baseline of lower levels of engagement with computing lessons.

"*I felt like some of my girls were really quite [makes bored noise] at the beginning, and by the end of it, especially when we did the second unit about the animation, loved it. They absolutely loved it.(S03)*

We also noted that some of the proximal outcomes identified in the logic model did not emerge in the implementation and process evaluation. These included: girls feeling that computing is a subject for them; girls feeling that computing is relevant to them as individuals/interests; girls better understanding the language of computing. It's possible that the intervention did have an effect on these proximal outcomes but, if so, this did not come through in the qualitative data from the four case study schools.

c) Additional pathways observed

The implementation and process evaluation highlighted two additional potential pathways through which the Pair Programming approach could improve intended outcomes. These pathways were separate from the collaboration mechanism.

Sense of ownership

Teachers reported that pupils' ownership of their Scratch projects increased engagement. Both pupils and teachers noted that having the flexibility of making Scratch 'do what pupils wanted it to do' was one of the attractions of the Scratch units. Pupils talked about being creative and using their favourite sprites (characters) or building in backgrounds which linked to places in their own lives.

"Scratch is pretty amazing, because you can do all of those different things and you can take ownership of it. That will have helped." (S03)

"Learning about computing makes me feel happy because I can do my own thoughts and things - I don't have to do something in particular." (S02: female pupil)

It's possible that having this ownership was a particularly positive experience as pupils were jointly owning projects: in this sense, it might not have been an entirely 'separate' mechanism to the collaboration pathway.

"My favourite part is Scratch because you can build characters – it doesn't have to be one person's idea. Me and Girl K in the computing lesson today, I said I want to make the Santa speak and Girl K added on to that and said maybe we should say "First stop is Spain!" and we made him talk in Spanish." (S02: female pupil)

Content of the lessons

One experienced teacher suggested that the content of the lessons, rather than the Pair Programming approach, might have particularly appealed to girls. She suggested that by teaching coding within the context of animation, as opposed to a shooting game context which she felt was often how Scratch was taught when it was first used, might have contributed to girls' high engagement with the lessons.

"I don't know whether it [increased girls' engagement] is the Pair Programming or the content. I think it's more that it was about animation, it wasn't about shooting a baddie. I think that suits girls perhaps more. I hate to say it but we all know it tends to be boys that prefer making the games...when we all started with Scratch it was always make a game, shoot people. So you can't tell whether the reason is the girls don't like it because they're doing 'boyish' things or… because it's the programming anyway?" (S02)

This idea was supported by girls' comments that they particularly enjoyed animating and customising the sprites:

"You can go wild and basically make the characters and sprites do anything"(S02: female)

"My favourite part of the computing lesson today was customising the sprite because I love being creative." (S01: female)

d) Specificity of mechanism to girls

The interviewed teachers diverged on whether they felt that the teaching approach intervention had specifically addressed barriers to girls' engagement with computing, or whether the effects were balanced across boys and girls.

Some teachers, whilst acknowledging that the intervention had been designed to target girls' specifically, reported that they thought the Pair Programming approach led to the same outcomes for both boys and girls. These teachers felt that the intervention had increased engagement with computing lessons indiscriminately of gender.

“Nothing stands out in particular. I'm not going to try and kind of conjure something. I'm pleased to say that there's been equal engagement and an equal impact on both male and female." (S01)

Other interviewed teachers did feel that the intervention had particularly appealed to girls, through the mechanisms described in the sections above. One teacher suggested that the building of confidence was particularly relevant to her female pupils, as they had started with lower confidence:

“I think the girls would have come out better from it, because of their confidence towards the subject. The boys, they liked it, but I feel like the girls were more engaged with it. I don't know if I would have seen that level of engagement from the girls, if it wasn't taught that way, because I do think a lot of mine go in on themselves when they don't know.” (S03)

Both boys and girls referred to the opportunities for helping one another which Paired Programming provided, however girls seemed to focus more on the relationship building opportunities, implying they might have engaged more deeply with the collaborative element:

“If someone was your friend you’d make them be more of your friend because you’d be talking with them more, sharing their interest and knowing what they like.“ (S02: female pupil)

There was no evidence from the case study schools that elements of the Pair Programming approach were unappealing to boys. None of the teachers could think of any ways in which the lessons might have harmed boys’ engagement, and one suggested that the mixed gender pairs might have led to increased attainment.

