Gender Balance in Computing

Evaluation of the Relevance Intervention

July 2022

Steve Human, Rachel O'Boyle and Clément Bisserbe

THE
BEHAVIOURAL
INSIGHTS
TEAM

Contents

Executive summary

Overview of the project

The Gender Balance in Computing Project (GBIC) 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 "Relevance" strand of the programme, where the aim was to evaluate whether exposing female pupils to real world applications of computer science can increase the sense of relevance of computing and make female pupils more likely to choose computer science as a GCSE subject. This intervention with year 8 pupils lasted for up to 13 weeks and covered two half terms of computing lessons.

Evaluation approach

The intervention was evaluated using a mixed-methods approach. The impact evaluation investigated whether there was evidence that the intervention affected female pupils' intention to study computer science at GCSE level and female pupils' attitudes towards computing. In parallel, an 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 taught their usual computing lessons for a period of up to 13 weeks from January to April 2022.
2. Treatment schools received training materials and class plans to deliver during computing lessons over the same period.

The primary outcome, female pupils' stated intention to select computer science as a GCSE subject, was measured through a survey. As secondary outcomes, we measured attitudes toward computing using questions derived from the Student Computer Science Attitudes Survey (SCSAS[1]), a 5-scale tool for assessing attitudes toward computing for school pupils. The two secondary outcomes were the Usefulness subscale and the Interest subscale from the SCSAS, and the overall SCSAS score was an exploratory outcome. In total, 167 secondary schools were originally recruited to participate in the trial and 97 submitted endline surveys (45 on the control group and 52 in the treatment group).

---

[1] 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.

Implementation and process evaluation

The mixed-methods IPE aimed to answer the following research questions:

1. How do the teachers experience the training organised by RPF?
2. How are the lessons delivered and in what ways?
3. How did the pupils experience the intervention?
4. Did the intervention address the mechanisms anticipated to lead to the targeted behaviour?

BIT researchers spoke with teachers from five 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 female pupils. We visited four of these schools, during which we conducted group discussions and a worksheet task with pupils as well as 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 schools and teachers in the sample limits the breadth of experience which can be reported in the IPE. We also explored the responses to a teacher survey which provided feedback on the intervention and how they implemented it.

Key findings

Evidence of impact

Our main analysis did not find statistically significant evidence of impact. For our primary outcome variable, the estimated treatment effect is positive, but small (0.2 percentage points) in magnitude and not statistically significant, with this result holding across different specifications. We conducted robustness checks to test the sensitivity of this result to the way pupils were asked about their intention to select GCSE computer science. We found that when using an alternative outcome based on a slightly different survey question, the treatment effect estimate is larger (2.9 percentage points) and is statistically significant for some, but not all, of the analysis specifications.

For our two secondary outcomes (perceived relevance of and interest in computing) and exploratory outcome (overall attitudes towards computing), the estimated treatment effect is positive, but it is small and not statistically significant. The differential attrition observed across the two treatment groups led to a small imbalance in SCSAS scores at baseline. This differential attrition introduces a risk of bias in the estimates for all outcomes, though it is difficult to determine the possible direction of this bias.

The lack of recollection of participating in a course that involved designing, building and testing an app reported in the endline survey by many pupils across treatment schools suggests that not all pupils who completed the endline survey received the intervention as planned. If this is the case, this may have limited the impact of the intervention across the full treatment sample.

Findings regarding intervention design and delivery

The IPE findings suggest that the intervention was implemented well, though there was deviation from the suggested session activities and several challenges were identified.

Teachers were able to deliver the majority of the lessons, but varied in their use of the suggested session activities and adjustments to the resources. The two main implementation challenges were (1) delivering all the 11 core lessons and (2) delivering all the session activities within the time constraints of a lesson. Other factors such as volume of content to deliver per session, teachers' programming subject knowledge and pupil computing ability and language affected the ease with which teachers could implement the intervention. It is worth noting that given the small number of case study schools, and that teachers who agreed to be involved in case study activities may have been more likely to have implemented the intervention with high fidelity, these teachers are unlikely to be representative of teachers across all treatment schools.

Teachers reported that they had enjoyed delivering the sessions to their pupils and would recommend the intervention to other teachers. Some intended to repeat the intervention or reuse the resources next year with slight adjustments. Pupil engagement with the lessons was high, particularly for practical elements, such as the app building and design. Teachers thought that the resources were generally useful, however highlighted a couple of issues related with online usability and collaborative work. The majority of teachers completed the online training and thought it prepared them well.

We observed some of the hypothesised mechanisms through case study activities, including: problem selection based on interest, applicability or concern; a sense of agency from selecting their own idea; and an increased perception of computing as creative. Other envisioned mechanisms of impact were only apparent for some pupils or not observed, such as perceiving computing as relevant to their lives. Teaching style, the idea selection process and app examples may have contributed to the absence of some female pupils' ability to connect computing to real world issues or problems they personally care about.

Implications

There are a number of possible reasons related to the design of the intervention, its delivery, and measurement challenges, for the absence of evidence of a clear impact despite the largely positive experiences of the intervention reported in case study schools. These include:

- Delivery challenges, such as insufficient time to cover all lessons and all of their content, could have prevented the full implementation of the intervention at all treatment schools. The indicative evidence that many pupils in treatment schools reported not taking part in a course that involved designing, building and testing an app is consistent with this possibility.
- There were some indications at case study schools that while the intervention increased engagement and enjoyment of computing among female pupils, particularly through highlighting its creative elements, this did not fully translate in an increase in the perceived relevance of computing to their lives, and the envisioned mechanisms to increase perceived relevance were not all realised. If this was the case at other treatment schools, this may have limited the impact of the intervention.
- Addressing the hypothesised barrier of computing not being perceived as relevant to solving real-world problems and to females' lives may not be sufficient to overcome

the large and structural barriers which prevent females from choosing GCSE computer science. These barriers are reflected in the much lower proportion of female pupils reporting the intention to study GCSE computer science in the baseline survey (7% of female vs. 22% of male pupils).

- Survey measurement challenges, while mitigated by various survey design strategies and robustness checks in the analysis, introduce a risk of some measurement error that could limit the evaluation's ability to detect differences between the treatment and control group.

Recommendations

Despite the lack of clear evidence of a significant impact, some of the findings from the evaluation suggest that this intervention holds some promise as a relatively low-cost strategy to positively affect female pupils' attitudes towards computing and intention to select computer science as a GCSE subject.

In light of these results, and that teachers reported finding the resources useful and the intention to use them again, it could be valuable to offer the lesson resources and training to KS3 teachers who would like to use them. Given the limited impact of this intervention alone suggested by the impact evaluation results, teachers could also be encouraged to adapt and combine these resources with other approaches which they think could increase female interest in computing within the context of their school. While getting schools currently less motivated or dedicating less resources to encourage female pupils to study GCSE computer science to use these resources may be more challenging than already motivated schools, the potential impact of the intervention resources may be greater as well.

However, to maximise pupil understanding, pupil engagement and ease of delivering the intervention, the following recommendations should first be taken into account:

1. Make resources easier to use online and collaboratively
2. Reduce the number of lessons and add 'optional/core' labels to teacher slides, to help teachers select the most important lessons or slides if under time pressure
3. Work with teachers and/or pupils to refine the intervention such that the app examples are more relevant and concepts are easier to understand
4. Allocate more time to practical and creative elements in lesson plans/ resources
5. Make teacher access to additional support easier and more interactive

Addressing these adjustments to the design and delivery of the intervention could help to improve the effectiveness of the Relevance intervention. In addition, using school administrative data to measure GCSE subject selection and tracking whether schools are able to implement the intervention as planned could help to reduce the cost of data collection and understand the longer-term impact of the intervention. Conducting another impact evaluation after the adaptations to the intervention and with these changes to the evaluation strategy could build the evidence for this approach. If the impact among schools that signed up voluntarily is then found to be positive, this could inform decisions regarding a more proactive scale-up of the intervention to a large number of schools.

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[2]. Across England, only 21% of the GCSE computer science cohort is made up of female students[3]: 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[4] [5]. The Gender Balance in Computing Project (GBIC) aims to tackle a number of known and well-researched barriers to female pupils engaging with computing[6], including a disconnect between extra-curricular computing activities and subject choice; a lack of encouragement to study computing; a lack of familial and other role models in computing and a perceived lack of relevance of computing to pupils. These barriers are addressed in the five intervention strands that comprise GBIC, with the common goal of increasing the number of female pupils who study GCSE and A Level computer science.

1.2 Relevance approach

One of the hypothesised barriers preventing women from entering into computer science and STEM careers is the lack of perceived relevance of computing to one's life. For instance, research shows females choose not to enter into the computer sciences through job-related beliefs like the belief there are few jobs available in the sector or a limited knowledge of the range of jobs they could pursue.[7] Moreover, there is evidence that suggests there is an increase in female pupils' motivation to study computing when it is linked to their lives or to a real-world impact.[8,9]

Research in the United States has suggested that female computing undergraduates are more likely to have goals related to helping others compared to their male peers, and that

[2] Royal Society. (2017) After the reboot: computing education in UK schools.
https://royalsociety.org/~/media/policy/projects/computing-education/computing-education-report.pdf
[3] 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
[4] 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
[5] 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..
[6] 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
[7] Croasdell, D., McLeod, A., & Simkin, M. G. (2011). Why don't more women major in information systems?. Information Technology & People.
[8] Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. science, 326(5958), 1410-1412.
[9] Kemp, P.E.J., Berry, M.G., & Wong, B. (2017) The Roehampton Annual Computing Education Report. University of Roehampton, London.

they do not always recognise the opportunities that computer science affords to meet these goals.[10] Further evidence indicates that by increasing the perceived relevance of computing to pupils' lives, engagement of female undergraduates with the subject can be improved.[11]

The Relevance intervention

This intervention builds on the aforementioned evidence by intending to increase female pupils' sense of relevance of computing to real-world problems, which in turn is hypothesised to improve their attitudes to computing and interest in studying computing. Throughout the intervention teachers delivered sessions that aimed to support pupils to identify real-world problems and to then develop apps that provide a solution. In doing so, the intervention aims to use tangible examples to increase female pupils' awareness of the real-world applications of computing and the specific computing skills that are applicable to causes they care about. The intention is that this increased comprehension of the field's relevance will improve attitudes towards computing and motivate female pupils to select computer science at GCSE level. The logic model for this intervention is depicted in Figure 1.

The programme is designed to be delivered to year 8 (age 12-13) pupils for up to 13 weeks.[12] Intervention materials consist of training for computing teachers and session materials, which teachers are invited to adapt as needed. Over the course of the programme, pupils were taught how to research and identify real-world problems that can be addressed with apps as well as how to program an app in App Lab.

A pilot of the intervention was conducted between April and August 2021 at three schools. This pilot generated findings and recommendations about the delivery of the teacher training and lesson plans, based on which the intervention resources and training were refined.

---

[10] Lewis, C., Bruno, P., Raygoza, J., & Wang, J. (2019, July). Alignment of goals and perceptions of computing predicts students' sense of belonging in computing. In Proceedings of the 2019 ACM Conference on International Computing Education Research (pp. 11-19).

[11] Guzdial, M., & Tew, A. E. (2006). Imagineering inauthentic legitimate peripheral participation: an instructional design approach for motivating computing education. In *Proceedings of the second international workshop on Computing education research* (pp. 51-58).

[12] The majority of pupils were between seven and ten months away from choosing their GCSE subjects at the end of this period.
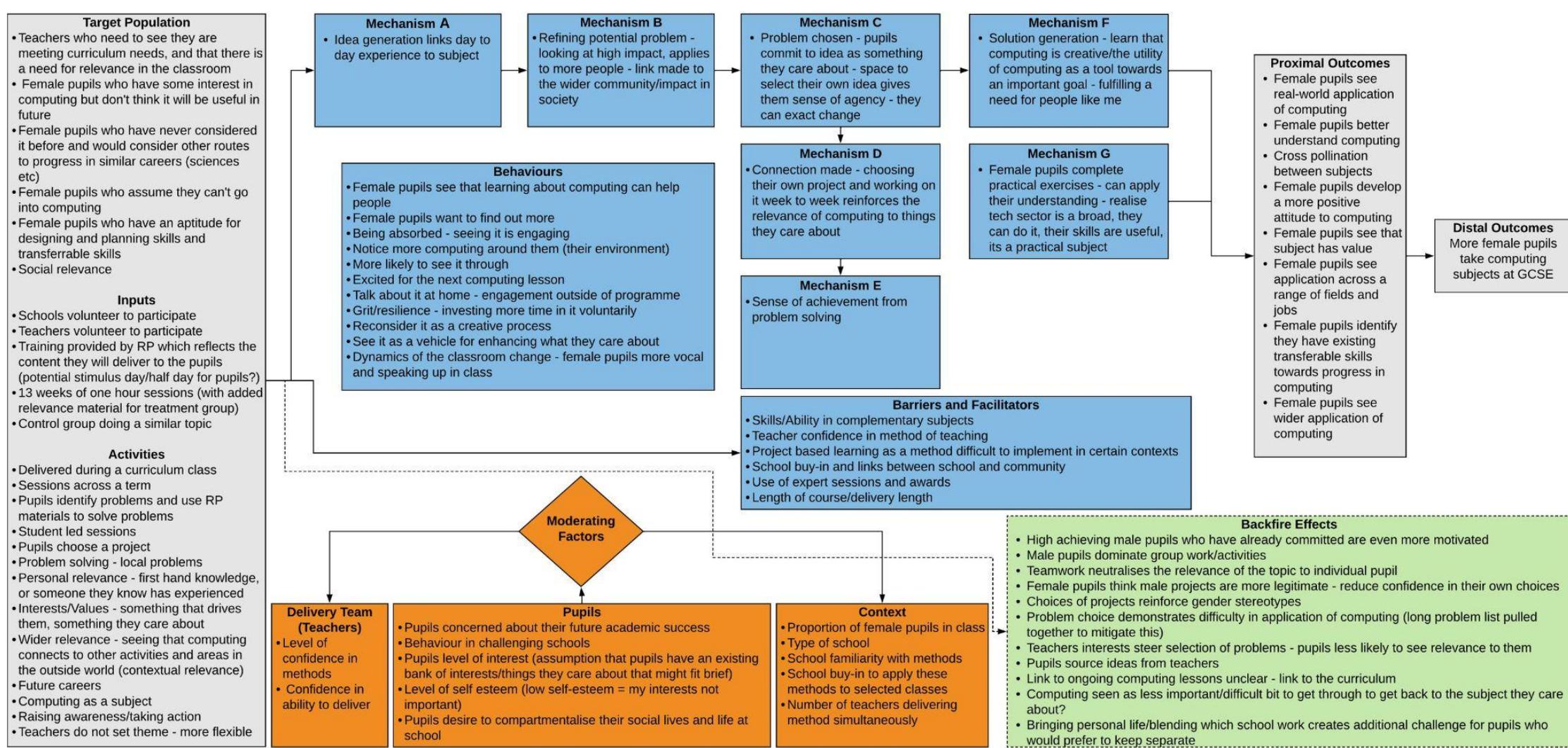
Figure 1: Logic model of the Relevance intervention

Online training

Teachers received online training on the material and how to deliver it. The online training for teachers consisted of three online tutorials for teachers to complete independently. "Part 1: Year 8 'Relevance' training – Introduction" provided an introduction to the GBIC programme and associated research, an overview of the two strands included in the 'Relevance trial' ('Identifying real-world problems that can be addressed with apps' and 'Programming an app in App Lab') and how data will be collected via an online survey. "Part 2: Year 8 Relevance' training – Getting started" provided an overview of the sessions, resources and two tasks for teachers to complete: one on app idea development and another on paper design and assets. "Part 3: Year 8 'Relevance' training – Investigating App Lab" aimed to help teachers to familiarise themselves with App Lab, the platform which pupils use to build their app on.

Sessions

The teachers were provided with the resources to deliver 13 hour-long lessons, of which two were optional (see Table 1). Each lesson consists of session activities which totalled one hour and ranged from teacher-led PowerPoint presentations to pupil-led practical activities.

Table 1: Summary of lessons

Lesson title	Objective
Technology in your life	Pupils understand the impact of technology on your life and the wider world
Getting to know App Lab	Explore App Lab's online interface and create screens for a sample app
Event-driven programming	Develop app for user input in an event-driven programming environment
Develop functionality through selection	Develop functionality of an app by creating selective code logic (code run only if particular conditions are met)
Changing your world with technology	Explain how apps can help address real-world challenges
Deciding which problem to solve	Be able to critically evaluate app ideas and select the strongest to take forward
Industry engagement (optional)	Pitch your app to an industry expert and obtain constructive feedback on the key concepts of your app
Understanding user needs	Explore user needs and features in mobile app design
UX design	Understand how the design of your app can affect the user experience
Build your app	Pupils start to build a technical prototype of an app using App Lab
Continue to build your app	Pupils continue to build a technical prototype of their app using App Lab[13]
Finish building your app (optional)	Pupils complete the technical prototype of their app using App Lab
Test and evaluate your app	Pupils plan and conduct user tests to evaluate how users interact with their app and refine their app based on findings

---

[13] This session is optional and only needed if the pupils require more time to build the app.

Resources

Teachers were provided with several resources. These included: (1) teacher slides for teachers to present to pupils at the start of the lesson, (2) lesson primers which provide guidance on how to teach the lesson, (3) lesson plans which provide a five-minute visual lesson plan, (4) a pupil checklist which is a one-page table where pupils can tick off completed tasks and (5) a pupil workbook and worksheets for pupils to document their learning. All of the resources are hosted on the project website, and schools received links to access the website.

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 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.

2. Methods

The evaluation used a mixed-methods approach. We conducted a randomised control trial (RCT) design and randomised at a school level. Quantitative data was collected via online surveys distributed pre and post intervention in both treatment and control schools to be completed by the year 8 pupils as part of their computing lessons. An implementation and process evaluation (IPE), which aimed to explore the mechanisms of change and complement the impact evaluation findings, was also carried out. This section describes the research questions, methods used and the limitations of this evaluation approach.

The sampling strategy and data collection methods and processes were informed by a pilot conducted in three schools from April to August 2021.

2.1 Impact evaluation

We used a cluster randomised RCT, with randomisation at the school level and outcome measurement at the pupil level.

