Gender Balance in Computing

Evaluation of Intervention 1b: Storytelling Approach

October 2021

Benji Horwell, Pujen Shrestha, Tim Hardy, Clément Bisserbe and Matthew Holt

THE BEHAVIOURAL INSIGHTS TEAM

Contents

Executive summary

Background

This evaluation focuses on an intervention implemented as part of the Gender Balance in Computing (GBIC) programme, which is designed to address barriers to female pupil's uptake of computing. This intervention, the Storytelling approach, was intended to address the barrier of a mismatch between teaching approaches and pupil learning styles by using storytelling to teach computing. These Storytelling lessons were delivered to primary school pupils in year 2 (ages 6-7).

Methods

The research was designed as a **pilot randomised controlled trial**, with two arms (one control, one treatment), and was randomised at the school level with outcomes at the pupil level.

The evaluation used a mixed-methods approach. Three methods were used as part of the quantitative and qualitative data collection:

- Pupil surveys - these measured pupils' attitudes toward computing and their intentions to study computer programming in the future.
- Case studies - three schools in the treatment group underwent an observation of a Storytelling lesson, a semi-structured interview with a teacher delivering the intervention, and a discussion with pupils.
- A teacher survey - this assessed teachers' experience of the programme.

30 schools were allocated to the treatment group, of which 9 completed the intervention and both the baseline (pre-intervention) and endline (post-intervention) survey; 30 schools were allocated to the control group, of which 10 completed the baseline and endline surveys. The intervention was implemented between March and July 2021, against a backdrop of fluctuating COVID-19 cases in the UK during the pandemic.

Findings

Evidence of promise

Quantitative evidence for this pilot was limited. There was no difference between treatment and control groups in terms of any changes to pupils' attitudes toward computing, their stated intention to study computing in future or their intention to study science and/or maths in future. However, these findings cannot be interpreted with any causal confidence due to the high attrition rates across both groups.

Interviewed teachers did not perceive a difference in the impact of the programme between male and female students, and instead felt that the Storytelling lessons had increased both male and female pupils' interest, confidence and skills equally. This was potentially because the lessons appealed to students across the board, rather than female students specifically.

Fidelity

Teachers were able to deliver all 12 of the Storytelling lessons and only made minor adjustments to the lesson content in response to pupil need, such as adapting terminology or the pace of lessons.

Acceptability

Teachers felt that the training was accessible and helped prepare them for delivering the lessons, but also felt that it took too much time and was less helpful for preparing them for teaching the more complex skills used in the later lessons.

Pupils enjoyed the lessons, particularly the ability to personalise their stories and display their work to others. Teachers were positive about the lessons and the use of storytelling to teach coding, and felt that their confidence and computing ability had improved during the programme.

Feasibility

Lessons were delivered in line with the lesson plans and teachers demonstrated good behaviour management skills. Teachers felt that lessons were appropriate for the age group and were well differentiated for different abilities, but some more confident teachers felt that the earlier lessons might be suitable for year 1 students.

Teachers found the Storytelling materials well-explained and easy to pick up, though some felt the slides occasionally lacked important detail or were overwhelming for pupils. Resource constraints created extra difficulties for teachers when delivering the lessons because, for example, some schools were unable to provide one tablet per pupil or a teaching assistant to help answer pupils' questions.

Readiness for trial

As stated, high attrition was observed in this trial. While this may indicate a risk that the requirements for school recruitment under a full-scale trial may not be met, it is difficult to determine the extent to which the COVID-19 context contributed to the attrition observed, and the extent to which school recruitment may be hindered by similar attrition levels in a different context.

From a methodological standpoint, the survey does not require adaptations prior to being used in a full-scale survey. However there were logistical concerns, with teachers commenting that pupils were unable to complete the surveys on their own because their reading and comprehension skills were not advanced enough. This meant that the surveys were extremely time consuming to administer as teachers had to provide extensive support to explain and answer the questions.