"*Without stereotyping, I think it was good for both [genders], because it was really good for me to assess them all...some boys might have not finished the task if they were on their own, and there would be nothing to show for it." (S03)*

The boys themselves spoke positively about the computing lessons, particularly the problem solving and creative elements. One boy noted that he felt he had to listen to the input for too long, but, as a lengthened main input is not a characteristic of the Pair Programming lessons, this is unlikely to be a concern of this specific programme.

Barriers and facilitators of the mechanisms

This section explores the factors which either acted as barriers to the mechanisms suggested in the logic model leading to the programme’s intended outcomes, or which acted as enablers.

Pupil skill in working in pairs

Where pupils had teamworking skills that enabled them to work together effectively as pairs (e.g. listening to one another, taking it in turns, understanding another's' perspective)**, they were able to engage well with the paired activities.**

"Well, I'm quite lucky, because I've got a really good class; really well behaved, so I don't have any issues there [how well the pupils worked together]." (S04)

Where pupils struggled with some elements of working together, this could have stopped the Pair Programming approach from leading to its intended proximal outcomes of improving girls' attitudes towards computing. For example, paired work might have led to girls experiencing more irritation in lessons as a result of disagreement with their partner. Both teachers and pupils noted that there were sometimes disagreements within the pairs: the theme of having an annoying partner seemed particularly strong amongst the girls' reports of the lessons. Behaviour that could be annoying included the partner being 'silly' e.g. by putting very large numbers into the size box, by not listening to instructions or by getting something 'wrong'.

"*If you have a really good idea and then they put the wrong thing in the wrong place, it's quite annoying." (S01: female)*

Within the lesson observations, there were a handful of instances in which pupils within pairs disagreed with one another. Where the teacher was actively identifying and praising pairs who were working well together (e.g. giving out stickers), there seemed to be less conflict between pairs.

School context

Participating schools seemed to be 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, coding week, displays of female computing role models and computing positions of responsibility given to girls such as 'digital leaders'.

"We did a computer science week and we had some [female] *computer scientists from Microsoft and the government come and talk about their jobs." (S04)*

The interviewed teachers also linked their school's values to the aims of the GBIC project, stating that they had felt the aim of encouraging girls' engagement with computing was in line with their schools' mission and values.

"We do a lot of PSHE [Personal, Social, Health and Economic Education] about aspirations, so that has come out in conversation. It's a very low economic area, so we are very big at aiming for whatever you want to do. Regardless of gender or ability, we want them to aim high, so it [girls being engaged with computing] could be a school thing." (S03)

It is possible that, in these environments where the teachers were 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 which has not been actively trying to improve girls' computing engagement.

Pupils’ original perceptions of computing

There was striking consistency in pupils’ responses that having computing as your favourite subject should not depend on gender. Pupils emphasised respecting one another's opinions and the equal input of computing lessons in schools.

“[The balance of boys and girls who choose computing as their favourite subject should be] equal because we’re learning the same things at the same time and we have the same amount of time and stuff” (S01: female pupil)

“[Boys and girls having computing as their favourite subjects is] equal - you can like the same things. Girls could like dragons and boys can like princesses but it doesn’t matter about if they’re a girl or a boy.” (S02: female pupil)

On the less abstract question of whether there is an imbalance in whether girls or boys spend more time on devices, there was a less strong sense from pupils that this was equal. It’s worth noting that, whilst the overall sense was a focus on equality, a handful of comments did refer to gendered stereotypes (e.g. when asked whether it’s true or false that ‘girls spend more time on electronic devices out of school than boys’, one girl responded with: *“false - girls learn to sew and boys go on devices” S02:female). Where pupils are holding these more gendered views, this could act as a barrier to the mechanism leading all the way to change in girls’ intention to study computing, even if the lessons themselves can increase girls’ perception of computing as a collaborative subject.