2.1.1 Research questions and outcome measures

The ultimate aim of this intervention is to increase the number of female pupils who study GCSE computer science. Therefore, the impact evaluation aimed to determine whether the intervention led to a change in:

1. *Primary outcome:* female pupil's stated intention to study computer science at GCSE level
2. *Secondary outcome*: female pupil's perceptions of the relevance of computing, as measured by the Usefulness subscale of the Student computer science Attitude Survey (SCSAS)
3. *Secondary outcomes*: female pupil's interest in computing as measured by the Interest subscale of the SCSAS.
4. *Exploratory outcome*: female pupil's overall attitudes towards computing as measured by their total SCSAS score.

All of the impact evaluation outcomes were measured through the pupil online survey. We used the Student computer science Attitudes Survey (SCSAS) to measure female pupil's attitudes towards computing (Haynie and Packman, 2017). The SCSAS is composed of five distinct subscales that represent different facets of attitudes towards computing as a subject and the perceived benefits of doing so, as well as the experience of computing lessons. It contains 25 questions across the 5 subscales (5 questions per subscale). These subscales are: Confidence, Interest, Belonging, Usefulness and Encouragement. Within each subscale,

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. The individual questions that make up these subscales can be found in Appendix 1.

The Usefulness and Interest subscales were used to measure the two secondary outcomes, and the overall SCSAS score was used to measure the exploratory outcome. Table 2 outlines the data that was collected to measure these outcomes listed above, and when it was collected.

Stated intention to select GCSE computer science was selected as the primary outcome as it is the most direct measure of the behaviour the intervention sought to influence[14]. The score for the Usefulness and Interest subscales of the SCSAS were selected as secondary outcomes. These subscales measure the proximal outcomes targeted by the intervention more directly than the overall SCSAS score (see the intervention logic model in Figure 1), and reflect the hypothesised pathways to impact on female pupils' choice to study GCSE computer science. We also examined the impact of the intervention on the overall SCSAS score as exploratory analysis.

Table 2: Method for collecting data on outcome measures

Outcome measures	Data collected	Point of collection
Primary: Intention to select GCSE computer science	Single item survey measure of whether the pupil plans to continue studying computing with possible responses "Yes", "No", or "I don't know"	Online surveys, completed on computers in class at baseline (Jan-Feb 2022) & immediately following the culmination of the programme (Apr-May 2022). These surveys were 'built in' to the first and last lesson plans of the 13-week intervention [15].
Secondary: Perception of relevance of computing	Usefulness score on the SCSAS	
Secondary: Interest in computing	Interest score on the SCSAS	
Exploratory: General attitudes towards computing	Overall score on the SCSAS	

2.1.2 Sampling and randomisation

The impact evaluation was designed as a two-armed cluster RCT, 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 between January 2022 and April 2022.

---

[14] We also explored the possibility of collecting actual GCSE subject choice data to validate this measure, but determined this was not possible given the constraints of this evaluation.
[15] Some schools completed the endline survey less than 13 weeks after completing the baseline survey, suggesting that not all schools completed the intervention within the planned 13-week period.

2. Treatment: In January 2022 schools in this group received online training and session resources, such as lesson plans and teacher slides, to enable them to deliver the Relevance intervention over a period of up to 13 weeks ending in April 2022. The resources and lessons delivered are outlined in section 1.2.

Recruitment of schools was conducted by RPF. All secondary schools in England that stated they could offer the full 13-week programme were eligible for the trial. Schools were also required to have female pupils (all-boys schools were excluded).

All schools that entered the sample did so voluntarily; this has implications for the external validity of the findings, as schools that volunteer are likely to be more interested in improving gender balance in computing than the average school, and may thus not be representative of all schools in England. Another factor that might limit external validity is that schools also had to agree to participate in the evaluation, which not all schools might be interested in doing even if they would generally be interested in the programme.

Randomisation was conducted by BIT, using school unique reference numbers (URNs) as unique identifiers. Schools were stratified by (1) overall Ofsted ratings, using four categories: "Outstanding", "Good", a combination of "Requires improvement" and "Inadequate" (this is to reduce the number of strata and also reflects the low number of schools rated "Inadequate") and a category for missing ratings; (2) Research Status[16] (binarised into accepted and pending).

We initially planned to stratify on the percentage of free school meals (FSM) status but chose not to in order to limit the number of strata. We conducted balance checks post randomisation and found this variable to be balanced. Groups were also found to be balanced in terms of the proportion of schools that were female pupils only.

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

The intervention was offered to both female pupils and male pupils and data was thus collected for both, but only data from female pupils was analysed for primary, secondary and exploratory analyses given the objectives of the intervention. Data collected from male pupils was included in the descriptive statistics and to contrast with female pupil's data (see next section).

2.1.3 Description of data

Table 3 presents the mean scores for each SCSAS subscale at baseline, combined across treatment and control groups and split by gender.

---

[16] Not all schools that had registered interest in the trial had confirmed their participation by the point at which schools were due to be randomised, thus this was included in the stratification to maximise the chances of equal sample sizes across treatment arms after all schools had confirmed their decision to participate.

Table 3: Pupil baseline survey data by gender (primary and secondary outcome indicators emboldened)

Outcome	Values	Gender	N (non-missing)[17]	Mean (SD)
Intention to study computer science at GCSE	1 = "Yes" 0 = "No", "Don't know"	Female	5,175	0.07 (0.26)
		Male	4,305	0.22 (0.41)
Total SCSAS score	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Female	5,175	2.47 (0.48)
		Male	4,305	2.74 (0.52)
SCSAS: Confidence subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Female	5,175	2.53 (0.56)
		Male	4,305	2.78 (0.59)
SCSAS: Interest subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Female	5,162	2.32 (0.64)
		Male	4,291	2.72 (0.69)
SCSAS: Belonging subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Female	5,151	2.65 (0.52)
		Male	4,275	2.89 (0.56)
SCSAS: Usefulness subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Female	5,128	2.61 (0.59)
		Male	4,251	2.86 (0.60)
SCSAS: Encouragement subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Female	5,104	2.24 (0.64)
		Male	4,233	2.47 (0.70)

Table 3 shows that at baseline, stated intent to study GCSE computer science was higher for male pupils than female pupils (22% of male pupils indicating they would like to study computer science as a subject for their GCSEs, compared to 7% of female pupils). This baseline rate for female pupils is close to an estimate provided by the DfE of 6% and similar to the baseline rate of 7% for the sample for the other GBIC evaluation targeting year 8 pupils ('Teaching Approach - Peer instruction').

Table 3 also shows that at baseline, male pupils scored higher than female pupils in the Usefulness (male pupils' mean = 2.86 out of 4, female pupils' mean = 2.61 out of 4), and Interest (male pupils' mean = 2.72 out of 4, female pupils' mean = 2.32 out of 4) subscales, as well as in overall attitudes toward computing (male pupils' mean = 2.74 out of 4, female pupils' mean = 2.47 out of 4)[18]. Table 3 illustrates that we observe gender differences in the same direction across all five subscales.

---

[17] All questions in the survey were mandatory, but if a pupil exited the survey early all the completed questions were saved as a partial response which we kept if this was the pupil's only response.
[18] We conducted mean comparison tests on all outcome variables and found that baseline rates for male pupils were all statistically higher than baseline rates for female pupils ($p < 0.01$).

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. Of the 167 schools recruited, 28 schools formally withdrew from the evaluation, 16 schools from the treatment group and 12 from the control group.

At both baseline and endline points of pupil survey data collection, RPF attempted to minimise attrition (across both treatment and control groups) by extending the window for data collection to account for delays in completing surveys, and by sending reminder emails to schools that had not completed the surveys by the expected time.

Despite these efforts, attrition was observed between randomisation and completion of baseline surveys, with proportionally more schools dropping out from the control group than the intervention group (39% of control schools vs 31% of treatment schools). This differential attrition brings a risk of bias at the analysis stage, which is explored later in this report through balance checks on the baseline data received from these schools post-attrition.

Attrition was also observed between baseline and endline, in terms of both schools failing to complete the endline survey and pupils within schools not completing the endline survey. At a school level we saw similar rates of attrition for control and treatment groups (10% of control schools vs 12% of treatment schools); this was the same at a pupil level (25% of pupils from control schools who completed the baseline did not complete the endline vs 23% of pupils from treatment schools). There were also 5 schools that completed the end line survey but did not complete the baseline survey (2 of these schools were in the control group and 3 were in the treatment group).
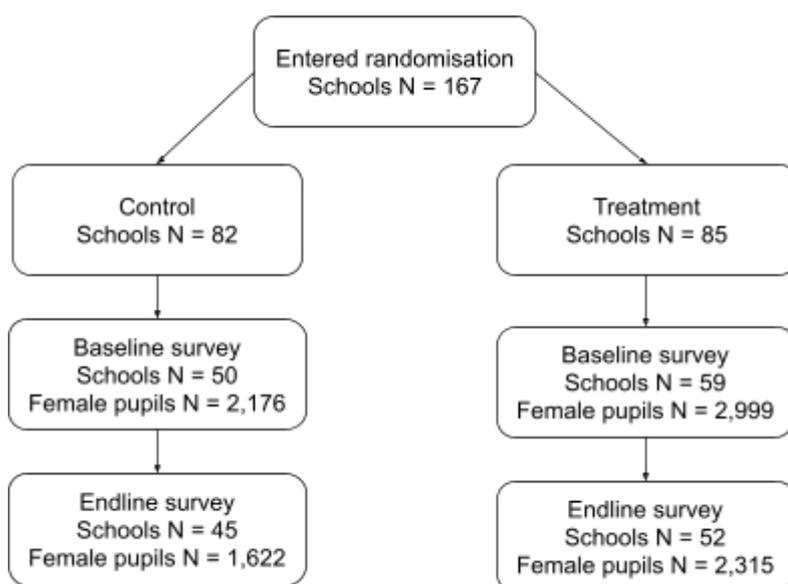
Figure 2: School level attrition by treatment group

This attrition at each stage results in very similar overall attrition rates across experimental groups (45% in the control arm 39% for the treatment arm) and between the point of randomisation and endline survey completion.

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 222 pupils from the total baseline sample of 10,169 (male pupils and female pupils, not necessarily matched to any endline observations) and 258 pupils from the endline sample of 7,638 (male pupils and female pupils). In cases where there were duplicate observations (the same pupil entering the survey twice), we kept only the first complete survey with no straight answering pattern from the pupil. If a pupil never fully completed the survey, we retained their first partially complete entry.

The final data set consisted of (1) data from female pupils who had completed the endline survey matched to their baseline data and (2) data from female pupils who had completed only the endline survey. We used a multistep matching process to match as many baseline and endline surveys as possible. Responses were matched on a combination of school name, full name and date of birth; we also used survey completion dates when manually reviewing possible matches. To improve the number of matches we used both exact matching and 'fuzzy matching[19]' to account for the use of different name spellings or orders and data entry errors. We used an iterative process which involved loosening the matching criteria on different matching variables to identify possible matches at each stage, and manual review to confirm these possible matches (match rates at each matching step are described in table 4). The final analytical sample consists of 3,937 female pupils, 1,622 in the control group and 2,315 in the treatment group.

Table 4: Matching status of endline observations in the final sample

Matching status	Number of observations (girls only)
Endline observations within schools where the baseline survey was completed by some pupils	
Perfect matches	2,976
Fuzzy matches	160
Unmatched	107
Endline observations within schools where no baseline survey was completed	
Unmatched	694
Total observations	
Final sample	3,937

---

[19]Fuzzy matching refers to a matching technique which uses a matching score to identify possible matches across two datasets on a given characteristic (e.g., name), and can thereby guide manual review of possible matches.

Baseline differences in outcome measures

Table 5 shows the results of balance checks for all outcome measures for the baseline survey sample. When balance was checked across control and treatment for data for female pupils and male pupils, the groups were imbalanced on intent to study computer science at GCSE level and both Usefulness and 'Relevance' subscales, but balanced on overall attitudes towards computing. When restricting the sample to female pupils only, all outcome measures were balanced across control and treatment groups.

Table 5: Balance checks for baseline data of female pupils

Covariate	Percentage (or mean) per arm		P-value (OLS)	Balanced?
	Control (n = 4,159)	Intervention (n = 5,799)		
Gender				
Female	52.3%	51.7%	>0.10	Yes
Intention to select computer science as a GCSE subject				
Baseline intention	6.3%	7.7%	<0.10	Yes
Usefulness				
Baseline Usefulness score	2.60	2.62	>0.10	Yes
Interest				
Baseline Interest score	2.31	2.33	>0.10	Yes
Total SCSAS				
Baseline SCSAS score	2.46	2.48	<0.10	Yes

Baseline differences in outcome measures for final analytical sample[20]

Further attrition was observed between baseline and endline data collection. Table 6 shows that at the point of baseline data collection, the final groups (the composition of which is outlined in section 2.5.1) were balanced for the primary outcome (stated intention to select GCSE computer science) but imbalanced across all other outcome measures. The difference in the primary outcome at baseline in the final analytical sample of 1.4 percentage points (6.2% in the control group vs 7.6% in the treatment group) is not statistically significant.

This imbalance on secondary outcomes as well as the differential attrition above suggests that there might be a risk of bias and has implications for analysis, outlined below. While pupils were allocated to control and treatment groups without their knowledge, the teachers were aware of the group their schools were assigned to. This may have affected school level attrition rates in each group if teachers in the group that received the intervention were more engaged with the evaluation and thus more likely to administer the survey with their pupils. Furthermore, in treatment schools the surveys were a component of the first and last lessons of the intervention, whereas control schools did not have these lessons and received survey links through a project pack and reminder emails sent at the start and end of the intervention. Teachers in the control group received the survey links and were asked to administer the

---

[20] This does not include baseline data for female pupils who did not complete the baseline survey.

survey with pupils during normal classes, which teachers may have been less likely to do than if they were delivering the intervention lessons with survey completion incorporated into the lesson.

Table 6: Balance checks for baseline data of female pupils who completed the endline survey

Covariate	Percentage (or mean) per am		p-value	Balanced?
	Control (n = 1,247)	Intervention (n = 1,889)		
Intention to select computer science as a GCSE subject				
Baseline intention	6.2%	7.6%	>0.10	Yes
Usefulness				
Baseline Usefulness score	2.58	2.62	<0.05	No
Interest				
Baseline Interest score	2.29	2.34	<0.05	No
Total SCSAS				
Baseline SCSAS score	2.45	2.49	<0.01	No

Implications for final analysis

In the final sample, we observe a positive difference in the treatment group relative to the control group in the mean baseline value for all outcomes. Although this difference is not statistically significant for the primary outcome (stated intention to select computer science as a GCSE subject), the overall imbalance at baseline suggests a risk of bias in the results: 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:

- The higher average mean for all outcomes in the treatment group suggests that teachers or pupils within the treatment schools that completed the endline survey had more positive attitudes towards computing already at baseline, and may have thus been more engaged with the intervention than treatment schools that did not complete the endline survey and than the control schools that completed the endline. This would imply a risk that our treatment estimates may be biassed upwards (i.e. the treatment effect estimate is higher than it would have been in the absence of attrition).
- On the other hand, these treatment schools with more positive attitudes towards computing at baseline may have been less affected by the intervention if a form of ceiling effect exists, such that the intervention would have less scope for impact in these schools than in treatment schools that did not complete the endline survey and than the control schools that completed the endline. This would imply a risk of downward bias of our treatment estimates.

While it is difficult to determine the extent to which either of the two scenarios above or random chance explain the imbalance observed at baseline in the final sample, there is a risk that this differential attrition may bias the results. This risk is addressed in the interpretation of the results in section 3.

2.1.5 Analytical approach

The full model is presented in Appendix 2. The primary and secondary analyses were both Intention to Treat (ITT) estimates. This means that outcomes were analysed on the basis of the groups that schools and pupils were randomly allocated to, regardless of their compliance with the intervention. Baseline SCSAS score, school Ofsted rating, and school proportion of pupils with FSM eligibility were used as covariates as they could potentially influence the outcomes, thus controlling for these variables could increase the precision of the estimates. We also ran the main specification of the analysis for the primary outcome with baseline stated intention to select GCSE computer science as an additional covariate, to understand whether the results are sensitive to the inclusion of this covariate.

All planned covariates were checked for missing data pre-analysis. For 32 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[21].

Given that the endline data would likely include some pupils who did not complete the baseline survey, we specified pre-trial decision rules for dealing with missing data as baseline scores on the SCSAS were to be used as a covariate in the analysis. In the final sample, approximately 20%[22] of pupils were missing baseline SCSAS scores (above the threshold of 5% for listwise deletion), and multiple imputation[23] was performed, whereby predicted values were substituted where data was missing. Imputation for SCSAS outcomes were done for total SCSAS score as well as total Usefulness and total Interest scores. No values were imputed within subscales of SCSAS.

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 indicators 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[24]
3. Complete case analysis

We also conducted robustness checks to understand the sensitivity of the results to alternative specifications[25] and to how the primary outcome was measured. We also

---

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

[22] See table 4 for the breakdown of endline, baseline and matched survey data.

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

[24] 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.

[25] This check was carried out on the main specification and the primary outcome variable only.

conducted additional analysis to investigate whether all treatment school pupils received the intervention as planned.

2.1.6 Limitations

Attrition

An implication of generally high attrition is that the analysis may not be powered to detect a change in outcome measures of the targeted effect size specified pre-trial. In power calculations conducted prior to the start of the trial, an attrition rate of 35% was predicted, leaving a predicted 54 schools per treatment arm, and 44 pupils[26] completing the survey per school was assumed. Based on these assumptions we estimated a minimum detectable effect size (MDES) of 4 percentage points if the treatment effect was an increase in intention to select computer science as a GCSE subject. The final sample consisted of 45 control schools and 52 intervention schools with an average of 41 responses per school. This final sample was thus closely in line with our assumptions, and based on the final sample size we would still expect to be powered to detect a 4 percentage point effect size.

However, differential attrition across treatment groups may introduce a bias in the results. This risk is discussed in section 2.1.4.

Pupil survey outcome measures

Given that the pupils in this trial were in year 8 for the duration of the intervention and that at pupils at most sample schools select GCSE subjects in year 9,[27] most pupils are over a year away from making their GCSE subject choices. Because of this, we rely on self-reported intention to study GCSE computer science as our primary outcome measure, rather than actual GCSE subject choices. While the correlation between this measure and actual subject choices is unknown, it is promising that the baseline rate of our analytical sample (7.1%) is close to an estimate provided by the DfE of 6% of female pupils selecting computer science. It is possible that the sample rate may also be marginally higher than the national rate given that these pupils are from schools that volunteered to take part in this intervention.