Conclusions and recommendations

- Based on preliminary qualitative evidence of the intervention being acceptable for pupils as well as feasible for teachers to deliver, we recommend that the intervention move to a full trial.
- Teachers requested that the Raspberry Pi Foundation (RPF) find ways to shorten the training or allow teachers to skip content they have covered.
- Less experienced teachers highlighted the need for RPF to ensure the training introduces teachers to all of the content covered in the lessons.
- More experienced teachers asked future iterations of the intervention to consider allowing teachers to decide which lessons to deliver based on the ability of their class.
- In order to go to trial, the school recruitment team will need to review attrition strategies, focusing particularly on schools that are randomised into the control group.
- In the interest of ensuring data accuracy, evaluators should conduct further cognitive testing for the survey.
- It is also recommended that future evaluators conduct further validity testing for the survey.

1. Intervention background

1.1. The GBIC programme

The Gender Balance in Computing (GBIC) programme is funded by the Department for Education (DfE) and involves the delivery of five interventions by the Coalition for a World-Leading Computing Education, a consortium of STEM Learning, the British Computer Society, and the Raspberry Pi Foundation (RPF). Each of the interventions is designed to address a barrier to female pupil's uptake of computing.

1.2. The Storytelling intervention

Rationale

This arm of the GBIC programme - the Storytelling approach - was intended to address the barrier of a mismatch between teaching approaches and pupil learning styles. The intervention used storytelling and story-writing to introduce children in Year 2 (ages 6-7) to computing principles such as sequencing and repetition. This approach had demonstrated some effectiveness with 10-15 year olds[1], but had not been tried with younger children.

Training

Teachers completed mandatory training and were given access to step-by-step lesson plans and age-appropriate software for computer programming (ScratchJr). Training was originally due to take place face-to-face, but due to COVID-19 restrictions it was completed online. Teachers were initially invited to a live webinar where they were introduced to the project and given an opportunity to ask questions. The training itself was structured into three self-directed sessions completed on a RPF training website. Schools were offered £100 per staff member to arrange supply cover while teachers completed the training.

Lesson delivery

The intervention was delivered by teachers between April and July 2021 and was made up of 12 lessons, each designed to last one hour. Unit 1 comprised six lessons introducing pupils to the ScratchJR programme and the fundamental coding 'blocks' they would need to write their stories. Unit 2 involved a further six lessons introducing more complex coding blocks to enable pupils to develop a full digital story.

The lessons were intended to be delivered weekly, although teachers could choose to vary the regularity of sessions and their length as long as all content was covered. The intervention was delivered to all pupils in mixed-gender classes, but the primary focus of our evaluation is the experience of female pupils.

---

[1] Caitlin Kelleher, Randy Pausch, and Sara Kiesler. 2007. Storytelling Alice motivates middle school girls to learn computer programming. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1455–1464. DOI:https://doi.org/10.1145/1240624.1240844

2. Research questions

The pilot evaluation was designed to address five main questions:

1. **Evidence of promise:** is there initial evidence that the intervention has an effect on the target outcomes?
2. **Fidelity:** is the content set out in the 12 lesson plans covered, and what affects teachers' ability to deliver the intervention as intended?
3. **Acceptability:** how do teachers and pupils experience the intervention, and what affects their perceptions of it?
4. **Feasibility**: how easy do teachers find it to deliver the intervention and what are the barriers and facilitators to delivery?
5. **Readiness for trial**: are data collection and intervention procedures sufficiently developed for the intervention to undergo an impact evaluation?