Pupils’ computing attainment

Across the case study schools, teachers noted that pupils who had previously been lower attaining in computing tended to particularly benefit from the Pair Programming lessons. Interviewed teachers suggested this could have been because the support from a partner meant that the pupils were able to access the tasks, in a way which would not have been possible, had they been working independently.

“*I tried to put them mixed abilities - have someone that's a little bit higher than them, help them out, explaining to them, giving them vocab. They found it a bit easier, because then there was another person navigating them through their problems" (S04)*

The age of the pupils

One teacher suggested that ‘sewing a seed’ at this point could lead to greater engagement once the pupils are older. It’s possible that, whilst we did not see a change in girls’ attitudes to computing at the end of the intervention, the experience might contribute to longer term changes in attitudes, potentially interacting with future computing experiences.

“I think anything, as early as you can, to get them [girls] involved, even if they don’t take an interest now, it’s still implanted and that learning is in their brains, so it might be an interest that they take up later down the line.” (S04)

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 conclusive evidence that the intervention increased girls' attitudes towards computing or intention to study computing in the future. While the estimated effect of the intervention on each outcome was positive, it was not statistically significant and small in magnitude. Issues related to the limited sample size, differential attrition and baseline imbalances in both outcome measures introduce some risk of bias in the results, which should thus be interpreted cautiously. However, given the small size of the differences between the two groups and that the potential bias caused by attrition is more likely to result in an over- rather than under-estimation of the impact of the intervention, there is no evidence indicating that the low sample size or potential biases are obscuring a substantive positive impact of the intervention.

The intervention was implemented well at IPE case study schools, though delivery challenges were identified

In the four case study schools which we collected data for, the Pair Programming intervention was implemented well, with high fidelity to the paired approach and the Scratch activities which children completed within the lessons. 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 and the small number who agreed to give feedback are unlikely to represent the full breadth of experience of the teachers and pupils in the treatment group. Adaptations made to the number of lessons taught and content of the lessons was mainly due to timetabling constraints, and was done in a way which preserved the central focus on collaboration.

The three main implementation challenges identified were: 1) including all of the content provided within the time constraints of a lesson; 2) pupil disagreement within pairs; and 3) the subject knowledge of less experienced teachers. The intervention is currently easiest to implement in schools in which the teacher delivering the lessons is confident in their computing subject knowledge, the pupils are skilled in working in pairs, and where there is broader support from the senior leadership team.

Multiple factors could be contributing to the lack of evidence of impact despite positive reported school experiences

Whilst the implementation and process evaluation did identify mechanisms through which the intervention might have led to the intended effect on girls' attitudes towards computing, no evidence of this was found in the impact evaluation based on the pupil surveys. There are a number of possible reasons for this contrast between the qualitative and quantitative findings related to the design of the intervention, its delivery, and measurement challenges:

- **Dosage:** COVID-19 challenges could have prevented the intervention from being fully implemented. It is possible that the 'dose' of the intervention was reduced to such an extent that it didn't have a discernible effect. Without having implementation information on all of the schools, it's difficult to estimate the extent to which lack of or uneven implementation across schools could have contributed to our findings.

- **Online nature of training constraining impact:** As the training had to be delivered online, RPF were not able to complete all of the training activities that they had intended to do with the teachers, for example modelling the Pair Programming element of the lesson. This lack of in-person training might have led to more variation in how teachers implemented the intervention, compared to more consistent implementation which might have resulted from in-person training.

- **Sensitivity of outcome measures:** As discussed in section 2.1.6, the nature and objectives of the intervention implied challenges in reliably measuring the intended outcomes, which created a need to rely on short-term and pupil-reported proxy indicators that may not capture the full and longer-term impact of the intervention. While strategies were implemented as part of the evaluation to mitigate these challenges, they are unlikely to fully address them, and some teachers reported pupil challenges in understanding and completing the SCSAS. This may have led to some measurement error and limited the ability to measure the constructs precisely enough to identify variation between the treatment and control group.