We included an alternative outcome variable which we used as a robustness check[28]. This variable was gathered in the same survey but through a multiple choice question (MCQ) rather than single-choice ('yes', 'no', 'don't know'). It appeared on the next screen after the binary question. A list of 8 GCSE subjects/subject groups[29] were listed (including computer science) and respondents were able to select any of the subjects they were planning to study. We then constructed the binary outcome variable by coding any responses that included computer science as a chosen subject as 1. This additional survey question was intended to simulate the decision that pupils will have to make between possible GCSE subjects, rather than a decision to select GCSE computer science independently of other

---

[26] Our initial baseline checks found an average of 44 pupils per school completing the survey.
[27] We collected data on this using a survey shared together with regular RPF communications with schools following the endline survey, and received responses from teachers at 117 of our sample schools. Of these 98 reported their current year 8 pupils would be selecting their GCSE subjects in the start of year 9 (January to March 2023).
[28] This robustness check was not prespecified in the evaluation protocol.
[29] Some subjects were grouped together: geography and history, french german and spanish and music, art and drama.

subject selections. Presenting other GCSE subjects may make pupils think harder about which subjects they may want to prioritise. However, given the limitations above with the survey measures, there is a limit to the extent to which a survey question can simulate this decision environment and the set of options pupils will face at their respective schools, and it could be that presenting multiple options could make respondents more likely to select any one subject. Therefore, the intention of this robustness check was mainly to test for the sensitivity of the results to the way the question was being asked, and check for consistency across different versions of the outcome. The proportion of female pupils selecting GCSE computer science in the multiple choice question was higher than the proportion of female pupils who selected "Yes'' in the binary variable, 14% of female pupils reported the intention to select GCSE computer science in both survey questions (8% reported this intention for the single-choice question, and 14% for the multiple choice question). Intention to study computer science in both questions was found to be strongly correlated using Pearson's correlation ($r > 0.6$).

In the multiple choice question, 75% of respondents selected 5 subjects/subject groups or less (out of a possible 8). Pupils are generally expected to choose four GCSE subjects and one reserve, and the distribution of number of options selected indicates that this outcome adheres to that expectation (even though we did not force a minimum or maximum number of choices). This suggests that the slightly higher proportion of pupils reporting an intention to study GCSE computer science in the multiple choice question was not driven by ignoring the limit to the number of subjects that can be selected.

2.2 Implementation and process evaluation

Alongside the impact evaluation, we conducted a mixed-methods IPE, which examined how the intervention was implemented and the mechanisms of change.

2.2.1 Research questions

The IPE aimed to address the following research questions:

1. How do the teachers experience the training organised by RPF?
2. How are the lessons delivered?
3. How do the pupils experience the intervention?
4. Does the intervention address the mechanisms anticipated to lead to changes in the targeted behaviour?

2.2.2 Research design

We used a mixed-methods design to address the research questions described above. This consisted of case studies of the intervention schools, which included lesson observations, pupil focus groups and teacher interviews as well as an online survey sent to all the treatment schools.

Each activity provided information relevant to the research questions:

1. The teacher interviews allowed us to explore the context of each school, teacher experience of the intervention, how lessons were delivered and possible mechanisms of change.
2. The focus groups helped us to understand how the pupils experience the intervention and to identify possible mechanisms of change.
3. The lesson observations allowed us to explore how the teachers deliver the lessons and use the materials, as well as how the pupils respond to and engage with the lessons.
4. The teacher survey enabled us to capture a broader range of teachers' experience by gathering a greater number of teacher perspectives than possible through the case studies alone.

2.2.3 Sampling and recruitment

Across the participating schools, we selected case study schools to represent range and diversity. 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 (North/South), ii) proportion of pupils eligible for FSM (above or below average[30]) and (iii) Ofsted rating. We recruited two schools in the North of England and two in the South of England. The proportion of FSM eligible pupils at each school was retrieved from the DfE national information about schools[31]; we recruited two schools that had an above-average and two schools that had a below-average proportion of pupils eligible for FSM. Given that no schools in the sample were rated '4: Inadequate', we had to adapt our approach and aim for at least one school each with a '1: Outstanding', '2: Good' or '3: Requires Improvement' rating. We also met our adapted sampling criteria for Ofsted rating.

Teacher sampling for interviews

For selecting teachers within the case study schools, the primary criterion was teacher gender and the secondary criterion was teaching experience (more or less than five years of teaching experience). We recruited five teachers: one at each of the case study schools and one teacher for a teacher interview whose school had already completed the intervention[32]. Information on computing teacher gender and experience was not readily available so we did not recruit schools based on gender. However, the teacher sample from the four case study

---

[30]At the time of sampling, 27.7% was the average pupils eligible for FSMs in England at any time during the past six years, so this was used as the threshold for above and below average. This was retrieved from the government school search website.

[31] This was retrieved from the government school search website.

[32] This teacher had already finished the intervention so their school was unable to take part in the lesson observations and pupil focus groups. Therefore, we conducted only a teacher interview.

schools who agreed to take part in the evaluation happened to consist of two female teachers and two male teachers.

Pupil sampling for focus groups

We aimed to recruit a sample of two groups of four pupils from each of the four case study schools to take part in pupil focus groups. For pupils, the primary criteria included gender and level of engagement with computing (as determined by their teacher). Considering the intervention focused on female pupils, we aimed for a majority of all female pupil focus groups and at least one male pupil focus group to help identify the extent to which mechanisms for change are specific to female pupils.

We conducted seven pupil focus groups, two in each school apart from one school (S05) where only one could be conducted due to time constraints. Five of the pupil focus groups were female focus groups and two were male focus groups. The number of pupils in each pupil focus group varied from three to seven pupils. In total we spoke to 28 pupils across four schools.

Table 7 below details the achieved sample of case study schools and the additional online teacher interview (S03).

Table 7: Achieved case study sample

School	Profile	Teacher	Pupils	Data collection
S01	- Located in the North - Below average FSM - Ofsted rating: Outstanding	- Joint Head of Department - More than 15 years of experience - Male	- 8 year 8 pupils - 8 female pupils - Range of confidence in computing	- Teacher interview - Lesson observation - Two pupil focus group discussions - All activities were in-person
S02	- Located in the South - Below average FSM - Ofsted rating: Good	- Head of Department - 5 years of experience teaching computing - Male	- 7 pupils - 4 female pupils - 3 male pupils - Range in level of engagement	- Teacher interview - Lesson observation - Two pupil focus group discussion - All activities were in-person
S03	- Located in the North - Below average FSM - Ofsted rating: Good	- Head of faculty - 10 years of teaching experience - Male	None	-Teacher interview -Conducted virtually
S04	- Located in the North - Above average FSM - Ofsted rating: Requires improvement	- Head of Department - 10 years of experience - Female	- 6 year 8 pupils - 3 female pupils - 3 male pupils - Range of confidence in computing	- Teacher interview - Lesson observation - Two pupil focus group discussion - All activities were in-person
S05	- Located in the South - Above average FSM - Ofsted rating: Good	- Head of Department - Unknown experience - Female	- 7 year 8 pupils - 7 female pupils - Range of confidence in computing	- Teacher interview - Lesson observation - One pupil focus group discussion - All activities were in-person

School recruitment

RPF reached out to teachers[33] in 16 schools that we had identified and could allow us to meet our sampling criteria, to ask whether they would like to be involved in the IPE. Five schools that collectively met these criteria indicated their interest and either agreed to a call to schedule the school visit or scheduled the visit via email.

Online feedback survey teacher sample

An online survey was sent to 69 teachers from all the schools in the treatment group, with the exception of those who had informed RPF of their withdrawal from the evaluation. It was sent by RPF to one teacher per school who was the 'lead' teacher in RPF's database. RPF also monitored online training completion via a survey embedded in the training and shared the data collected with us.

---

[33] RPF sent an email to the lead teacher from each school.

2.2.4 Data collection methods

Teacher interviews

Individual, in depth semi-structured interviews were conducted with five teachers who delivered the Relevance lessons to explore their experiences of the intervention and any factors that influenced their ability to implement the intervention with their pupils. One interview was conducted in each of the four case study schools and one online. We were able to conduct 45-60 minute interviews with four of the teachers, and a 14 minute interview at one school (S05) due to time constraints.

Pupil focus groups

Group discussions were held with pupils at four of the case study schools and lasted approximately 20 to 30 minutes. They consisted of a semi-structured interview and an independent worksheet activity where pupils individually circled 'Yes', 'I don't know' or 'No' to a series of statements and then explained their answer. After the first two pupil focus groups the worksheet was slightly amended.[34] These activities were designed to create the opportunity for pupils to share their current perceptions of computing, as well as their experience of the Relevance lessons, and how this compared to 'normal' computing lessons.

Lesson observations

At each school, we observed teachers delivering a Relevance lesson and pupil reactions before conducting the pupil focus groups. At two schools we observed lesson 13 and at the other schools lesson 9 and lesson 10, respectively. Each observation was approximately 45 to 60 minutes long, apart from in one school (S05) where the lesson observation was 25 minutes long[35] due to time constraints. These observations were designed to independently assess pupil engagement, lesson fidelity and facilitators and barriers to lesson delivery.

Online feedback survey teacher sample

Teachers in treatment schools were asked to complete a short online feedback survey (see Appendix 3) (n=28 respondents).[36] The survey asked a mix of open-ended and closed-ended questions. The purpose of the survey was to gather a broader range of perspectives and additional information on certain aspects of the intervention, such as which lessons teachers delivered.

---

[34] After reviewing worksheets from the first two schools, the worksheet was slightly adapted to include (1) a statement to encourage honest feedback (*'Please answer honestly. There are no right or wrong answers'),* (2) an extra statement to gather opinions on whether pupils think computing is important in their everyday lives, and (3) additional text to the statement: 'I think computing can be used to solve problems *[I care about*]'.

[35] We observed the first half of the lesson.

[36] Whilst the survey was only sent to one teacher per school, teachers may have forwarded it to other teachers who were involved in the intervention so it is possible that more than one teacher from the same school filled out the survey. We also do not know whether teachers in our case study schools were among those that had completed the survey.

2.2.5 Analysis

Case study data

Interview transcripts and fieldnotes were managed using the Framework Approach[37]. 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.

There are several considerations to keep in mind when interpreting the findings from the analysis:

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 sample size was small, with only four case study schools and one additional online teacher interview. Most teachers led extra-curricular activities associated with computing and some had taken part in projects that aimed to get more females involved in computing before, which could indicate they were more motivated to increase the number of female pupils in computing than the average computing teachers at the trial schools.[38] The findings might thus not be representative of the experiences of all teachers who implemented the intervention.

Online feedback survey data

Open-ended answers were analysed using the Framework Approach summarised above and closed-ended questions, such as multiple choice questions, were used to generate descriptive statistics.

---

[37] Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (Eds.). (2013). *Qualitative research practice: A guide for social science pupils and researchers*. Sage.
[38] We reached out to 16 schools to take part in the IPE and not all schools responded which indicates that perhaps the schools that took part in the IPE were even more motivated than the schools in the overall sample.

3. Impact evaluation findings

Key findings:

- Our main analysis did not find a statistically significant impact of the intervention on female pupils' intention to study GCSE computer science. When using an alternative outcome based on a slightly different question about intention to select GCSE computer science, the treatment effect is larger and is statistically significant for some, though not all, of the analysis specifications.
- We did not find statistically significant evidence of a positive impact on female pupils' attitudes towards computing across all treatment schools for our two secondary outcomes (perceived relevance of and interest in computing) and our exploratory outcome (overall attitudes towards computing). While the treatment effect is positive, it is small in magnitude and not statistically significant.
- There is suggestive evidence that not all treatment school pupils who completed the endline survey received the intervention as intended, which could have limited the impact of the intervention across all treatment school pupils.

3.1 Primary analysis: intervention impact on female pupils' intention to select GCSE computer science

There was no statistically significant evidence that the intervention positively impacted female pupils' intention to study GCSE computer science relative to the control group. The proportion of female pupils stating they intended to study computer science at GCSE for the full sample of female pupils was 8.4%. For the control group this proportion was 7.6% and for the intervention group it was 8.9%. Compared to the control group, the pre-specified multiple imputation model found a treatment effect estimate of 0.2 percentage points (p=0.852) for the intervention, which is not statistically significant[39]. The estimated treatment effect of 0.2 percentage points is small in magnitude, and lower than the anticipated effect of 4 percentage points for this intervention[40].

This finding was consistent across the missingness indicator and complete case analysis model specifications, where we detected positive treatment effects of 0.2 percentage points and 0.1 percentage points respectively and neither estimate was significant. All estimates are presented after controlling for the prespecified covariates (baseline SCSAS score, school Ofsted rating, school proportion of pupils with FSM eligibility and proportion of female pupils in the school). When running the main specification of the analysis for the primary outcome with baseline stated intention to select GCSE computer science included as an additional

---

[39] The analysis controls for baseline differences in SCSAS scores and other covariates, which explains why the estimated treatment effect is much smaller than the difference in raw means.
[40] This was the effect size that we had discussed with x to be achievable for this intervention.

covariate, the estimated treatment effect size becomes -0.01pp, and is again not statistically significant (p=0.992). As described in section 2.1.4, the mean baseline SCSAS scores for the final sample were slightly higher in the treatment group than in the control group. While the differential attrition that led to this small imbalance introduces a risk of bias in the estimates, it is difficult to determine the possible direction of this bias.

The results of the primary analysis are presented in Table 8. Primary model specifications, along with full regression tables, can be found in Appendix 2.

Table 8: Impact evaluation results for stated intention to select GCSE computer science

Outcome: Intention to select GCSE computer science	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group mean	7.6%		7.8%
Treatment group mean	8.9%		9.1%
Estimated treatment effect (in percentage points)	0.2pp	0.2pp	0.1pp
p-value	0.852	0.845	0.917
N	3,937[41]		3,038

Figure 3 shows the raw control mean and estimated treatment effect of the intervention on female pupils' intent to study computer science at GCSE level using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention.

---

[41] There were no observations in the final sample missing this outcome variable.
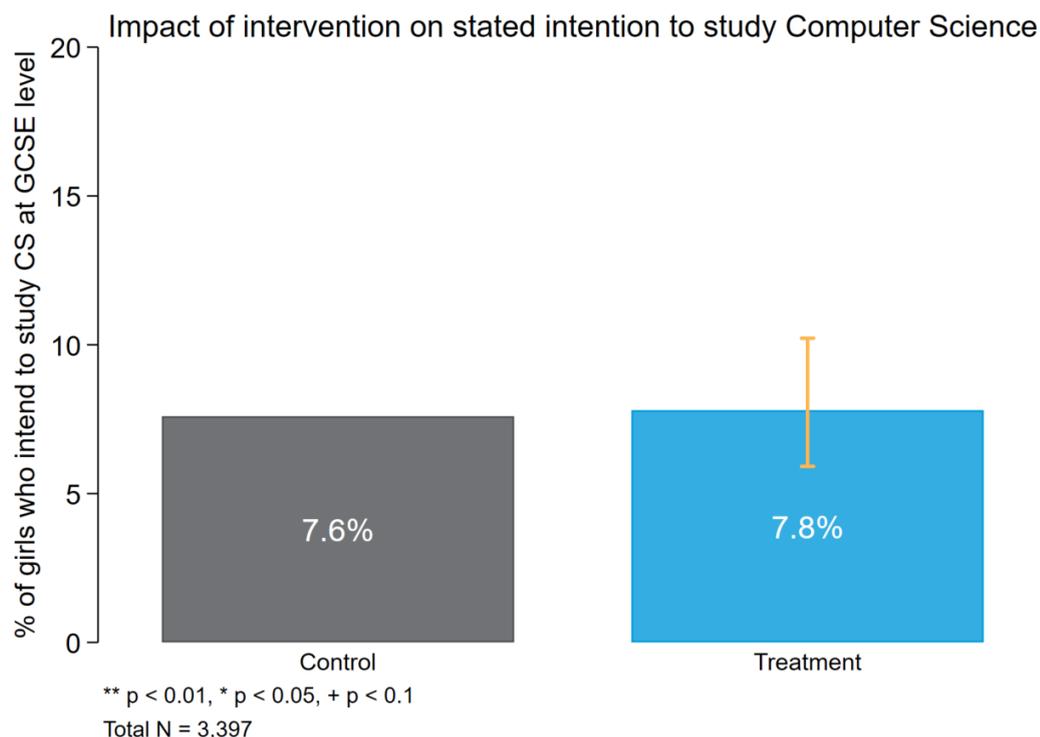
Figure 3: Model-adjusted intention scores by treatment group

3.2 Secondary analysis: intervention impact on female pupils' specific attitudes towards computing

3.2.1 Relevance

There was no evidence that the intervention positively impacted female pupils' perception of the relevance of computing, as measured by scores on the SCSAS Usefulness subscale. The mean score of this subscale (range 1-4) for the full analytical sample was 2.52 (SD=0.642). For the control group it was 2.48 (SD=0.632) and for the intervention group it was 2.55 (SD=0.649). Compared to the control group, the pre-specified multiple imputation model found a treatment effect of 0.03 points (p=0.269) on a 1-4 scale for the intervention, which is not statistically significant at conventional significance levels[42].

This finding was consistent across the missingness indicator and complete case analysis model specifications, which also found a positive treatment effect of 0.03 on a 1-4 scale for the intervention. All estimates control for the prespecified covariates (baseline SCSAS score, school Ofsted rating, school proportion of pupils with FSM eligibility and proportion of female pupils in the school). As with the primary analysis, the differential attrition that led to the small

---

[42] The analysis controls for baseline differences in SCSAS scores and other covariates, which explains why the estimated treatment effect is much smaller than the difference in raw means.

baseline imbalance introduces a risk of bias in the estimates; however, it is difficult to determine the possible direction of this bias.

The results of the analysis for this outcome are presented in Tables 9. Secondary model specifications, along with full regression tables, can be found in Appendix 2.

Table 9: Impact evaluation results for the Usefulness subscale score

Outcome: Usefulness subscale score	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group unadjusted mean	2.48	2.48	2.48
Treatment group unadjusted mean	2.55	2.55	2.55
Estimated treatment effect (standard error)	0.03 (0.02)	0.03 (0.03)	0.03 (0.02)
p-value	0.269	0.231	0.267
N	3,937	3,877[43]	3,003

Figure 4 shows the raw control mean and estimated treatment effect of the intervention on Usefulness subscale scores from the SCSAS using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention.