3. Methods

3.1. Research design

Evaluation design

The evaluation was designed as a **pilot randomised controlled trial (RCT)**, with two arms (one control, one treatment), and was randomised at the school level with outcomes at the pupil level, in addition to other measures at the teacher and school level. The aim of this pilot was to assess the feasibility of conducting a trial at scale, as well as any evidence of promise that the intervention may benefit the target group of pupils. Pilot RCTs allow for the testing of implementation processes, randomisation and data collection procedures, particularly when the feasibility of running a large-scale impact evaluation (e.g., a full-scale randomised controlled trial) is unknown. Including a pilot stage can thus help to assess whether an impact evaluation is viable, what approach is most suitable for conducting an impact evaluation, whether any adaptations are required prior to it taking place, and whether there is an indication that the intervention could have an impact on the outcomes of interest (i.e., 'evidence of promise').

The research questions were investigated through a **mixed-methods approach** involving both quantitative and qualitative data collection. Fidelity (question 2), Acceptability (question 3) and feasibility (question 4) were primarily assessed using qualitative data, whereas evidence of promise (question 1) and readiness for trial (question 5) were assessed using both methods.

Sampling and randomisation

Sampling

Recruitment for this pilot was conducted by RPF. All primary schools in the UK were eligible for this pilot, however pupils within schools were ineligible to take part in more than one GBIC intervention to avoid contamination between trials. RPF was responsible for anticipating and managing such 'conflicts' between trials during recruitment.

To enter the sample, the primary school had to volunteer to participate in the intervention, and to complete a memorandum of understanding.

Figure 1: School level attrition[2]

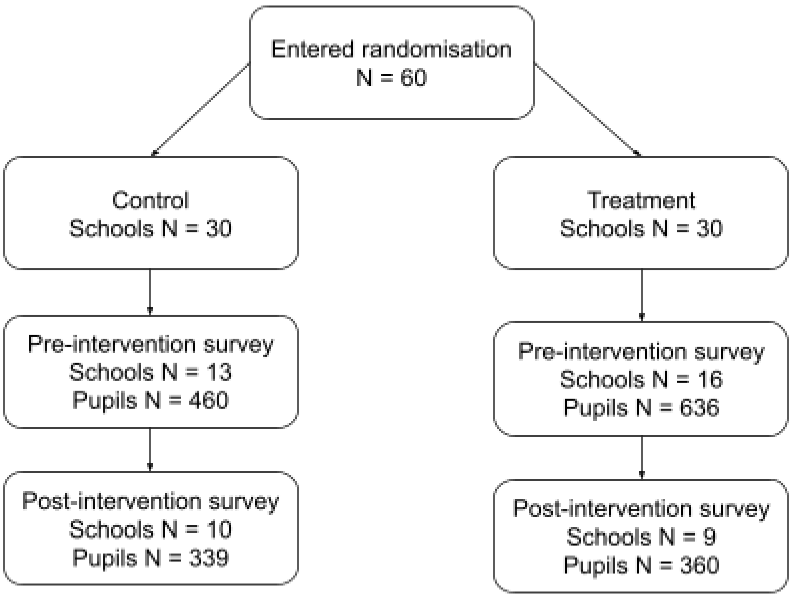


Figure 1 below describes school level attrition at the different stages between recruitment and the completion of the endline survey. There was considerable school attrition throughout the stages between recruitment and the completion of the trial. This is likely due to the timing of the pilot, which was implemented during COVID-19 school disruptions and closures (March-May 2021). It should be noted that we did not observe within school attrition, but rather entire schools withdrawing from the programme. Implications of this attrition for the interpretation of the results are discussed in section 4.3.

Randomisation

Schools recruited for the trial were randomised into a treatment group, which received the intervention, and a control group, wherein business-as-usual computing lessons were taught to pupils. Randomisation was stratified by the *proportion of pupils eligible for free school meals (FSM) in the school*, binarised into above and below the median, and calculated

[2] Among the 60 schools that were randomised, 7 had expressed interest but not yet signed an MoU for their participation in the trial. These schools ended up not participating in the trial, such that only 53 schools were formally recruited for the trial. From these 53 schools, 29 completed the pre-intervention survey, and 19 completed the post-intervention survey.

across the schools in the sample. This was done to ensure balance on these variables at the randomisation stage, given its possible correlation with the outcomes of interest for this pilot evaluation.