- **High engagement with computing at baseline:** Due to the characteristics of the sample (schools which contain a teacher interested in taking part in a gender balance in computing project), the sample could have had particularly high baseline engagement with computing, leaving less scope for improvement as a result of the intervention and making it more difficult for the evaluation to detect an impact on these outcomes. The high baseline SCSAS score and stated intention to study computing are consistent with this hypothesis.

- **Hypothesised barrier not most critical to intended outcome:** It is possible that the intervention did indeed lead to greater collaboration in computing, but that lifting the barrier of limited collaboration is insufficient to meaningfully improve girls' attitudes towards computing or intention to study it in future. If this causal link does not exist (between collaboration in computing and girls' attitudes towards computing) then an intervention which increases collaboration in computing would not causally affect girls' attitudes towards computing or intention to continue with the subject.

- **Changes in intended outcomes are not observable immediately after the intervention:** Given the measurement challenges inherent to this intervention, it is possible that this unit of Pair Programming lessons set off a chain of mechanisms which will eventually contribute to improved attitudes towards computing or increased intention to study computing, but that changes in these outcomes were not yet apparent when measured immediately after the intervention has finished.

There is not sufficient evidence to confidently determine whether and how these factors may have contributed to the evaluation results.

5.2 Recommendations

Recommendations to support implementation

The intervention could be made easier to implement in a broader range of schools through the following adaptations of the intervention in response to the main implementation challenges identified:

1. **Cut down content within each lesson plan**
   The starter and end-of lesson activities which teachers did not always have time to complete could either be removed, or it could be made clear in lesson plans that it is acceptable for teachers to remove content from the plans themselves. The paper-based starters could be replaced by similar activities which can be done as a whole class through the power-point.

2. **Provide additional support for teachers who lack confidence teaching these National Curriculum objectives**
   Training videos could be made to demonstrate how to create the code to complete certain projects, or how to deal with frequently faced challenges. The arrangement of links to different parts of the training and resources could also be simplified. In future, face-to-face training might provide opportunities for support to be more personalised based on teacher confidence.

These changes could make the Pair Programming intervention easier to implement for a broader range of schools. Whilst they may not fully tackle the identified barriers of 1) rushed lesson times for teachers and 2) lower confidence and subject knowledge in non-specialist computing teachers, they could increase the potential for the Pair Programming lessons to improve girls' attitudes towards computing and intention to study computing in the future.

Recommendations for future use of intervention resources

Given that the interviewed teachers reported that delivering the Pair Programming lessons had been a positive experience for both their pupils and themselves, and that they felt the lessons increased pupils' enjoyment of computing and collaboration, it would be valuable to:

1. **Offer the lesson resources and training to all KS2 teachers (and teaching assistants) who would like to use the resources in their lessons**
   As the units cover curriculum content that must be taught in all primary schools, we do not foresee any significant opportunity costs to teachers using these lessons. This is because the developers of these materials are likely to have more computing expertise than most non-specialist primary class teachers who might be expected to produce resources from scratch. Given that the majority of teachers delivering computing lessons in KS2 are unlikely to have a subject specialism in computing, it's plausible that they would benefit from resources which are both a) developed by subject specialists and b) produced in a way that is accessible for non-specialist teachers.

As teachers reported that they were keen to try the Pair Programming approach with other groups of pupils and whilst teaching different topics, it could be helpful for RPF to:

2. **Make general Pair Programming resources available**
   RPF could make general Pair Programming resources (non specific to topics within computing) available to teachers, who could then integrate the approach into other lesson plans.