---

[43] Missingness indicators were created for all independent variables but in the case where the dependent variable was missing this was excluded from the analysis which is why the multiple imputation and missingness indicator models have differing observations.
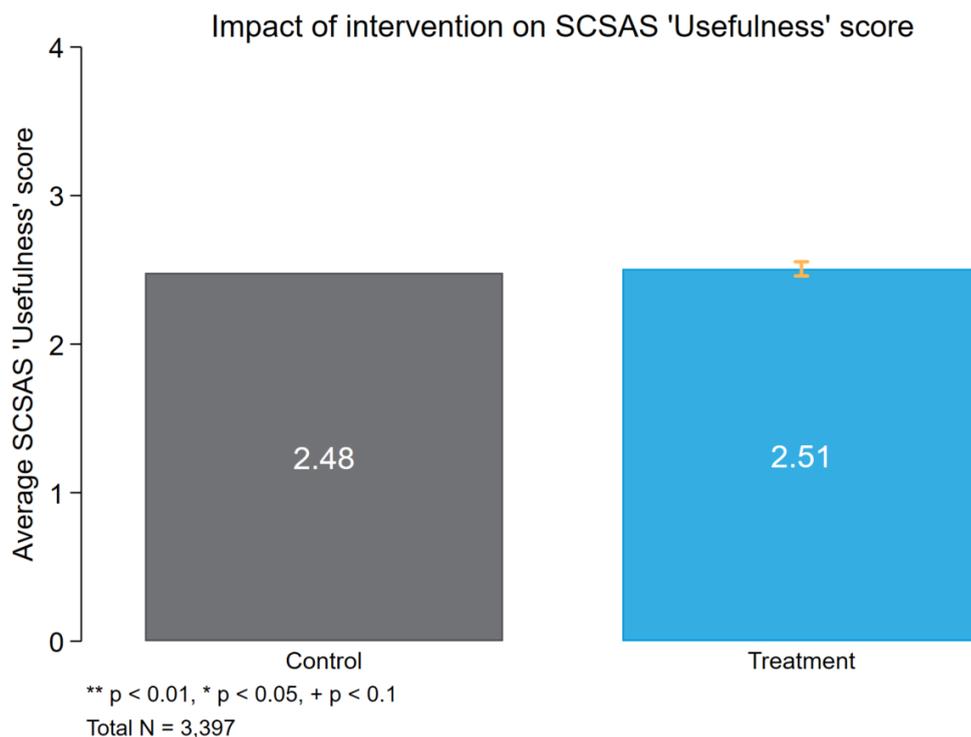
Figure 4: Model-adjusted SCSAS Usefulness scores by treatment group

3.2.2 Interest

There was no evidence that the intervention positively impacted female pupils' interest in computing, as measured by scores on the SCSAS Interest subscale. The mean score of this subscale (range 1-4) for the full analytical sample was 2.24 (SD=0.690). For the control group it was 2.19 (SD=0.690) and for the intervention group it was 2.26 (SD=0.689). Compared to the control group, the pre-specified multiple imputation model found a treatment effect of 0.02 points (p=0.415) on a 1-4 scale for the intervention, which is not statistically significant at conventional significance levels[44].

This finding was consistent across the missingness indicator and complete case analysis model specifications, which found a positive treatment effect of 0.04 and 0.02 points on a 1-4 point scale, with neither being significant. All estimates control for the prespecified covariates (baseline SCSAS score, school Ofsted rating, school proportion of pupils with FSM eligibility and proportion of female pupils in the school). As with the primary analysis, the differential attrition that led to the small baseline imbalance introduces a risk of bias in the estimates; however, it is difficult to determine the possible direction of this bias.

The results of the analysis for this outcome are presented in Tables 10. The model specifications, along with full regression tables, can be found in Appendix 2.

---

[44] The analysis controls for baseline differences in SCSAS scores and other covariates, which explains why the estimated treatment effect is much smaller than the difference in raw means.

Table 10: Impact evaluation results for Interest subscale score

Outcome: Interest subscale score	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group unadjusted mean	2.19	2.20	2.19
Treatment group unadjusted mean	2.26	2.26	2.25
Estimated treatment effect (standard error)	0.02 (0.03)	0.04 (0.03)	0.02 (0.03)
p-value	0.415	0.272	0.412
N	3,937	3,919[45]	3,026

Figure 5 shows the raw control mean and estimated treatment effect of the intervention on Interest subscale scores from the SCSAS using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention.
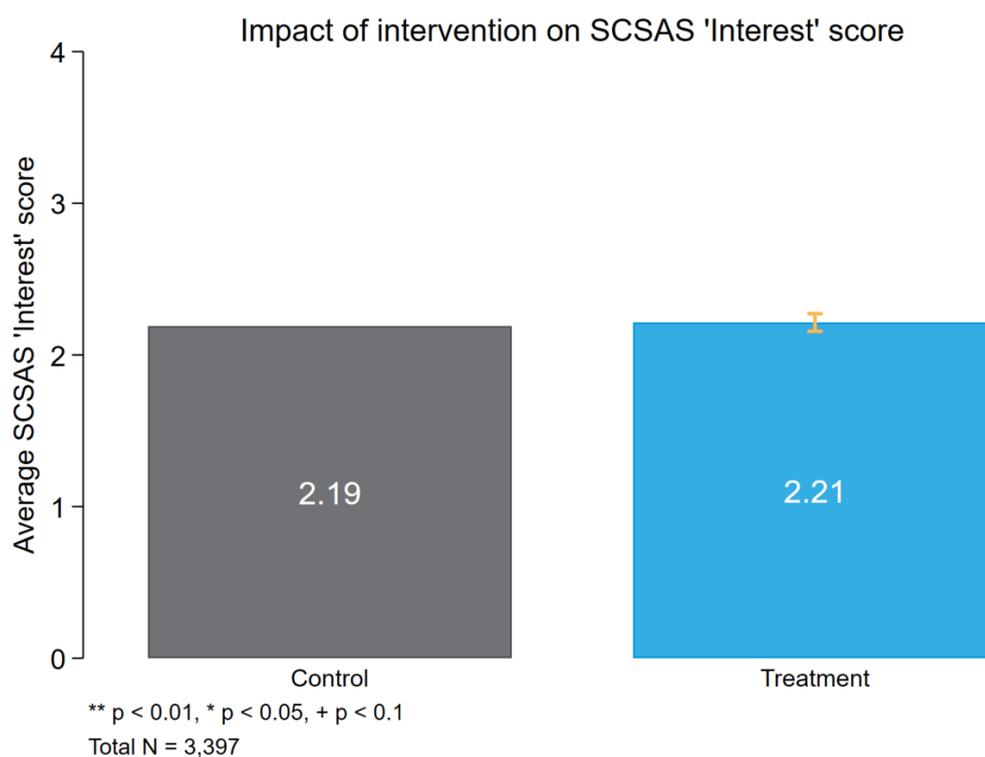
Figure 5: Model-adjusted SCSAS Interest scores by treatment group

---

[45] Missingness indicators were created for all independent variables but in the case where the dependent variable was missing this was excluded from the analysis which is why the multiple imputation and missingness indicator models have differing observations.

3.3 Exploratory analysis: intervention impact on female pupils' overall attitudes towards computing

There was no evidence that the intervention positively impacted female pupils' overall attitudes towards computing relative to the control group, as measured by scores on the SCSAS. The mean score on the SCSAS scale (range 1-4) for the full analytical sample was 2.42 (SD=0.533). For the control group it was 2.38 (SD=0.524) and for the intervention group it was 2.45 (SD=0.537). Compared to the control group, the pre-specified multiple imputation model found an estimated treatment effect of 0.03 points (p=0.198) on a 1-4 scale for the intervention, which is not statistically significant at conventional significance levels. This finding was also consistent across the missingness indicator and complete case analysis model specifications. As with the primary analysis, the differential attrition that led to a small baseline imbalance introduces a risk of bias in the estimates, however it is difficult to determine the possible direction of this bias.

The results of the exploratory analysis are presented in Table 11. Exploratory model specifications, along with full regression tables, can be found in Appendix 2.

Table 11: Impact evaluation results for exploratory outcome total SCSAS score

Outcome: Total SCSAS Score	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group unadjusted mean	2.38		2.38
Treatment group unadjusted mean	2.45		2.45
Estimated treatment effect (standard error)	0.03 (0.02)	0.04 (0.03)	0.03 (0.02)
p-value	0.198	0.144	0.215
N	3,937		3,038

Figure 6 shows the raw control mean and estimated treatment effect of the intervention on total scores from the SCSAS using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention.
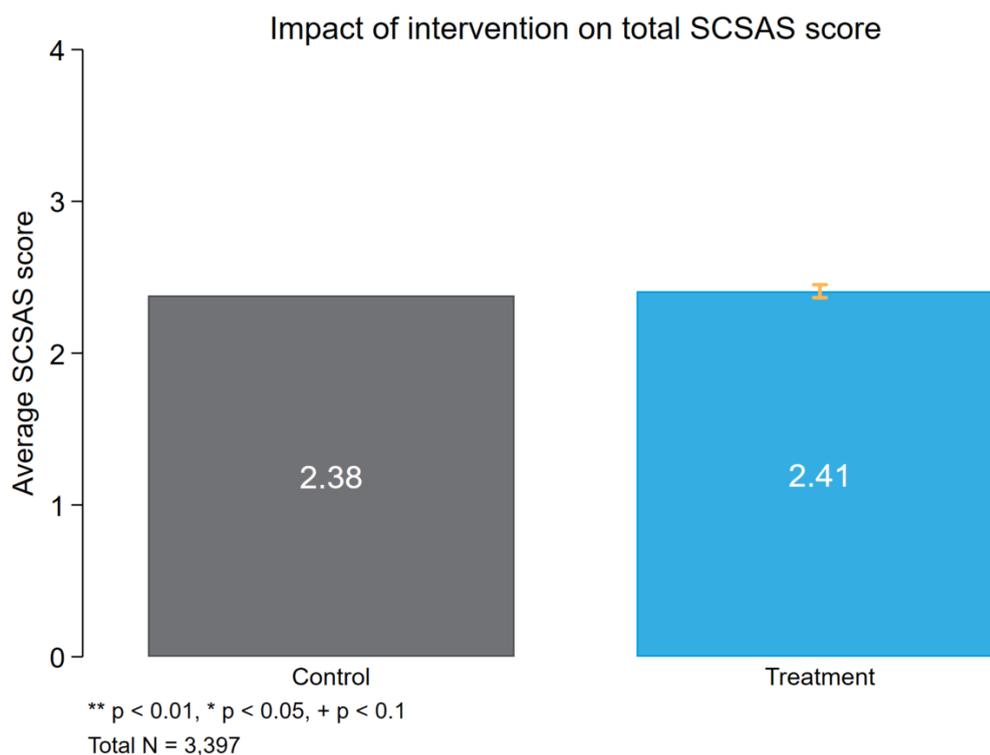
Figure 6: Model-adjusted SCSAS scores by treatment group

3.4 Robustness checks

In section 3.1, we included results from running the analysis for the primary outcome with baseline stated intention to select GCSE computer science as an additional covariate, to understand the sensitivity of the results to the inclusion of this covariate. This section describes additional robustness checks conducted as part of the primary and secondary analysis.

3.4.1 Alternative primary outcome variable

We carried out robustness checks using an alternative primary outcome variable to understand how sensitive the results were to the way pupils are asked to report their intention to select computer science as a GCSE subject. This alternative outcome was based on a survey question that asked pupils to select a set of subjects, rather than a decision to select GCSE computer science independently of other subject selections. The potential advantages and limitations of the two outcome variables are discussed in section 2.1.6. The results of this robustness check are presented in Table 12.

When using this alternative outcome measure, we find statistically significant results and a positive treatment effect estimate of the intervention on female pupils' stated intention to select computer science as a GCSE subject. The proportion of female pupils stating they intended to study computer science at GCSE for the full sample of female pupils was 14%. For the control group this proportion was 11.8% and for the intervention group it was 15.2%.

Compared to the control group, the pre-specified multiple imputation model found a treatment effect estimate of 2.9 percentage points (p=0.03) for the intervention, which is statistically significant at conventional significance levels. This finding was consistent with the missingness indicator specification but not the complete case analysis model specification, which found a smaller effect estimate that was not statistically significant. When running the main specification of the analysis for this alternative primary outcome with baseline stated intention to select GCSE computer science included as an additional covariate, the estimated treatment effect size was the same as in the complete case analysis specification (1.8pp), and similarly not statistically significant (p=0.213).

While it is unclear that this alternative indicator more accurately captures our outcome of interest, and the estimated treatment effect is not robust across specifications, these additional results provide suggestive evidence that the intervention may have a small positive effect on female pupils' intention to study GCSE computer science. As with the analysis using the main primary outcome and described in section 2.1.4, the differential attrition observed introduces a risk of bias in the estimates, however it is difficult to determine the possible direction of this bias.

Table 12: Robustness check on primary outcome using alternative variable

Outcome: Intention to select GCSE computer science (multiple choice)	(1) Multiple imputation model	(2) Baseline missingness indicator	(3) Complete case analysis
Control group mean	11.8%		12.8%
Treatment group mean	15.2%		15.5%
Estimated treatment effect (in percentage points)	2.9pp	2.7pp	1.8pp
p-value	0.030	0.047	0.161
N	3,937		3,038

3.4.2 Compliance

We included two questions in the endline survey to investigate whether pupils in treatment schools are likely to have received the intervention as planned. The first question asked pupils whether they had taken part in a course that involved designing, building and testing an app[46], and the second asked them which part of the course they preferred[47].

We used the responses to both questions to construct a variable intended to capture whether female pupils at a given school received the intervention as planned. Table 13 shows the number of schools in the treatment group in which more than 75%, between 50% and 75%, and less than 50% of female pupils who completed the endline who reported 'yes' to the first

---

[46] Survey question: Have you recently taken part in a course which involved designing, building and testing an app?
[47] Both questions and corresponding answer options can be found in Appendix 1.

question, and did not respond 'I didn't take part in any of these lessons' to the second question.

There are various reasons why pupil responses to these questions may not reliably indicate whether the intervention was implemented as planned at their school. Pupils in treatment schools may not recall the intervention, which could be exacerbated by absences from some lessons. In addition, their responses cannot distinguish between partial and full implementation of the intervention (e.g., the pupil may report not having done the intervention because parts of the intervention were skipped or covered in less detail by the teacher). In control schools, pupils may have received computing lessons or additional programmes with similar sounding content as the one described in the survey question, which could lead them to report they have taken part in a course that involved designing, building and testing an app.

Despite these limitations, the lack of recollection reported by many female pupils across treatment schools suggests that not all pupils who completed the endline survey received the intervention as planned or at all. In the four schools where less than half the female pupils reported taking part in a course that involved designing, building and testing an app, survey responses suggest that most pupils may not have received the intervention. If some pupils or schools did not receive the intervention as planned, this may have limited the impact of the intervention.

Table 13: Proportion of pupils in treatment schools who recalled partaking in a course that involved designing, building and testing an app

Share of pupils (girls only)	Number of schools	Number of pupils	Mean proportion of pupils who recalled taking part
Treatment schools where >75% of pupils reported taking part	17 (33%)	669 (29%)	82%
Treatment schools where >50% and =<75% of pupils reported taking part	31 (59%)	1,433 (62%)	63%
Treatment schools where =<50% of pupils reported taking part	4 (8%)	203 (9%)	40%
Total	52	2,315	

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 of pupils and teachers (or engagement with the intervention elements) and 3) mechanisms through which the intervention could have affected the target outcomes. Survey results throughout both sections are not necessarily representative of all schools due to self selection, instead they provide a broader range of teacher experiences.

4.1 Implementation

4.1.1 Fidelity

This section describes how closely the implementation of the intervention at the IPE case study schools matched the intended implementation, as described in the training and materials from RPF.

> **Key findings:**
> - Teachers found it challenging to deliver all 11 core sessions due to term length and most opted to skip sessions in the latter half of the intervention.
> - Teachers adapted the lesson structure to increase time for specific or difficult activities.
> - The extent to which teachers completed the session activities and objectives varied.
> - Teachers found the resources useful but made small adjustments, such as skipping or adapting the teacher slides and utilising the worksheets in different ways to meet the needs of the pupils.
> - According to the teacher survey, completion rates of the three online training sessions were high.

Lesson delivery

Number of sessions[48] delivered

Many teachers did not deliver all 11 core sessions. The intervention consisted of 11 core sessions and two additional, optional sessions (session 7 and session 11, see Table 16).

---

[48] Sessions refer to the sessions which RPF organised the intervention into, one session is intended to be delivered in one lesson.

Eight out of 26 teachers[49] who responded to the survey completed all 11 core sessions. An additional 5 teachers delivered at least 11 sessions, but skipped at least one core session. In addition:

- Five survey respondents indicated completing 12 sessions and none reported completing 13 lessons in the survey, though one teacher we interviewed reported delivering all 13 sessions, which suggests that other teachers in the treatment group may have also delivered 13 sessions.
- Among the 12 teachers who reported delivering less than 11 sessions, the number of sessions delivered varied from 3 to 10.

Teachers tended to skip core sessions scheduled for later on in the intervention, suggesting that teachers ran out of time to deliver all the sessions. According to teacher survey responses, implementation of the first half of the intervention (sessions 1 to 6) was high in comparison to implementation of the second half of the intervention (sessions 7 to 13). The majority of teachers were able to complete most of the sessions, but a lower proportion of teachers reported delivering the last three sessions (11, 12 and 13) relative to the earlier sessions. Pupils therefore might not have accessed the full range of sessions originally intended. Table 15 shows the breakdown of sessions delivered. It is worth noting that not all teachers completed the survey, and this data may thus not be representative of the full sample of intervention schools.

Table 15: Results from a survey question asking teachers which lessons they delivered[50]

Lesson	Delivered	Didn't deliver [51]
Lesson 1: Technology in your life	25	1
Lesson 2: Getting to know App Lab	26	0
Lesson 3: Event-driven programming	26	0
Lesson 4: Develop functionality through selection	23	3
Lesson 5: Changing your world with technology	24	2
Lesson 6: Deciding which problem to solve	23	3
Lesson 7: Industry engagement (optional)	3	23
Lesson 8: Understanding user needs	19	7
Lesson 9: UX design	20	6
Lesson 10: Build your app	23	3
Lesson 11: Continue to build your app (optional)	18	8
Lesson 12: Finish building your app	16	10
Lesson 13: Test and evaluate your app	13	13

In general, teachers skipped session 7, which requires teachers to organise a visit from an industry expert. Only three teachers who completed the survey delivered session

---

[49]One teacher who completed the survey did not deliver any lessons and another dropped out after delivering four lessons. Therefore, the majority of the descriptive statistics omit the two teachers which did not complete the intervention apart from sections on online training as they still reported completing the training.