Pupils were blind to allocation during the programme and during outcome data collection. Teachers were not blind to allocation, as they were responsible for delivering the materials, and as they had registered interest in participating in the trial, would have been aware of a control group. We used school unique reference numbers (URNs) as unique identifiers. BIT conducted the randomisation.

Balance checks

As specified in the Evaluation plan, we conducted balance checks after randomisation on the following variables:

1. *The proportion of girls in each school*, by checking if the normalised difference between arms is smaller than 0.25[3].
2. *Ofsted ratings*, by checking whether differences between arms in any rating category are less than 5 schools (~17% of arm size).

We also conducted balance checks on these variables for the schools with at least one response in both the baseline and endline surveys.

Table 1 shows the results of the balance checks for the proportion of girls in each school for each stage (randomisation, baseline survey, endline survey). The normalised difference is less than 0.25 at randomisation but not for the baseline or endline surveys. We also perform t-tests at each time point; the corresponding p-values are all above 0.1, though this is not surprising given the small sample sizes. In any case, the actual magnitude of the differences for each survey is small (47.7% vs. 48.9% at the baseline survey, and 47.7% vs. 49.7% for the endline survey), thus, as specified in the TP, we do not include this variable as a covariate in our main specifications.

Table 1: Mean proportion of girls in each group

	Control	**Treatment**	**Overall**	**Normalised difference**	**p-value from t-test**
Mean proportion of girls by school					
Randomisation	0.498 (N=30)	0.494 (N=30)	0.496 (N=60)	0.136	0.6006
Baseline survey	0.477 (N=13)	0.489 (N=16)	0.483 (N=29)	0.400	0.2792
Endline survey	0.477 (N=10)	0.497 (N=9)	0.486 (N=19)	0.753	0.1283

[3] This is *(mean_1 - mean_2) / sqrt((var_1 + var_2)/2)*, as per Imbens and Ruben (2015), which also specifies the rule of thumb that a difference of less than 0.25 suggests good balance.

Tables 2-4 present the results of the balance checks for Ofsted rating at each of the three stages. According to the rule above, the arms are balanced on Ofsted ratings at each stage (i.e. there are no differences in any rating category of at least 5 schools). We also run chi-square tests to test whether the distribution of values for this variable significantly differed in the treatment and control group, and do not find a significant difference at any point (p=0.5987, p=0.2877, p=3679).

Table 2: Distribution of values for school Ofsted rating in each group, post randomisation

	Control (N=30)	**Treatment (N=30)**	**Overall (N=60)**	**p-value**
Ofsted Rating				0.5987
Outstanding	7 (23.3%)	8 (26.7%)	15 (25.0%)	
Good	19 (63.3%)	18 (60.0%)	37 (61.7%)	
Requires Improvement	2 (6.7%)	4 (13.3%)	6 (10.0%)	
Inadequate	1 (3.3%)	0 (0.0%)	1 (1.7%)	
Missing	1 (3.3%)	0 (0.0%)	1 (1.7%)	

Table 3: Distribution of values for school Ofsted rating in each group, baseline survey (school level)

	control (N=13)	**treatment (N=16)**	**Overall (N=29)**	**p-value**
Ofsted Rating				0.2877
Outstanding	2 (15.4%)	3 (18.8%)	5 (17.2%)	
Good	9 (69.2%)	10 (62.5%)	19 (65.5%)	
Requires Improvement	0 (0.0%)	3 (18.8%)	3 (10.3%)	
Inadequate	1 (7.7%)	0 (0.0%)	1 (3.4%)	
Missing	1 (7.7%)	0 (0.0%)	1 (3.4%)	