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 COVID-19 context, in addition to the challenges often associated with evaluating school-based interventions and attrition in particular. 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 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. **Identify strategies to measure outcomes targeted by the intervention further into the future**
   Tracking relevant behavioural outcomes (in this case, subject choice from Year 10 onwards) multiple years after the intervention requires planning, collaboration with schools, and longer evaluation timelines. However, it would also greatly enhance the ability to evaluate the impact of early interventions over a time horizon in line with the mechanisms and barriers hypothesised, and thus identify the most impactful ones. In this case, attempting to collect and analyse data on whether pupils in the evaluation sample select computer science as a GCSE subject once the choice arises would enable the estimation of the impact of the intervention on the long-term outcomes targeted, 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 Computing 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, though this would be easier to achieve for interventions targeting older pupils closer to their GCSE subject selection. 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.

Reference list

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.

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.

Haynie, K.C. and Packman, S. (2017). *AP CS Principles Phase II: Broadening Participation in Computer Science Final Evaluation Report.* Prepared for The College Board and the National Science Foundation, February 12, 2017. Skillman, NJ.

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

McDowell, C., Werner, L., Bullock, H. E., & Fernald, J. (2006). Pair programming improves student retention, confidence, and program quality. *Communications of the ACM, 49(8), 90-95.*

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

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American educator*, *36*(1), 12.

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.

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: Survey measures

Baseline survey

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?			
	Computing	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]

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

2.1	I have self-confidence when it comes to 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 am good at learning computing skills on my own	Strongly disagree	Disagree	Agree	Strongly Agree
2.4	I am good at doing hard computing work	Strongly disagree	Disagree	Agree	Strongly Agree
2.5	I think I will do well in computing	Strongly disagree	Disagree	Agree	Strongly Agree

Page 3
How much do you agree or disagree with the following statements? There are no right or wrong answers.

2.6	I would take more computing lessons if I could	Strongly disagree	Disagree	Agree	Strongly Agree
2.7	In the future I'd like to do a job where I use 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	I like the challenge of using computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.10	I like writing computer programs	Strongly disagree	Disagree	Agree	Strongly Agree

Page 4

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

2.11	I feel happy in computing class	Strongly disagree	Disagree	Agree	Strongly Agree
2.12	Computing is for people like me	Strongly disagree	Disagree	Agree	Strongly Agree
2.13	I get on with the people in my computing class	Strongly disagree	Disagree	Agree	Strongly Agree
2.14	I know someone like me who uses computing in their work	Strongly disagree	Disagree	Agree	Strongly Agree
2.15	I know lots of people like me who think computing is interesting	Strongly disagree	Disagree	Agree	Strongly Agree

Page 5

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

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	Computing skills I use for class work can help me understand things I do in other classes and at home	Strongly disagree	Disagree	Agree	Strongly Agree
2.19	I'll need very good computing skills for my future job	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 like me said I should do computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.22	Someone I know has made me want to learn 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	Someone I know has talked with me about computing jobs	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

Endline survey

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
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		
1.6	When you do computing lessons, do you usually work on your own or in pairs?	On my own	In a pair	
1.7	Do you want to study any of these subjects in future?			
	Computing	Yes	No	Don't know
	Science	Yes	No	Don't know
	Technology	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]

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

2.1	I have self-confidence when it comes to 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 am good at learning computing skills on my own	Strongly disagree	Disagree	Agree	Strongly Agree
2.4	I am good at doing hard computing work	Strongly disagree	Disagree	Agree	Strongly Agree
2.5	I think I will do well in computing	Strongly disagree	Disagree	Agree	Strongly Agree

Page 3
How much do you agree or disagree with the following statements? There are no right or wrong answers.

2.6	I would take more computing lessons if I could	Strongly disagree	Disagree	Agree	Strongly Agree
2.7	In the future I'd like to do a job where I use 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	I like the challenge of using computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.10	I like writing computer programs	Strongly disagree	Disagree	Agree	Strongly Agree

Page 4

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

2.11	I feel happy in computing class	Strongly disagree	Disagree	Agree	Strongly Agree
2.12	Computing is for people like me	Strongly disagree	Disagree	Agree	Strongly Agree
2.13	I get on with the people in my computing class	Strongly disagree	Disagree	Agree	Strongly Agree
2.14	I know someone like me who uses computing in their work	Strongly disagree	Disagree	Agree	Strongly Agree
2.15	I know lots of people like me who think computing is interesting	Strongly disagree	Disagree	Agree	Strongly Agree