[50] Teachers referred to sessions as lessons and so 'lesson' was used in the survey.

[51] Note that one teacher who completed the survey did not deliver any lessons and another dropped out after delivering four lessons.

7. Teachers reported in both the survey and interviews that time constraints made it difficult to deliver session 7; one teacher explained in an interview that they skipped session 7 due to anticipated lack of pupil engagement.

"We just did not have enough time in six weeks to get the organisations in." (S05: Teacher interview)

"Where we could skip over that lesson seven where they're marketing, I don't think they'd have engaged with it as much." (S04: Teacher interview)

Teachers skipped lessons when they found it difficult to fit the intervention into one term, especially in schools where computing lessons were shorter or less frequent. Surveyed teachers reported that they aimed to complete the intervention in one term so they could return to their school curriculum in the next term. This constrained the number of the 11 core sessions they could get through. (See the 'Time and practical challenges' in the 'Feasibility' section for more explanations).

"We didn't want to go on longer than the term and so had to fit it into that." (S16: Teacher survey)

Lesson structure

Some teachers condensed sessions or added extra lessons to suit the needs of their class. Whilst there were 11 core sessions to deliver, teachers adapted the delivery of the lessons to meet pupil needs. For example, an interviewed teacher condensed session 8 and 9 into one lesson to give pupils more time to make and design their app. In contrast, another surveyed teacher added extra lessons which focused on learning how to use App Lab to increase pupil understanding of the app making tool.

Teachers proactively set session content as homework when they couldn't deliver the session in one lesson. The lesson primers and lesson plans didn't refer to homework assignments, except lesson plan 13 which specified that user testing and feedback could be completed as homework. However, interviewed teachers reported assigning lesson content as homework when they couldn't finish the session content in one lesson. For example, one teacher explained that they set part of session 9 (which focused on user experience) as homework because they were unable to complete the activities within the lesson.

"The design lesson definitely took longer, and I had to get them to meet as groups outside of lessons and come up with their design plans as homework because I found that we weren't getting all of it done in the lesson." (S05: Teacher interview)

Lesson activities

The extent to which teachers carried out the intended session activities varied from completing all of the session activities to significantly modifying the session activities. Both the lesson primer and teacher slides[52] outlined the activities for each

---

[52] In each teacher slide deck there is a slide titled 'Session activities' which lists the activities to be completed in the lesson.

session. The lesson primer also outlined the suggested timings for each activity, which totalled one hour, and the teacher slides outlined the session objectives.

In session 13, pupils are asked to (1) plan a user test, (2) conduct a user test, (3) refine their prototype and (4) reflect on their prototype in order to *'Demonstrate your understanding of the importance of testing and evaluation in the app-development lifecycle with App Lab'*. The user testing involved watching the tester carefully to see what they do and say when using the app. However, during lesson observations a teacher slightly deviated by setting a five-minute timer and asked pupils to go around the classroom and spend five minutes testing each other's app, thus omitting the watching element. Another teacher decided to do group presentations on their app followed by questions and thus did not cover the testing and evaluation stage of the intervention. However, it is possible that the teacher incorporated this into another lesson.

Group allocation

Teachers allocated pupils into groups based on the current seating plan, or gender. This was a deviation from the session 1 lesson primer, which suggests that teams consist of an equal mix of genders, a range of talents and mixed aptitude. During lesson observations, teachers based groups on the class seating plan or gender rather than the aforementioned criteria. The teacher who assigned groups by gender thought that was instructed and gender-separated groups would help demonstrate the female pupil's ability to come up with good ideas. Lack of adherence to allocation recommendations may have influenced whether the groups represented a range of talents and aptitudes as specified in the intervention instructions.

"I wanted to prove to the girls that the girls were better than the boys. I think, largely, as a whole, across the year groups - across the classes - the girls came up with better ideas." (S02: Teacher interview)

Use of resources

Resources included: (1) teacher slides for teachers to present to pupils at the start of the lesson, which they were free to to *'Adapt — Remix, transform, and build upon'*, (2) lesson primers which provide guidance on how to teach the lesson, (3) lesson plans which provide a five-minute visual lesson plan, (4) a pupil checklist which is a one-page table where pupils can tick off completed tasks and (5) a pupil workbook and worksheets for pupils to document their learning.

Teacher slides

The use of slides varied from not skipping any slides to skipping and condensing some slides or slide decks. This variation was related to the lesson time teachers had available to deliver the content and session activities. As mentioned earlier, some lesson primers, which included suggested activity timings, totalled one hour, but not all teachers had one-hour lessons and some thus skipped slides to fit one session into their lesson. For example, one teacher who completed the survey reported that they had 40 minutes to deliver

the session and therefore skipped and condensed slides to ensure they conveyed the main points of the session. Other teachers reported not skipping any slides.

Teachers chose to skip slides which they perceived to be less important or when they found them difficult to understand themselves. However, perceived importance and difficulty depended on the teacher, meaning teachers chose to skip different slides and thus pupil access to content may have differed across schools. Teachers skipped or removed content, such as videos, which they didn't consider 'core content'. An interviewed teacher explained they skipped repetitive slides, such as the 'objectives and success criteria' slides, as they believed the content had already been covered in previous slides. Another teacher explained that they skipped the cognitive surplus and crisis management app part of session 5 because they lacked a good understanding of the concepts themselves.

"I skipped a few of the slides and videos - not what I considered core content.'" (S07: Teacher survey)

Teachers were free to adapt the slides, and survey respondents reported making minor adjustments to the slides to ensure consistency with school format and learning values. For example, during a lesson observation, we noticed the teacher had changed the slides to match the style and format of the school's typical slides as well as adding memory retrieval tasks for learning purposes. Additionally, a surveyed teacher reported adding content, like 'Success Criteria', to maintain consistency with the schools learning expectations.

Worksheets

Teachers used the workbooks in different ways to meet the needs of the class. The use of workbooks varied from always completing all tasks in the workbook, to sometimes or never completing the task in the workbook. For example, during a lesson observation, we saw pupils completing a worksheet in a workbook which the teacher had printed and provided. Other interviewed teachers reported not using the worksheet or using them in some, but not other lessons. One teacher reported asking pupils to think about the questions on the worksheets instead of writing answers because they wanted to increase the time for practical tasks. The variation in the uses of the worksheets suggests that teachers decided what activities were most valuable to their class and focused on those.

"I didn't do so much recording in the student workbook type things. I tended to say to them, 'Right these are here as leading questions to get you to intelligently think about what you're trying to do rather than as worksheets, everybody's filling this in for five minutes, because I want to try and maximise the time they're actually talking and doing, rather than writing about it." (S01: Teacher interview)

App Lab

Teachers consistently reported using App Lab, the platform provided to teachers for pupils to build and develop their apps on, as intended. Interviewed teachers consistently referred to using the tool and provided positive feedback in general (see 'Quality' section).

Online training

Completion rates of survey within the online training, shared by RPF, indicated that 72% of schools in the treatment group had at least one member of staff who completed the online training[53]. Moreover, according to the teacher survey, completion rates of all three online training sessions were high. The training sessions consisted of three tutorials to be completed independently. Twenty-four out of 26 teachers completed all three online sessions, and 2 out of 26 teachers completed one online session. The two teachers who reported not completing all the training sessions and one interviewed teacher referred to time constraints as the main reason for not completing the training.

"If I'm honest, I didn't have the time to look at them." (S03: Teacher interview)

4.1.2 Feasibility

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

Key findings:
- Pupil computing ability and comfort with English affected the ease with which teachers could implement the intervention.
- Teachers with more programming experience reported that they found it easy to implement the intervention, while their colleagues with less experience would have preferred more support with the advanced coding sections of the intervention.
- Practical considerations such as timetabling and the volume of content led to difficulties delivering the content within the allocated lesson time.

Pupil ability

Computing

Pupils' computing ability and level of difficulty influenced how easy it was for teachers to deliver sessions in the expected timeframe. Teachers who taught classes with mixed-ability pupils highlighted that lower-ability pupils required more support. For example, one interviewed teacher mentioned that the lower-ability groups found it difficult to come up with their own ideas and needed help with this stage of the intervention. A teacher who completed the survey explained that this, alongside the resources being too challenging, led to sessions taking longer than expected to deliver.

"The lessons were planned very well but often took more than one lesson to deliver with mixed ability groups. Extra time was needed at points to allow pupils to investigate theory or practical tasks in sufficient depth to secure their understanding before moving on." (S12: Teacher survey)

---

[53] Fifty schools had at least one teacher who completed the training. Given that 85 schools were originally in the treatment group and 16 schools formally withdrew, the 72% figure is based on the 69 schools that did not formally withdraw. RPF monitored the proportion of completion of online training using a survey embedded in the last section of the training, and measured the completion at the school level rather than at the teacher level.

"Most of my students apart from the top 10% who are gifted and talented in this subject found the resources too challenging. We are also in an area where students have very low aspirations and the behaviour is quite challenging. So the time that I took moving through the slides was a lot longer than anticipated." (S04: Teacher survey)

Moreover, one teacher who completed the survey couldn't deliver the intervention because the pupils found the content too difficult. They explained that they decided to stop teaching the course as it seemed too complicated for pupils.

"[I decided] to stop early as the course seemed to be too complicated for students." (S14: Teacher survey)

Language

Pupils whose first language isn't English required more support from the teachers. In one school where a higher percentage of pupils whose first language isn't English, an interviewed teacher reported that pupils found the lessons more difficult to understand when the content was information-heavy. Consequently, pupils required more support from the teachers to complete tasks set in the intervention.

"If you've got pupils who don't speak English as a first language, it can be a struggle for them. You find that we're doing more of the legwork just to help them." (S04: Teacher interview)

Teacher programming knowledge

Teacher programming knowledge and experience influenced how satisfied teachers were with the level of support. During teacher interviews, teachers who had ample experience teaching programming or had worked as a programmer before teaching found the content easy to deliver and the level of support satisfactory. In contrast, teachers with less experience expressed that more support or more time allocated to learning the advanced programming content would have been useful.

"With the programming aspect I think I've got a lot more confident with them because I teach it consistently. If I'm not teaching it with Key Stage Three, I'm generally teaching it with year 10s and year 11s, so that's increased my confidence a lot more when I'm teaching it" (S04: Teacher interview)

"I think [I would have liked] more support on the actual app, basically, of the block-based programming, and not so much the block-based programme, the easy stuff, the more advanced stuff." (S03: Teacher interview)

"Teachers without programming knowledge will struggle delivering [the intervention]." (S24: Teacher survey)

Time and practical challenges

Timetabling

Lesson length and frequency varied across schools, which influenced how easy it was to deliver one session per lesson and the 11 core lessons. As mentioned in the 'Fidelity' section, the session activities in the lesson primer totalled one hour. Some schools had as

little as 35 minutes to deliver the lessons and others only had a lesson every other week. For example, one surveyed teacher only had 35 minutes to deliver each session and it thus took 15 weeks to deliver the intervention. Another teacher only had one session every fortnight and thus could only deliver eight sessions over 14 weeks.

"We only have a [computer science] lesson once a fortnight for our Y8 pupils as the school operates a 2 week timetable." (S22: Teacher survey)

Volume of content per lesson

Teachers found it challenging to deliver the volume of the content in one lesson. Some teachers expressed that there was too much to achieve in every lesson. For example, an interviewed teacher expressed that it was challenging to get through all the different types of material in 55 minutes:

"To try and do all the booklet in the lesson, do all the paper-based activities, watch all the videos, for me, I found it difficult to get it done in a 55-minute lesson." (S05: Teacher interview)

"There was too much to do in each lesson so we did parts of it and left other bits." (S21: Teacher survey)

As mentioned in the 'Fidelity' section, some teachers used homework as a way to reduce the amount of content to cover in each session. However, one surveyed teacher highlighted that they were not able to do this because some pupils lacked access to technology at home and this would thus have disadvantaged those pupils.

"Access at home is varied for many pupils meaning that extended time required could not be set as homework as it would disadvantage pupils in the groups." (S12: Teacher survey)

Other practical challenges

COVID-19 related and practical challenges were referred to as obstacles to delivering the 11 core sessions. Staff sickness with COVID-19 led to missed lessons; for instance one teacher had to isolate for two weeks and thus missed two lessons. It is worth noting that this is likely to have affected other schools too as the timing of the implementation coincided with a spike in cases of COVID-19 related teacher absences.[54] Other practical challenges including lessons lost to extra-curricular challenges, late starts and the need to get back to the school curriculum, all contributed to the inability to deliver the 11 core sessions.

"We have a shorter term being an independent school at 10-11 weeks (depending on the class). Also, some sessions were disrupted by staff sickness with COVID." (S20: Teacher survey)

"We started it a bit late. Also, there was a need to return to our own Y8 KS3 curriculum." (S03: Teacher survey)

---

[54] See GOV.UK (2022). *Attendance in education and early years settings during the coronavirus (COVID-19) pandemic, Week 30 2022*. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak

4.2 Intervention

4.2.1 Quality

This section explores teachers and pupils' perceptions of the quality of the Relevance intervention resources.

> **Key findings:**
> - Teachers thought that the resources were useful, but highlighted different preferences for the style and structure of the slides and the online accessibility of the worksheets.
> - Teachers thought the quality of the online resources (App Lab and Code.org) was high apart from the inability to work simultaneously on App Lab.
> - Teachers generally thought the training prepared them well for the intervention.

Teacher slides

In general, teachers thought the teacher slides were of good quality but some highlighted minor suggestions for improvement with the chronology and content. Some teachers who completed the survey reported that they would reuse the content and one interviewed teacher reported they would keep 90% of the intervention the same. Other interviewed teachers identified minor issues including: thinking that the style could have been more engaging; repetition across slides; and finding the chronology of the slides difficult to follow. However, teachers were free to adapt the slides.[55]

"The slides are quite nice." (S01: Teacher interview)

"The resources were good, however, there was too much theory to cover that was above the ability of Year 8 students." (S02: Teacher survey)

"It just seemed to be a lot of repetition about functions in that certain slide. The earlier slides about creating an app, again, it went from one extreme - it went from the basics, which were really good, to the extreme." (S03: Teacher interview)

Worksheets

Teachers found the worksheets difficult to use digitally and time-consuming to fill out. In classes where teachers conducted all their work online, pupils found the worksheets difficult to use because they could not easily fill in the text boxes. Interviewed teachers also reported spending time adapting the worksheets so pupils could input data into the boxes. In comparison, teachers who provided paper worksheets to pupils in lessons didn't mention any technical issues with the worksheets but did mention that they took a long time to fill out.

---

[55] At the end of each session slide deck, the last slide stated that teachers were free to '*Adapt —– Remix, transform, and build upon the material*'.

"Unless you actually clicked on the text box, you couldn't actually write in the text box, so that in itself caused frustration." (S03: Teacher interview)

"It meant that I had to go through these things and make them more user friendly to be filled in online." (S02: Teacher interview)

"Filling in the entire booklet, sometimes to get them to fill in all those activities would have taken up most of the lessons" (S05: Teacher interview)

Online resources

App Lab

As mentioned in the 'Fidelity' section, feedback on App Lab was generally positive. However, teachers found group work challenging when using App Lab because it didn't allow for multiple pupils to code simultaneously. Interviewed teachers thought App Lab was a good tool and easier to use than other app making tools (e.g. Google's App Maker). However, teachers consistently reported that it was frustrating that pupils could not work on the code at the same time. Although one teacher developed a work-around by developing a master copy of the code and then cloning it, teachers reported a preference for a tool that could allow the whole group to work on an app at the same time.

"We used to do app building before with another tool and this is slicker, less clunky." (S01: Teacher interview)

"There were issues with the app tool because of course they all wanted to be able to simultaneously, they wanted three people to be able to code the app at the same time" (S01: Teacher interview)

"Once again, nice tool. It would have been nicer if more than one person could work on it at the same time" (S02: Teacher interview)

Code.org

Tutorials on Code.org, the platform which provides 'App Lab', were good resources and helped the pupils to learn how to code. The session 2 lesson primer provided a link to "Recommended courses" on Code.org for teachers. Both interviewed teachers and pupils thought that Code.org was a useful tool. For example, in one lesson observation the pupils explained that they struggled with coding and that using the YouTube videos from Code.org helped with their learning. Teachers also referred to how easy it was to use Code.org and one expressed how it allowed them to learn to code at their own pace.

"I introduced them to ways they can learn how to code by going on to the YouTube tutorials and working at their own pace and giving them their headphones, everything just really took off and the interest really grew." (S05: Teacher interview)

"Code.org was quite easy" (S03: teacher interview)

Lesson guidance (primers and plans)

Teachers tended to use teacher slides to prepare for the lessons instead of the lesson primers and plans. Interviewed teachers described looking at the lesson slides before the lessons rather than reading the lesson primers or lesson plans because they found the slides sufficient for lesson preparation. In contrast, the lesson primer was a good resource when it was used because it contained links to useful documents.

"I watched a video, just normally and looked through the slides and tried to interrelate it to my own experience." (S02: interview)

"I looked at the PowerPoint presentations, the pupils' worksheets, and went from there, because to me that was what I needed." (S03: Teacher interview)

"[The lesson primer is] quite nice because it's one document to open that opens everything rather than finding all the bits." (S01: Teacher interview)

Online training

Teachers generally thought the training prepared them well for the intervention. Out of 28[56] teachers who completed the online survey, 23 thought the training prepared them 'Well' and five out of 28 'Not well'. During teacher interviews, teachers also reported that it prepared them well, but one mentioned that they already knew 75% of the content and another that it took longer than expected.

"It took more than four hours, definitely, but I think it really prepared me for the sessions that I needed to teach." (S05: Teacher interview)

Some teachers struggled to recollect doing the training or parts of the training content. During teacher interviews, one teacher mentioned that they completed the training but couldn't remember it. However, it is worth noting most interviews with the teachers took place more than 10 weeks after they had started implementing the intervention, which may have contributed to the lack of recollection.

Teachers expressed a preference for a hands-on approach and more time for learning how to use App Lab. As mentioned in the 'Feasibility' section, one interviewed teacher, who hadn't used an app making tool before, would have appreciated more time to familiarise themselves with App Lab. Another expressed that their preferred learning style was to learn by doing and looking through the lesson content. It is worth mentioning that RPF had originally planned an in-person one-day training course, which could have provided teachers with more hands-on support, but had to adapt due to the COVID-19 context.

"When I was looking at AppLab, my colleague next-door was looking at it as well, we like to get hands-on on to it, and we were trying it out before reading the information." (S04: Teacher interview)

---

[56] We used the full sample of 28 here because the two teachers who dropped out completed the training.

"It would have been good to have a bit more time set aside for the coding aspect but, yes, it's getting used to the App Inventor. It's not something I've used before, so I would have liked to have some more time with that." (S05: Teacher interview)

4.2.2 Responsiveness

In this section we explore the extent to which pupils and teachers engaged with the various elements of the Relevance intervention.

Key findings:
- According to teachers, pupil engagement was generally high.
- Pupils responded well to and were excited for the lessons. They enjoyed the practical elements, such as the app building and design, more than the written work.
- Pupils also enjoyed and engaged less with activities or content that were challenging or difficult for them.
- Pupils associated working in groups on intervention activities with perceiving computing as more fun, but also with feelings of frustration when there was an imbalance of team member contribution. This sometimes led to disengagement, team reorganisation by teachers and requiring help from teachers to work effectively as a team.
- Teachers reported that they would recommend the intervention to other schools and that they intended to independently repeat it next year with slight adjustments.

Enjoyment

Pupils generally enjoyed the intervention lessons in comparison to the previous term's computing lessons. During pupil focus groups, when pupils were asked how they found the Relevance lessons, most female pupils reported that they were different, more exciting or more fun. However, some expressed a neutral feeling towards the intervention lesson and one pupil said that the lessons were increasingly monotonous.

"Last term was [more] exciting than before, because before we just used to do work and they used to give us a little test or something and these ones are more fun than the other ones." (S01: female pupil)

"They were more fun than the other ones!" (S02: Female pupil)

"It felt a bit monotonous to me. At the start it was quite cool, I guess, but it got a bit monotonous" (S01: Female pupil)

The type of work (practical versus written) influenced the level of enjoyment. Pupils tended to dislike written work and like practical work. During pupil focus groups, pupils expressed that they didn't enjoy the written tasks such as filling out the worksheets. Although dislike for an activity doesn't necessarily mean it doesn't add value to the lesson, one teacher also expressed that filling out worksheets did not meet pupil expectations who anticipated computer-based work. Another reported thinking that the worksheets hindered female pupil engagement because it resulted in loss of attention and less time for developing their idea.

"[I didn't enjoy] the worksheets, you had to write down everything you had to do, it wasn't much fun."
(S02: Female pupil)

"I feel that they've engaged more when we're doing the practical hands-on as opposed to the written work." (S04: Teacher interview)

"I know the planning is important for their learning, but I always find filling in all those activities sometimes could result in me losing them [the pupils]." (S05: Teacher interview)

In contrast, pupils enjoyed practical and hands-on tasks, especially designing and building the app. Pupils mentioned they thought designing and building the app was fun and expressed that they would have liked to have spent more time on that part.

"The making of the actual thing, like the designing of the app was fun." (S02: Female pupil)

"I think we only had two lessons to design and think about our app and I would like it if we had more time, I guess." (S01: Female pupil)

"I like the designing parts where - because when you add images and stuff, you get to choose from all the different pictures and it's fun." (S04: Female pupil)

Level of difficulty

Abstract concepts

Teachers reported that pupils found abstract concepts difficult to grasp and engage with. During teacher interviews, some teachers explained that some pupils found it difficult to engage in the lessons which focused on theoretical or abstract concepts. For example, pupils 'struggled' with identifying and expressing the problem they wanted to solve or understanding what a 'community' is, which was introduced in session 5.

"I think there needs to be a bit more focus on what they meant by communities. That was quite difficult for the pupils to actually engage with that bit. They didn't really understand what that meant, if you're talking about communities as in their local communities or world community." (S03: Teacher interview)

"Quite a few of them found it difficult to engage in that abstract part of the process. It was quite difficult for them to think about what problem do I want to solve?" (S02: Teacher interview)

Coding

Responsiveness to coding depended on pupils' sense of improvement or perceived complexity. Some pupils expressed that they felt a sense of achievement when overcoming coding challenges or they felt that it became easier over time. In contrast, others disliked coding because they found it difficult and complicated. One pupil explained that this was because they felt they had inadequate knowledge and were unable to easily translate coding skills they had previously learnt on Python to App Lab as they were both very different.

"After accomplishing something or finally mastering tough code it can feel like an achievement and it's quite rewarding." (S01: Female pupil worksheet answer)

"I think at first, it was a little bit challenging with the coding and everything, but as time went by you got used to it, so it got easier." (S05: Female pupil)

"I don't really like coding, and it's not very fun to put stuff together. It takes quite a long time." (S02: Female pupil)

"I don't like coding. I would have liked it, I don't know, if it wasn't complicated." (S04: Female pupil)

"Sometimes you have a feature in your app which you're not really sure how to make the coding. So it gets really hard when you're trying to find it online but the answers may not be there, because we were taught a bit of coding, but not enough for us to fully take advantage of it and use it in this term." (S01: Female pupil)

Group dynamics

Teamwork

Responses to teamwork varied from increasing enjoyment of the lesson to frustration due to unequal contribution to the workload. During focus groups, pupils explained that they enjoyed teamwork, collaborating with people with different skills and with their friends during computing this term, suggesting that working collaboratively contributed to their engagement with the sessions.

"With what we've done this term you're around your friends more in it, so you can speak to them and get ideas from them and ask for help and help them." (S02: Female pupil)

"I enjoy teamwork. Like we can actually discuss stuff. Like, today, we're like designing stuff, so I like that." (S04: Female pupil)

"It was really fun to work together as a team, because we have a designing team which designed the logo and we had a coding team, and that's really good because some people have different strengths." (S01: Female pupil)

However, one teacher reported in the survey that female pupils didn't enjoy working in groups as not all group members contributed to the workload equally. Pupils also reported frustration about uneven levels of contribution. Yet this may be a common issue with group work and not specific to the intervention.

"It's not really to do with computers, but it's probably the whole being in groups. So sometimes it didn't really work out as well as we hoped because everyone had different ideas, and then to put it all into one app, it was quite difficult to make it all work together." (S01: Female pupil)

"I think there were some people who didn't contribute as much as others and it was a bit annoying." (S01: Female pupil)

Teachers found that pupils responded well to the freedom to choose to work together or in a team. Some teachers explicitly instructed pupils to work together or separately;

others gave them the choice to work independently on developing and designing their app For example, in two lesson observations, pupils within each team were allowed to each create a design for their user interface rather than creating just one design.

"They all worked as a team, but in terms of building their projects, they had a choice, and most of them chose to use the ideas that they developed as a group, but they built individual projects following the same idea." (S05: Teacher interview)

Lack of group cohesion resulted in pupils feeling left out or pupils dominating specific tasks. In teams where arguments occurred this led to pupils feeling excluded. When this occurred later on in the intervention, it made it difficult for teachers to rearrange groups to mitigate pupil exclusion, as team members had already invested time in their idea. Teachers described that this led to pupil disengagement, required reorganising or creating new groups, or having to teach groups about effective teamwork. However, it is worth noting that this is a challenge that is likely inherent to pupils working together and thus not necessarily specific to the intervention.

"So I added a little bit of, during this lesson can you say what everybody in the group is going to be doing and get them all to do it." (S01: Teacher interview)

Teacher recommendation and intention to repeat the intervention

Interviewed teachers reported that they had enjoyed delivering the sessions to their pupils and would recommend the intervention to other schools. Whilst some interviewed teachers would recommend the intervention as it stands, others would recommend the intervention if minor adaptations were made, such as including fewer lessons.

[when asked if they would recommend the programme to other schools] "Yes, absolutely. Yes." (S02: Teacher interview)

"I think it's really engaging because we want to know why females don't choose computing, so I think it'd be really good in other schools so they could find out why, [...]. I wouldn't have as many lessons as we've been given, because I just find it disengaging for some of our students." (S04: Teacher interview)

Moreover, a teacher explained that they would encourage other schools to consider their school context. They highlighted that they would ask other schools to consider whether they would be able to deliver the intervention effectively within their school.

"I would tell other schools about it and ask them to go do the training, look at the lessons and then decide if they think the context of their school would allow them to be able to deliver it effectively in ten weeks, just because it's down to individual schools." (S05: Teacher interview)

Some teachers reported intention to use some variation of the intervention again. Some teachers indicated that they would run the intervention again with minor adaptations and provided positive feedback on the intervention. Teachers also suggested that shorter units and the ability to fit the intervention into a timetable block would increase their intention

to deliver the intervention again. This aligns with the difficulties with fitting the sessions into one term outlined in the 'Fidelity' section.They did not link this to the intended outcome of the programme, that is, increasing female pupils' participation in computing but rather the enjoyment and quality of the resources.

"I will be adapting the resources and delivering the app project again. Overall, I think the programme was really good and allowed me to link app creation to careers." (S02: Teacher survey)

"I really enjoyed delivering it, and learning new things myself with the pupils and am planning on using the resources again with next years' year 8 groups across the school." (S12: Teacher survey)

"13 weeks is a long project (1/3 of our timetabled lessons in a year!) Shorter units would be easier to integrate into our curriculum and make us more likely to participate in future studies." (S06: Teacher survey)

"Assuming I could fit it into a sensible block, yes, I would definitely do it." (S01: Teacher interview)

4.2.3 Mechanisms

Key findings:
- At case study schools, we found suggestive evidence for some of the mechanisms through which the intervention was hypothesised to increase female pupils' perceived relevance of computing, including: problem selection based on interest, applicability or concern; a sense of agency from selecting their own idea, and an increased perception of computing as creative.
- Other mechanisms were clearly apparent in some female pupils but absent in others, including: a sense of achievement; and recognising the general utility of computing.
- It was unclear that these mechanisms translated into female pupils perceiving computing as relevant to their lives. It is possible that this mechanism was at play for some female pupils through helping them identify the link between computing and issues they care about, but not for others. Teaching style, idea selection process and app examples may have impeded pupils' ability to connect computing to real world issues or problems they personally care about.
- The mechanisms observed did not seem specific to female pupils, and similar themes appeared in male and female pupils (though less data was collected for male pupils).
- Barriers to the working of the hypothesised mechanisms included the schools' pre-existing motivation to increase female pupil engagement in computing, limited pupil confidence in their computing ability, and negative experiences from peers or family members having studied GCSE computer science.

At its broadest level, this intervention aimed to use a 'relevance approach' to increase female pupils' awareness of the real world applications of computing and of its relevance through tangible examples, which would, in turn, improve female pupils' attitudes towards computing

and their intention to continue to study the subject. The mechanisms hypothesised for bringing out this change in pupil perceptions of computing are described in figure 7.

Mechanism A
- Idea generation links day to day experience to subject

Mechanism B
- Refining potential problem - looking at high impact, applies to more people - link made to the wider community/impact in society

Mechanism C
- Problem chosen - pupils commit to idea as something they care about - space to select their own idea gives them sense of agency - they can exact change

Mechanism F
- Solution generation - learn that computing is creative/the utility of computing as a tool towards an important goal - fulfilling a need for people like me

Behaviours
- Female pupils see that learning about computing can help people
- Female pupils want to find out more
- Being absorbed - seeing it is engaging
- Notice more computing around them (their environment)
- More likely to see it through
- Excited for the next computing lesson
- Talk about it at home - engagement outside of programme
- Grit/resilience - investing more time in it voluntarily
- Reconsider it as a creative process
- See it as a vehicle for enhancing what they care about
- Dynamics of the classroom change - female pupils more vocal and speaking up in class

Mechanism D
- Connection made - choosing their own project and working on it week to week reinforces the relevance of computing to things they care about

Mechanism G
- Female pupils complete practical exercises - can apply their understanding - realise tech sector is a broad, they can do it, their skills are useful, its a practical subject

Mechanism E
- Sense of achievement from problem solving

Figure 7: Relevance logic model mechanism and proximal outcomes

Mechanisms observed

Mechanisms observed that were hypothesised

Mechanism C: Problem chosen

Mechanism C
- Problem chosen - pupils commit to idea as something they care about - space to select their own idea gives them sense of agency - they can exact change
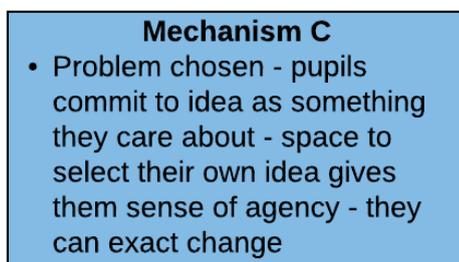
Figure 8: Mechanism C from logic model

In line with the logic model, female pupils tended to pick issues based on a cause they care about, applicability to daily life or general interest. During discussions with female pupils about their choice of app in lesson observations and focus groups, female pupils described their apps. These ranged from: (1) apps which aimed to tackle societal issues such as a sea life population and mental health; (2) apps which were applicable to their daily life such as homework timetabling and public transport and (3) apps based on interest such as a melody-making and interior design.

"I feel like it's an important subject and I feel like sea life is at risk right now and I want to help people realise that." (S05: Female pupil)

A sense of agency was apparent in female pupils due to the freedom of working independently and selecting what to work on. In line with 'Mechanism C', female pupils described feelings of agency from being free to choose what to do. When comparing these lessons to the previous terms', they reported that they had more free range and independence.

'In these lessons it's more like independence where you do what you want to do, but not go outside limits and things, and I found that really nice because you kind of get to do what you think.' (S01: Female pupil)

"I think last term we had instructions and you follow them, whereas now it's like your own ideas and your own creativity and whatever you make." (S02: Female pupil)

Mechanism E: Sense of achievement

Mechanism E
• Sense of achievement from problem solving
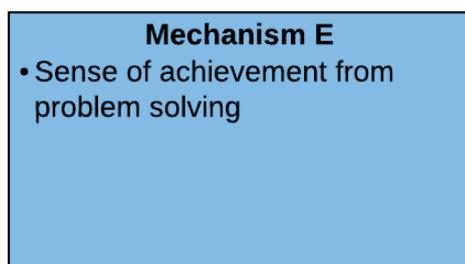
Figure 9: Mechanisms E from logic model

Female pupils' sense of achievement from completing the lessons varied from low to high depending on the pupil and the lesson. Low feelings of achievement were attributed to feeling inferior to other pupils or requiring too much help from the teacher. Female pupils explained that feelings of sense of achievement also varied according to the lesson.

"As sometimes I get help and then I feel like I didn't do it myself so I don't think I achieved something"I think it depends on the lesson." (S02: Female pupil worksheet answer)

"I think it depends on the lesson. Sometimes I feel very satisfied whilst other times I don't think about it at all." (S02: Female pupil worksheet answer)

When a high sense of achievement was apparent, this wasn't always attributed to problem solving and varied depending on the activity. As mentioned in the 'Responsiveness' section and in line with 'Mechanism E', some female pupils felt a sense of achievement when solving problems during coding activities. In contrast, other female pupils felt a sense of achievement when they learnt something new, improved their computing skills or when they had turned their idea into a working app.

"When creating the app it was a bit challenging however when we got to the final stage and our app was working it was really great seeing our hard work lead to something we could use." (S01: Female pupil worksheet answer)

"I feel a sense of achievement in computing when making your ideas a reality makes you proud of your creation which is rewarding" (S02: Female pupil worksheet answer)

Mechanisms F and G: Solution generation and practical exercises

Mechanism F
- Solution generation - learn that computing is creative/the utility of computing as a tool towards an important goal - fulfilling a need for people like me

Mechanism G
- Female pupils complete practical exercises - can apply their understanding - realise tech sector is a broad, they can do it, their skills are useful, its a practical subject

Figure 10: Mechanism F and G from logic model

a) Perceived creativity

A strong theme which emerged was that female pupils reported that they found computing more creative during the intervention. Female pupils reported (1) finding the intervention lesson more creative than previous lessons, (2) enjoying the creative elements of the intervention, such as app design and development, and (3) reporting that they thought computing in general was a creative subject when completing the pupil worksheet.

"It was more creative and we came up with our own ideas." (S01: Female pupil)

"I have to say [that I like] the designing part as well because it's just really creative." (S04: Female pupil)

b) Link to relevance

However, there was no clear evidence of this translating into perceived relevance. Whilst the three points above suggest that female pupils found computing more creative and that they enjoyed the creative aspects, they did not link this to increased relevance. Instead they linked creativity to freedom to choose and design their own ideas in the way they wanted.

"[I think computing is creative] because you can do what you like [..] and create whatever you like." (S05: Female pupil worksheet answer)

"I think computing is creative as we get to design our ideas from the ground up and have creative freedom to adapt and change things along the way." (S02: Female pupil worksheet answer)

c) Improving attitudes

It is possible that increased perceived creativity could improve attitudes towards computing independently of whether it increased perceived relevance. Whilst this mechanism is not entirely separate to '*Mechanism F and G: solution generation and practical*

exercises', points a) and b) suggest creative activities in computing lessons could improve attitudes towards computing independently. Alongside the IPE which found that female pupils engaged well with the creative aspects of the intervention (see the 'Responsiveness'), the pupil survey results found that female pupils' favourite part of the intervention was the design stage, as shown in table 16.

Table 16: Results from survey question on preferred part of the intervention[57]

Question: Which part of the course did you enjoy the most?	Number of times selected (%)
Designing my app	1,472 (60%)
Building my app	342 (14%)
Testing my app	244 (10%)
I didn't take part in these lessons	141 (6%)
Learning about potential careers	129 (5%)
Learning about how computers are relevant	112 (5%)
Total	**2,395**

Additionally, for some pupils GCSE subject choice may be influenced by perceived creativity of a subject. During the pupil focus groups, some female pupils expressed that creativity is an important factor which partially motivates them to pick a subject for GCSE.

"I don't know, certain subjects that require maybe more creativity, I have a pull towards them. I like them more, for example, graphics design, I may not be the best at it but I enjoy it because of creativity." (S01: Female pupil)

[when asked why they preferred IT in comparison to computer science] "Yes, the collaborative work and also just it's creative but on computers and I work decently well around computers and technology." (S02: Female pupil)

The perceived utility of computing varied from recognising the general importance of computing to utility depending on your intended career. To understand pupils' perception of the utility of computing, pupils were asked whether they thought computing was a useful subject to learn about in the pupil worksheet. Female pupils described how computing is applicable to modern life. For example, a female pupil reported that computing is becoming more prevalent and was important during COVID-19 and another that basic knowledge of computing is useful. In comparison, other female pupils explained that computing is a useful subject to learn about only if a job requires computing or you want a career in computing.