Page 5

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

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	Computing skills I use for class work can help me understand things I do in other classes and at home	Strongly disagree	Disagree	Agree	Strongly Agree
2.19	I'll need very good computing skills for my future job	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 like me said I should do computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.22	Someone I know has made me want to learn 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	Someone I know has talked with me about computing jobs	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

Appendix 2: Model specification

Primary outcome: SCSAS scores

The primary 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, and because we randomised at the cluster level, 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*, = 1 if pupil *i* attends a treatment school
- 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* (using "Outstanding" as a , comprising (i) "Good"; (ii) "Below good" (the combination of "Requires improvement" and "Inadequate"); and (iii) Missing Ofsted rating/No rating available;
- 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: OLS regression coefficients for primary outcome (standard errors in parentheses)

Outcome: SCSAS score	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Treatment group (reference category is control)			
Intervention	0.051 (0.047)	0.030 (0.045)	0.059 (0.046)
Baseline SCSAS score	0.596** (0.057)	0.630** (0.037)	0.624** (0.038)
Missing Baseline SCSAS	-	1.699** (0.116)	-
Ofsted rating *(reference category is Outstanding)*			
Good	0.078 (0.061)	0.060 (0.056)	0.109* (0.051)
Below Good	-0.028 (0.104)	-0.053 (0.102)	-0.030 (0.111)
Missing	0.155+ (0.084)	0.147+ (0.082)	0.194* (0.076)
Percentage FSM	0.000 (0.002)	-0.001 (0.002)	0.001 (0.002)
Constant	1.157** (0.174)	1.094** (0.119)	1.033** (0.117)
Control group mean	2.80	2.80	2.77
Observations	990	990	785
R^2		0.284	0.369

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 binary, and therefore we will use 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 will use cluster-robust standard errors in analysis, clustering at the school level.

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

Where:

- Y_{is} is a binary indicator for pupil *i* reflecting intention to study computing 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,* = 1 if pupil *i* attends a treatment school
- 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*
- $Ofsted_s$ s a tertiary indicator of Ofsted rating in school *s* (using “Outstanding” as a , comprising (i) “Good”; (ii) “Below good” (the combination of “Requires improvement” and “Inadequate”); and (iii) Missing Ofsted rating/No rating available.

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: Logistic regression coefficients for secondary outcome (standard errors in parentheses)

Outcome: SCSAS score	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Treatment group (reference category is control)			
Intervention	0.179 (0.239)	0.132 (0.232)	0.211 (0.258)
Baseline SCSAS score	1.649** (0.215)	1.793** (0.175)	1.803** (0.179)
Missing Baseline SCSAS	-	4.999** (0.513)	-
Ofsted rating *(reference category is Outstanding)*			
Good	0.381 (0.282)	0.325 (0.271)	0.485+ (0.264)
Below Good	0.394 (0.345)	0.336 (0.335)	0.599* (0.289)
Missing	-0.181 (0.308)	-0.219 (0.293)	-0.040 (0.287)
Percentage FSM	0.017+ (0.009)	0.015+ (0.009)	0.018* (0.009)
Constant	-5.258** (0.739)	-5.582** (0.612)	-5.858** (0.638)
Control group mean	0.450	0.450	0.428
Observations	994	995	789
R^2		0.103	0.132

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

Appendix 3: Online teacher training feedback survey

Thank you for completing the online training course for the Pair Programming Teaching Approach project. Please answer the following questions, which should take around 5 minutes, to confirm you have completed the training and provide some feedback.

Your data will be processed in line with the National Centre for Computing Education's privacy policy (https://teachcomputing.org/privacy). If you have any queries about how your data will be used, please contact the STEM Learning Data Protection Officer at datasecurity@stem.org.uk.

Data processing

▢ I understand that my data will be processed in accordance with the National Centre for Computing Education's privacy policy, which contains information about how my personal data is used.

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 2 pupils?

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

Please confirm that you have completed the online training for the Pair Programming 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