"Now there are more things which are based on technology, so having knowledge in computing would be useful." (S01: Female pupil worksheet answer)

"Computing comes up more and more as time goes by and computing is extremely important during 2022 due to covid." (S02: Female pupil worksheet answer)

---

[57] Only asked respondents who reported they had received the intervention.

"Because it's only useful if you want to choose a career in this field." (S01: Female pupil worksheet answer)
"It really depends on what your future is. If you are an IT person you would find it useful but for an artist maybe not so much." (S01: Female pupil worksheet answer)

Mechanism B and D: Refining potential problem and connection made

Mechanism B	Mechanism D
• Refining potential problem - looking at high impact, applies to more people - link made to the wider community/impact in society	• Connection made - choosing their own project and working on it week to week reinforces the relevance of computing to things they care about
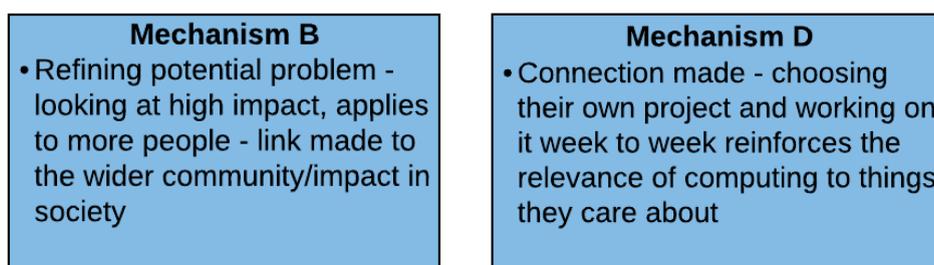
Figure 11: Mechanisms B and D from logic model

It is unclear whether refining and working on a problem to solve via an app every week reinforced the relevance of computing to things they care about and helped to link computing to wider impact in society. Interviewed teachers' perspectives ranged from aligning with the logic model by reporting that perceived relevance increased female pupil interest, to being uncertain as to whether this mechanism was occurring and whether female pupils viewed apps as tools for enhancing social benefit.

"The girls were more engaged when the projects were relevant. Its relevancy was the thing that made the girls far more interested in it." (S02: Teacher interview)

"Whether they really picked up on the 'apps for good' aspect or whether they were just thinking about making an app, there's a lot of stuff that tries to get them to think about the social benefit and that sort of side. Whether they really picked up on that and thinking I'm going to change the world by making an app or whether they're thinking, wouldn't it be fun to have something that plays a bit of K-pop I don't know." (S01: Teacher interview)

Female pupil perspectives varied from thinking computing could be used to solve real world problems to thinking computing can be used to solve technical issues. In responses shared by female pupils in the worksheet activity, three themes emerged: (1) thinking that computing could be used to solve problems related to issues like mental health or fitness; (2) that computing could mainly be applied to fixing technology rather than real-world problems; and (3) that computing could not be used to solve issues that they care about.

"I feel like computing can create apps to do with solving mental health problems which I think are very important and personally need a lot of improvement on the way we can cope with mental health." (S04: Female pupil worksheet answer)

"If your device is broken you can fix it because you know how to do computing." (S05: Female pupil worksheet answer)

"Because I am not going to be solving coding when I go home. It can help to learn about your computer and how it works but I don't think it can solve problems I care about." (S05: Female pupil worksheet answer)

How the teachers taught the content and what they focused on may have influenced whether this mechanism was at play. When reflecting on how they taught the course, an interviewed teacher felt they may have slightly overlooked encouraging pupils to develop an app which had a positive impact. Moreover, as mentioned in the 'Responsiveness' section, perhaps pupil difficulty understanding concepts, like communities and society, may have also hindered their ability to link computing to wider society.

"I know that was probably, if I'm honest, the main crux of the whole app creation, to think about where - what I've learned about gender, the research for girls, they need a real positive to focus their app creation for, and I think that slightly missed the point." (S03: Teacher interview)

The examples of how computing was used to solve real world problems in the session plans may have influenced pupil ability to connect computing to issues they care about. Teachers explained that the examples which were provided could have been more relevant to issues year 8 female pupils encounter every day. For instance, a teacher explained that they wouldn't have used the Ushahidi app[58] as an example and instead would have opted for problems year 8 pupils could relate to.

"I think that could have done something more appropriate to the pupils." (S03: Teacher interview)

"Examples relate more to topics they may be interested in, in more of an engaging way e.g. topical events such as global warming." (S01: Teacher survey)

"More relevant topics of interest to get them engaged. Some discussions/tasks did not need to be dragged out too much so they do not lose interest." (S09: Teacher survey)

How pupils picked their ideas may have also influenced whether they connected computing to issues they care about. For example, an interviewed teacher mentioned providing some pupils with their ideas when they were struggling and to save time. This may have resulted in pupils working on creating a solution to a problem they don't personally care about. Moreover, whilst each pupil may have put forward a problem to tackle, another pupil's problem may have been chosen to focus on by the team. Consequently, pupils may not have been working on solving an issue they each personally care about.

"Some of these ideas that we've got are ones that I've handpicked for other pupils, so I've said this is what we've been doing in this group, and this is what we've been doing in that group." (S04: Teacher interview)

---

[58] A company which helps communities to develop tools and services which enable people to generate solutions and mobilise communities.

Specificity of mechanism to female pupils

Interviewed teachers did not feel that the intervention had appealed to female pupils specifically, and similar themes emerged across female and male pupil focus groups. Pupils of both genders enjoyed creating and designing the app and referred to how computing was practical for later life. However, two of the teachers we interviewed were from all-female schools and we only conducted two male focus groups out of seven; we thus captured a broader range of female than male pupil experiences.

"I actually thought they were very good, and the students were generally engaged regardless of gender." (S22: Teacher survey)

Barriers to mechanisms

Negative experience from people around them

Feedback from friends or family who had taken computer science at GCSE influenced pupils' perspective of the subject. Female pupils referred to feedback from friends and family when asked why they thought they wouldn't enjoy computing if they took it for GCSE.

"Well, my siblings, one of them taking, one at the moment and one of them has, they both don't really enjoy it." (S02: Female pupil)

"I don't really like it because I can remember one of my brothers taking it. He said that it was fun in year 7 and 8 and then it apparently changed to just coding and apparently it's less fun." (S02: Female pupil)

"I have a friend in a higher year who's always telling me about computing and I'm so tired of her because she won't stop going on about it. Apparently it's very, very hard and very difficult and very boring." (S01: Female pupil)

Pupil confidence in computing ability

Perceived difficulty and chance of passing computer science if they took it as a subject influenced whether or not pupils intended to select it and how they felt during computing lessons. Some pupils expressed that they found computing difficult and a key barrier to taking computer science as a subject was the perception that they would not pass it if they did.

"I don't know, I don't know how to describe it, but I find it difficult and it's not something that I feel with that physics, computing, maths, stem stuff. I don't know, they all count poorly, I don't really like that field in general, just because it's just a bit too difficult for me." (S01: Female pupil)

"If I did computing, I'll fail it." (S05: Female pupil)

School context

Interviewed teachers were already trying to engage female pupils in computing. Most interviewed teachers were highly motivated to get more female pupils into computing. They led extra-curricular activities associated with computing and some had taken part in projects that aimed to get more females involved in computing before. This could be an important

indicator of the types of schools we interviewed and the schools in the sample overall, which were motivated and keen to increase the number of females in computing.

"On a Tuesday I do a girls' school Lego robot club, so there's a table over there for Lego robots." (S01: Teacher interview)

"I'm always looking at different opportunities for girls in particular to come into our subject." (S04: Teacher interview)

It is possible that, in these school contexts, where teachers were already actively trying to promote female pupils' engagement with computing, there was less scope for the intervention to improve female attitudes and interest in computing, compared to a school which has not been actively trying to improve female pupils' engagement with computing.

5. Conclusions and recommendations

5.1 Summary and interpretation of findings

Our main analysis did not find a statistically significant impact of the intervention on female pupils' intention to select computer science as a GCSE subject

For our primary outcome of stated intention to select computer science as a GCSE subject, we did not find a statistically significant effect of the intervention, with this result holding for all our prespecificed specifications. The estimated treatment effect is positive, but small in magnitude (the 0.2 percentage point effect size implies that for every 1,000 female pupils, an additional 2 would intend to select computer science as a GCSE subject after the intervention) and not statistically significant.

We conducted additional robustness checks to test the sensitivity of this result to the way pupils were asked about their intention to select GCSE computer science. While it is unclear that this alternative indicator more accurately captures our outcome of interest, this analysis found that when using an alternative outcome based on a slightly different question, the treatment effect is larger (2.9 percentage points in the main specification) and is statistically significant for some, but not all, of the analysis specifications. The differential attrition that led to the small imbalance across the two treatment groups introduces a risk of bias in the estimates for both the main analysis and robustness checks, though it is difficult to determine the possible direction of this bias.

The estimated treatment effect is positive for the secondary and exploratory outcomes related to attitudes towards computing, but it is small and not statistically significant

For our two secondary outcomes (perceived relevance of and interest in computing) and our exploratory outcome (overall attitudes towards computing), we find a small positive difference between the intervention and control group after controlling for covariates (between 0.02 and 0.03 points on a 1-4 scale), but this difference is not statistically significant.

There are indications that not all treatment school pupils received the intervention as intended, which could have limited the intervention's impact

The lack of recollection of participating in a course that involved designing, building and testing an app reported in the endline survey by many pupils across treatment schools suggests that not all pupils who completed the endline survey received the intervention as planned or at all. If this is the case, this may have limited the impact of the intervention across the full treatment sample.

The IPE findings suggest that the intervention was implemented well, though deviations from suggested session activities and several challenges were identified

Teachers were able to deliver the majority of the lessons, but varied in their use of the suggested session activities and adjustments to the resources. The two main implementation challenges were (1) delivering all the 11 core lessons and (2) delivering all the session

activities within the time constraints of a lesson. Other factors such as volume of content to deliver per session, teachers' programming subject knowledge and pupil computing ability and language affected the ease with which teachers could implement the intervention. It is worth noting that given the small number of case study schools, and that teachers who agreed to be involved in case study activities may have been more likely to have implemented the intervention with high fidelity, these teachers are unlikely to be representative of teachers across all treatment schools.

Case studies highlighted that teachers' and pupils' perceptions of the lessons were generally positive, and some of the envisioned impact mechanisms were observed

Teachers reported that they had enjoyed delivering the sessions to their pupils, and that they would recommend the intervention to other teachers. Some intended to repeat the intervention or reuse the resources next year with slight adjustments. Pupil engagement with the lessons was high, particularly for practical elements, such as the app building and design. Teachers thought that the resources were generally useful, however they highlighted a few issues related with online usability and collaborative work. The majority of teachers completed the online training and thought it prepared them well.

We observed some of the hypothesised mechanisms of impact through case study activities, including: problem selection based on interest, applicability or concern; a sense of agency from selecting their own idea; and an increased perception of computing as creative. Other envisioned mechanisms of impact were only apparent for some pupils or not observed, such as perceiving computing as relevant to their lives. Teaching style, the idea selection process and app examples may have contributed to the absence of some female pupils' ability to connect computing to real world issues or problems they personally care about.

Multiple factors could have contributed to the lack of stronger evidence of impact despite positive reported school experiences

There are a number of possible reasons related to the design of the intervention, its delivery, and measurement challenges, for the absence of evidence of a more significant impact despite the largely positive experiences of the intervention reported in case study schools:

- **Dosage and/or partial compliance:** Delivery challenges, such as insufficient time to cover all lessons and all of their content, could have prevented the full implementation of the intervention at all treatment schools. The indicative evidence that many pupils in treatment schools reported not taking part in a course that involved designing, building and testing an app is consistent with this possibility.

- **The hypothesised barrier was only partially addressed, or may not be the most critical to the intended outcome:** At case study schools, there were some indications that while the intervention increased engagement and enjoyment of computing among female pupils, particularly through highlighting its creative elements, this did not fully translate in an increase in the perceived relevance of computing to their lives, and the envisioned mechanisms to increase perceived relevance were not all realised. Alternatively, in the event that the intervention was more successful in increasing perceived relevance of computing at other schools that did not participate in the case study activities, it is possible that lifting this barrier was

insufficient to more meaningfully improve female pupils' attitudes towards computing or intention to study it in the future. Other barriers may include limited pupil confidence in their computing ability and the perception of computer science as a difficult GCSE subject.

- **The large and structural barriers preventing female pupils from choosing GCSE computer science cannot be overcome by a single strategy:** It is possible that this intervention might need to take place alongside other efforts by the school (or possibly even broader social changes) in order to provide sufficient motivation or support for female pupils to overcome the barriers currently preventing them from choosing GCSE computer science. These barriers are reflected in the much lower proportion of female pupils reporting the intention to study GCSE computer science in the baseline survey (7% of female vs. 22% of male pupils).

- **Survey measurement challenges:** As outlined in section 2.1.6, the nature of the outcomes measured and limitations of reported data, including the possible gap between stated intent and observable behaviour and the difficulty of measuring attitudes, imply some challenges in measuring the outcomes of interest for the impact evaluation. While various survey design strategies and robustness checks were implemented as part of the analysis to mitigate these challenges, they may not have fully addressed them, and there is a risk of some measurement error that could limit the evaluation's ability to detect differences between the treatment and control group.

There is not sufficient evidence to confidently determine the extent to which each of these factors may have influenced the evaluation results.

5.2 Recommendations

Recommendations for future use of intervention resources

Despite the lack of clear evidence of significant impact, some of the findings from the evaluation suggest some promise of the Relevance intervention in encouraging female pupils' intention to study computer science and positive attitudes towards computing. Given these results, it could be valuable to:

1. **Offer the lesson resources and training to KS3 teachers who would like to use them**
   Considering that no backfire effects were detected, intervention resources have already been developed, and teachers reported finding the resources useful and the intention to use them again, other KS3 teachers may also find them valuable.

However, to maximise pupil understanding, pupil engagement and ease of delivering the intervention, the recommendations to support implementation and improve potential impact below should first be taken into account. Given the limited impact of this intervention alone suggested by the impact evaluation results, teachers could also be encouraged to adapt and combine these resources with other approaches which they think could increase female interest in computing within the context of their school.

Recommendations to support implementation

The intervention could be made easier to implement in a broader range of schools through the following adaptations:

1. **Make resources easier to use online and collaboratively**
   a. Teachers reported that it was difficult to use the intervention resources online. To overcome this, resource templates could be adapted to be suitable for online platforms, such as Google Docs, such that pupils can easily input text on the computer and teachers do not spend time adapting the content.
   b. Pupils were unable to code simultaneously on App Lab, which made it difficult for all pupils to learn and develop their coding skills. Considering a platform similar to App Lab with this capability could help to alleviate this problem.

2. **Reduce the number of lessons and add 'optional/core' labels to teacher slides**
   a. Teachers did not always have time to deliver all the lessons or deliver the content within each lesson. Reducing the number of lessons or content within each session could enable teachers to fit the intervention within one term (this is what they commonly expressed was the amount of time they allocated to the intervention before they had to return to their school curriculum).
   b. Additionally, labelling lesson content as 'optional/core' to indicate how important the content is could help to (1) signpost which content is necessary to deliver and which can be dropped under time pressure, (2) ensure teachers do not skip core content, (3) reduce pressure on lower ability pupils who are struggling with the workload, yet provide additional activities for pupils who are more confident in computing.

3. **Work with teachers and/or pupils to refine the intervention such that the app examples are more relevant and concepts are easier to understand**
   Teachers reported that some pupils found abstract concepts or processes difficult to understand and that app examples were not relatable to the day to day lives of year 8 female pupils. Working with teachers and/or pupils to refine the intervention and app examples such that they include problems they encounter in their lives could help increase the relevance of the examples provided in the slides.

4. **Allocate more time to practical and creative elements in lesson plans and resources**
   Some teachers reported that if they delivered the intervention again, they would reduce lesson time spent on planning and increase time spent on practical elements. This aligns with some pupils reporting they enjoyed the more practical and creative parts of the intervention, such as designing and developing their app. Therefore, re-allocating time spent on different types of tasks could help increase pupil engagement and improve perceptions of computing as creative and enjoyable. Moreover, considering that some pupils expressed a preference for creative subjects at GCSE, further exploration into the link between creativity and subject desirability as an avenue to increase female pupil engagement with computing could be beneficial.

5. **Make teacher access to additional support easier and more interactive**
   Since, teachers reported using the teacher slides to prepare for lessons, including a link in the first slide of each teacher slide deck could be a strategy to signpost additional support. Moreover, there could also be online 'drop-in' sessions where teachers could ask questions regarding App Lab or programming to an expert in the area. These steps would help address teacher preference for more support with App Lab and hands-on approaches to training. They may also indirectly support teachers to help pupils who struggle with coding by making them more confident in their programming abilities and better equipped to teach and answer pupil questions.

Recommendations for future evaluations

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

1. **Measure the outcomes targeted by the intervention further into the future**
   Tracking relevant behavioural outcomes (in this case, actual GCSE subject choice) after the end of the intervention would require planning, greater collaboration with schools and longer evaluation timelines. However, it would also enable more direct measurement of the behavioural outcome targeted by this intervention (female pupils selecting GCSE computer science), in addition to the shorter-term survey-based indicators used in this evaluation. Setting up systems to facilitate the collection and sharing of this data would also likely have benefits beyond a given evaluation, such as helping to understand pupil trajectories and trends across time and pupils.

2. **Track school implementation of the intervention to assess compliance and estimate its impact when implemented in full**
   Any school-based intervention implemented at a large scale is likely to be implemented with some variability across schools depending on resources and context. However, designing a strategy to measure and track the extent to which schools implement the intervention in full and as intended would enable an evaluation to understand its impact when implemented as planned. This evidence can then be used to inform decisions regarding how much resources to invest in facilitating implementation to achieve this impact across a larger number of schools.

In light of the evaluation findings, this intervention holds some promise as a relatively low-cost strategy to positively affect female pupils' attitudes towards computing and intention to select computer science as a GCSE subject. Addressing the adjustments to the design and delivery of the intervention recommended above could help to improve the effectiveness of this intervention. The lesson resources and training could also be offered to KS3 teachers who would be interested in using them.

In addition, using school administrative data to measure GCSE subject selection and tracking whether schools are able to implement the intervention as planned could help to reduce the cost of data collection and to better understand the longer-term impact of the intervention. Conducting another impact evaluation after the adaptations to the intervention and with these changes to the evaluation strategy could contribute to building the evidence around the

Relevance approach. If the impact among schools that signed up voluntarily is then found to be positive, this could inform decisions regarding a more proactive scale-up of the intervention to a large number of schools. While getting schools currently less motivated or dedicating less resources to encourage female pupils to study GCSE computer science to use these resources may be more challenging than already motivated schools, the potential impact may be greater as well.

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.

Croasdell, D., McLeod, A., & Simkin, M. G. (2011). *Why don't more women major in information systems?*. Information Technology & People.

GOV.UK (2022)*. Attendance in education and early years settings during the coronavirus (COVID-19) pandemic, Week 30 2022.* Available at:
https://explore-education-statistics.service.gov.uk/find-statistics/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak

Guzdial, M., & Tew, A. E. (2006). *Imagineering inauthentic legitimate peripheral participation: an instructional design approach for motivating computing education*. In Proceedings of the second international workshop on Computing education research (pp. 51-58).

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. Available at:
https://csedresearch.org/resources/evaluation-instruments/tool/?id=156

Hulleman, C. S., & Harackiewicz, J. M. (2009). *Promoting interest and performance in high school science classes. science, 326(5958), 1410-1412*.

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

Lewis, C., Bruno, P., Raygoza, J., & Wang, J. (2019, July). *Alignment of goals and perceptions of computing predicts students' sense of belonging in computing.* In Proceedings of the 2019 ACM Conference on International Computing Education Research (pp. 11-19).

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

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

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

Appendices

Appendix 1: Quantitative survey including SCSAS[59]

[SURVEY START]

Hi! Thank you for taking part in this survey.

If you have any questions about this survey or the study please contact project_i4@bi.team before starting the survey.

If you are ready to begin the survey...

Please complete the first section, check that all personal information is correctly typed and accurate, and click the "next page" button.

[NEW PAGE: PAGE 1]

1	Please type your first name	Text entry		
2	Please type your last name	Text entry		
3	Please select the gender you identify with	Female	Male	Non-binary/ Other
4-6	Please select the day you were born/month you were born/year you were born[60]	Drop down selection boxes		
7	Please pick the name of your school from the list below[61]	Drop down selection boxes		
8	Do you intend to choose computer science as one of your GCSE subjects?	Yes	No	Don't know

---

[59] The survey was identical in its entirety at baseline and endline and the only difference in surveys between the control and treatment groups was the list of schools to select from in question 7
[60] A limited list of the years 2008-2012 and an additional text box below if the year was not in this list.
[61] An additional text box was provided below to capture any schools that weren't in the list.

[NEW PAGE: PAGE 2]

9	Do you intend to choose any of these GCSE subjects? *Please select all that apply*	Checkbox multiple choice question (MCQ)	
9.1	Chemistry	Checked	Unchecked
9.2	Music/Art/Drama	Checked	Unchecked
9.3	Technology	Checked	Unchecked
9.4	Engineering	Checked	Unchecked
9.5	Maths	Checked	Unchecked

[NEW PAGE: PAGE 3]
Thanks! Now it's time for the rest of the questions.

[NEW PAGE: PAGE 4]
[**Not shown to pupils: 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.

10.1	I am confident that I can do computing	Strongly disagree	Disagree	Agree	Strongly Agree
10.2	I am confident that I can solve problems by using computing	Strongly disagree	Disagree	Agree	Strongly Agree
10.3	I can learn computing skills without much help	Strongly disagree	Disagree	Agree	Strongly Agree
10.4	I am good at solving hard questions in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
10.5	I think I will do well in computing	Strongly disagree	Disagree	Agree	Strongly Agree

[NEW PAGE: PAGE 5]

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

11.1	I would choose more computing lessons if I could	Strongly disagree	Disagree	Agree	Strongly Agree
11.2	In the future I'd like to do more computing	Strongly disagree	Disagree	Agree	Strongly Agree
11.3	I like to use computing to solve problems	Strongly disagree	Disagree	Agree	Strongly Agree
11.4	Solving questions in computing lessons makes me feel happy	Strongly disagree	Disagree	Agree	Strongly Agree
11.5	I like computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree

[NEW PAGE: PAGE 6]

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

12.1	I feel happy in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
12.2	I feel like I belong in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
12.3	I have lots of friends in my computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
12.4	I know someone who uses computing in their job	Strongly disagree	Disagree	Agree	Strongly Agree
12.5	I have friends who think computing is interesting	Strongly disagree	Disagree	Agree	Strongly Agree

[NEW PAGE: PAGE 7]
Great job you're over halfway!

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

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

[NEW PAGE: PAGE 8]
How much do you agree or disagree with the following statements? There are no right or wrong answers.

14.1	A friend, or someone I know said I should do computing	Strongly disagree	Disagree	Agree	Strongly Agree
14.2	Someone I know has made me feel interested in computing	Strongly disagree	Disagree	Agree	Strongly Agree
14.3	Someone I know has said my work in computing is good	Strongly disagree	Disagree	Agree	Strongly Agree
14.4	I have been taught about how computing is used outside of lessons	Strongly disagree	Disagree	Agree	Strongly Agree
14.5	Someone in my family has made me want to learn computing	Strongly disagree	Disagree	Agree	Strongly Agree

[NEW PAGE: PAGE 9]
Almost done!

15	Have you recently taken part in a course which involved designing, building and testing an app?	Yes	No	Maybe

[NEW PAGE: PAGE 10 - Only shown to those who answer "Yes" to 15]

16	Which part of the course did you enjoy most?	Radio List Question (can only select one)	
	Building my app	Checked	Unchecked
	Learning about how the computers are relevant in my life	Checked	Unchecked
	Learning about potential careers	Checked	Unchecked
	Testing my app	Checked	Unchecked
	Designing my app	Checked	Unchecked
	I didn't take part in any of these lessons	Checked	Unchecked

[SURVEY END]

Appendix 2: Model specification

Primary outcome: Intention to study computer science at GCSE level

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

$$Y_{is} \sim bernoulli(p_{is}); logit(p_{is}) = \alpha \; + \; \beta_1 T_i \; + \; \beta_2 BL_i \; + \; \beta_{3prop} FSM_s \; + \; \beta_4 Ofsted_s \; + \; \beta_{5prop} Girls_s \; + \; \epsilon_{is}$$

Where:

- Y_{is} is a binary indicator for student *i* reflecting intention to study computer science in school *s*

- p_{is} is the probability of a positive intention for student *i* in school *s*

- α is the constant

- $\beta_1 T_i$ is a binary indicator of treatment assignment for student *i in school s*

- $\beta_2 BL_i$ is the baseline SCSAS score for student *i* in school *s* collected before the intervention

- $\beta_{3prop} FSM_s$ is the proportion of pupils eligible for FSM in school *s*

- $\beta_4 Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) "Outstanding"; (ii) "Good"; and (iii) "Below good" (the combination of "Requires improvement" and "Inadequate")

- $\beta_{5prop} Girls_s$ is the proportion of female pupils in school *s*

- ϵ_{is} is the error term for student *i* in school *s*

Table 17 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 17: Logistic regression coefficients for primary outcome (standard errors in parentheses)

Outcome: Intention to study computer science	(1) MI	(2) Miss. Ind.	(3) CCA
Treatment group (reference category is control)			
Intervention	0.028 (0.151)	0.2 (0.158)	0.016 (0.158)
Baseline SCSAS score	3.142** (0.201)	3.247** (0.205)	3.200** (0.207)
Ofsted rating (reference category is Outstanding)			
Good	0.047 (0.227)	-0.096 (0.308)	0.028 (0.228)
Below Good	-0.401 (0.277)	-0.363 (0.300)	-0.448[+] (0.269)
Missing	-0.240 (0.345)	-0.256 (0.379)	-0.392 (0.378)
Proportion FSM	-0.001 (0.008)	-0.001 (0.009)	-0.007 (0.008)
Proportion of female pupils	0.517 (0.406)	0.610 (0.463)	0.675 (0.491)
Constant	-11.169** (0.671)	-11.516** (0.690)	-11.254** (0.690)
Control group mean	0.076	0.076	0.078
Observations	3,937	3,937	3,038
R^2	-	0.197	0.235

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

Secondary outcome: SCSAS scores - Usefulness Subscale - 4-point mean score

The secondary outcome is continuous and therefore we will use 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 will use cluster-robust standard errors in analysis, clustering at the school level.

$$Y_{is} = \alpha + \beta_1 T_i + \beta_2 BL_i + \beta_{3prop} FSM_s + \beta_4 Ofsted_s + \beta_{5prop} Girls_s + \epsilon_{is}$$

Where:

- Y_{is} is the total Usefulness subscale survey mean score for pupil *i* in school *s*

- α is the constant

- $\beta_1 T_i$ is a binary indicator of treatment assignment for student *i in school s*

- $\beta_2 BL_i$ is the baseline SCSAS score for student *i* in school *s* collected before the intervention

- $\beta_{3prop} FSM_s$ is the proportion of pupils eligible for FSM in school *s*

- $\beta_4 Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) "Outstanding"; (ii) "Good"; and (iii) "Below good" (the combination of "Requires improvement" and "Inadequate")

- $\beta_{5prop} Girls_s$ is the proportion of female pupils in school *s*

- ϵ_{is} is the error term for student *i* in school *s*

Only female pupils will be used in this analysis.

Table 18 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 18: Linear regression coefficients for secondary outcome Usefulness (standard errors in parentheses)

Outcome: Usefulness subscale score	(1) MI	(2) Miss. Ind.	(3) CCA
Treatment group (reference category is control)			
Intervention	0.027 (0.024)	0.031 (0.026)	0.028 (0.025)
Baseline SCSAS score	0.830** (0.022)	0.835** (0.024)	0.838** (0.025)
Ofsted rating (reference category is Outstanding)			
Good	0.027 (0.029)	0.049 (0.040)	0.035 (0.029)
Below Good	0.002 (0.043)	0.017 (0.053)	0.006 (0.046)
Missing	0.035 (0.045)	0.050 (0.057)	0.030 (0.051)
Percentage FSM	0.001 (0.001)	0001 (0.001)	0.002 (0.001)
Proportion female pupils	0.126[+] (0.068)	0.173 (0.079)	0.168 (0.080)
Constant	0.340** (0.072)	0.278** (0.082)	0.286** (0.078)
Control group mean	2.48	2.48	2.48
Observations	3,937	3,877	3,003
R^2	-	0.320	0.386

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

Secondary outcome: SCSAS scores - Interest Subscale - 4-point mean score

The secondary outcome is continuous and therefore we will use 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 will use cluster-robust standard errors in analysis, clustering at the school level.

$$Y_{is} = \alpha + \beta_1 T_i + \beta_2 BL_i + \beta_{3prop}FSM_s + \beta_4 Ofsted_s + \beta_{5prop}Girls_s + \epsilon_{is}$$

Where:

- Y_{is} is the total Interest subscale survey mean score for pupil *i* in school *s*

- α is the constant

- $\beta_1 T_i$ is a binary indicator of treatment assignment for student *i in school s*

- $\beta_2 BL_i$ is the baseline SCSAS score for student *i* in school *s* collected before the intervention

- $\beta_{3prop}FSM_s$ is the proportion of pupils eligible for FSM in school *s*

- $\beta_4 Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) "Outstanding"; (ii) "Good"; and (iii) "Below good" (the combination of "Requires improvement" and "Inadequate")

- $\beta_{5prop}Girls_s$ is the proportion of female pupils in school *s*

- ϵ_{is} is the error term for student *i* in school *s*

Only female pupils will be used in this analysis.

Table 19 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 19: Linear regression coefficients for secondary outcome Interest (standard errors in parentheses)

Outcome: Interest subscale score	(1) MI	(2) Miss. Ind.	(3) CCA
Treatment group (reference category is control)			
Intervention	0.024 (0.029)	0.038 (0.035)	0.024 (0.029)
Baseline SCSAS score	0.968** (0.020)	0.973** (0.022)	0.970** (0.022)
Ofsted rating (reference category is Outstanding)			
Good	0.027 (0.035)	0.044 (0.308)	0.006 (0.324)
Below Good	0.076 (0.041)	0.099+ (0.048)	0.055 (0.634)
Missing	0.012 (0.050)	0.036 (0.063)	-0.013 (0.441)
Percentage FSM	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Proportion female pupils	0.087 (0.068)	0.095 (0.082)	0.081 (0.077)
Constant	-0.237** (0.070)	-0.293** (0.084)	-0.223** (0.075)
Control group mean	2.19	2.20	2.19
Observations	3,937	3,919	3,026
R²	-	0.373	0.212

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

Exploratory outcome: SCSAS scores: Total score - Four point mean scale

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

$$Y_{is} = \alpha + \beta_1 T_i + \beta_2 BL_i + \beta_{3prop} FSM_s + \beta_4 Ofsted_s + \beta_{5prop} Girls_s + \epsilon_{is}$$

Where:

- Y_{is} is the total SCSAS survey mean score for pupil *i* in school *s*

- α is the constant

- $\beta_1 T_i$ is a binary indicator of treatment assignment for student *i in school s*

- $\beta_2 BL_i$ is the baseline SCSAS score for student *i* in school *s* collected before the intervention

- $\beta_{3prop} FSM_s$ is the proportion of pupils eligible for FSM in school *s*

- $\beta_4 Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) "Outstanding"; (ii) "Good"; and (iii) "Below good" (the combination of "Requires improvement" and "Inadequate")

- $\beta_{5prop} Girls_s$ is the proportion of female pupils in school *s*

- ϵ_{is} is the error term for student *i* in school *s*

Only female pupils will be considered in this analysis.

Table 20 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 20: Linear regression coefficients for exploratory outcome total SCSAS score (standard errors in parentheses)

Outcome: Total SCSAS	(1) MI	(2) Miss. Ind.	(3) CCA
Treatment group (reference category is control)			
Intervention	0.028 (0.022)	0.039 (0.027)	0.027 (0.022)
Baseline SCSAS score	0.823** (0.016)	0.832** (0.018)	0.836** (0.018)
Ofsted rating (reference category is Outstanding)			
Good	0.014 (0.026)	0.031 (0.040)	-0.000 (0.023)
Below Good	0.031 (0.034)	0.051 (0.048)	0.016 (0.032)
Missing	0.018 (0.036)	0.037 (0.050)	-0.002 (0.001)
Percentage FSM	-0.002+ (0.001)	-0.001 (0.001)	0.006 (0.014)
Proportion female pupils	0.112 (0.047)	0.120 (0.061)	0.125* (0.054)
Constant	0.309** (0.054)	0.260* (0.065)	0.288** (0.054)
Control group mean	2.38	2.38	2.38
Observations	3,937	3,937	3,038
R²	-	0.461	0.558

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

Appendix 3: Online teacher feedback survey

Background and Consent

Thank you for participating in the GBIC Relevance project!

We'd be really grateful if you could provide some quick feedback on how the project went.

Your responses will be completely anonymous, so please answer as honestly as you can.

1. This survey needs to be completed by the teacher who has delivered the GBIC Relevance lessons in your school - please confirm whether this applies to you [check box]: *

◯ Yes, I am a classroom teacher who has delivered the GBIC Relevance lessons in my school

◯ No, I am NOT a classroom teacher who has delivered the GBIC Relevance lessons in my school

Your experience with the GBIC Relevance training

2. Did you complete the online training for the GBIC Relevance programme? Please tick one box.

◯ I completed all three online sessions (Part 1: Year 8 'Relevance' training - Introduction, Part 2: Year 8 'Relevance' training - Getting started, Part 3: Year 8 'Relevance' training - Investigating App Lab)

◯ I completed one of the online sessions.

◯ I completed two of the online sessions.

◯ None - I did not complete any of the online training sessions.

Reasons for Incompletion

3. Why were you not able to complete any of the training sessions?

4. Why were you not able to complete all of the training sessions?

Preparation for the GBIC Relevance lessons

5. How well do you think the training prepared you for delivering the Relevance lessons?

Not very well   Not well   Well   Very well
⚪   ⚪   ⚪   ⚪

Quantity of GBIC Relevance lessons delivered

6. Which of the 13 GBIC Relevance lessons were you able to deliver? [select all that apply, if you are not sure click other and specify the number of lessons you delivered (e.g. 12)

☐ Lesson 1: Technology in your life
☐ Lesson 2: Getting to know App Lab
☐ Lesson 3: Event-driven programming
☐ Lesson 4: Develop functionality through selection
☐ Lesson 5: Changing your world with technology
☐ Lesson 6: Deciding which problem to solve
☐ Lesson 7: Industry engagement
☐ Lesson 8: Understanding user needs
☐ Lesson 9: UX design
☐ Lesson 10: Build your app
☐ Lesson 11: Continue to build your app
☐ Lesson 12: Finish building your app
☐ Lesson 13: Test and evaluate your app

If you are not sure, please add anything else you remember here:

```

```

7. How many school weeks (i.e. not including school holidays or half-term) did it take you to deliver the lessons?

⚪ Less than 7 weeks
⚪ 7 weeks
⚪ 8 weeks
⚪ 9 weeks
⚪ 10 weeks
⚪ 11 weeks
⚪ 12 weeks
⚪ 13 weeks
⚪ 14 weeks

◯ 15 weeks
◯ More than 15 weeks

Reasons for length of delivery

8. What were the main reasons it took less than 13 weeks?

```
[                                                                    ]
```

9. What were the main reasons it took more than 13 weeks?

```
[                                                                    ]
```

Your experience with delivering the GBIC Relevance lessons

10. We're interested to know how easy it was to deliver the lesson content in the way that it was set out in the lesson plans, and what adaptations you might have needed to make.

What changes, if any, have you made to the content of the lessons (e.g. slides, worksheets, lesson plan)?

▢ I added content (example: I created a completely new worksheet or slide)
▢ I skipped content
▢ I adapted content (e.g. I adapted a slide or worksheet provided)
▢ None of the above (I delivered the content without alteration)

Reasons for changes

11. What changes did you make?

```
[                                                                    ]
```

12. Why did you make these changes?

```
[                                                                    ]
```

Reactions to the programme

13. Overall, how could the lessons be made more engaging for the female pupils?

Any other feedback?

Do you have any other feedback regarding the GBIC relevance programme?