Table 4: Distribution of values for school Ofsted rating in each group, endline survey (school level)

	control (N=10)	treatment (N=9)	Overall (N=19)	p-value
Ofsted Rating				0.3679
Outstanding	1 (10.0%)	3 (33.3%)	4 (21.1%)	
Good	7 (70.0%)	5 (55.6%)	12 (63.2%)	
Requires Improvement	0 (0.0%)	1 (11.1%)	1 (5.3%)	
Inadequate	1 (10.0%)	0 (0.0%)	1 (5.3%)	
Missing	1 (10.0%)	0 (0.0%)	1 (5.3%)	

3.2. Data collection

Data collection methods

We used multiple methods to collect data: pupil surveys, administrative data on school recruitment, teacher surveys, teacher interviews, lesson observations and pupil discussions. Table 5 describes methods used to collect data, when they were used to collect data, and the research questions this data was used to examine.

Table 5: Research questions examined through each data collection tool

Data collection tool	Point of collection	Research questions examined
Pupil survey	Pre- and post-intervention	1) Evidence of promise; 5) Readiness for trial
Teacher interviews	Post-intervention	2) Fidelity; 3) Acceptability; 4) Feasibility
Pupil discussions	Post-intervention	2) Fidelity; 3) Acceptability; 4) Feasibility
Lesson observations	During intervention	2) Fidelity; 3) Acceptability; 4) Feasibility
Teacher survey	Pre- and post- intervention	2) Fidelity
Administrative data	Pre-intervention	5) Readiness for trial

Pupil surveys

Pupils in treatment and control arms were asked to complete a survey prior to the start of the intervention, and after its completion. The baseline survey explored three outcomes; initial awareness of, experience with, and attitudes toward computing that pupils may have prior to commencing any formal education on the topic. It also measured pupils' stated intention to study computing, as well as the related topics of science and maths. As we could not assume that pupils had meaningful prior experience with computing before the intervention, the baseline was shorter and less detailed than the endline.

Attitudes to computing were captured by items asking about confidence in computing, interest in computing, feelings of belongingness in computing, and perceived usefulness of computing. The items used were adapted from the Student Computer Science Attitudes Survey (SCSAS), with substantial modifications for the much younger study sample in the current trial. The SCSAS was modified and cognitively tested with a sample of KS1 students and teachers. The findings of the cognitive testing led to the following modifications (see Annex A for the full baseline and endline surveys):

1. Attitudinal items were measured using smiley faces to indicate a positive, neutral and negative response to a question
2. Teachers were provided with instructions to assist pupils respond to survey items as pupils were unable to complete the surveys without teacher support
3. Teachers were advised to provide a concrete example of what computer programming is (e.g Beebots) due to the potential for pupils to misinterpret the concept or to interpret the application of the concept differently for different parts of the survey (e.g thinking that playing computer games can be considered programming)
4. Items that were misinterpreted were either modified or removed from the survey

Surveys were completed online by the pupils with assistance from teachers before the first lesson and after the sixth lesson. In an attempt to ease the burden of teachers, BIT staff prefilled classlists with pupil names, leaving blank columns for teachers to enter pupils' date of birth and URN. Gender data was also collected by BIT researchers and teachers after the endline surveys were administered. Teachers were then asked to send the completed classlists to BIT researchers who then merged classlist data with data from the surveys. Unfortunately, feedback from teachers highlighted the additional burdens placed on them to both administer the survey and to fill in the classlists, and they instead suggested that if run again, the data should be collected within the survey itself at the time of administration.

Qualitative methods

To help assess fidelity, acceptability, and feasibility, we implemented a case study design, conducting qualitative research activities in three schools (cases) in the treatment group. For each case, we conducted:

- **Semi-structured interviews with teachers,** lasting around 45 minutes, focused on teachers' background, their experience of training and lesson delivery, and their perceptions of pupils' engagement and any perceived impact of the intervention (see Annex B for the full interview guide).

- **Observations**,designed to independently assess pupil engagement, lesson fidelity, and facilitators and barriers to lesson delivery (see Annex C for the full observation pro-forma).

- **Pupil discussions,** involving four female and two male students in each of the three schools, and lasting no more than 15 minutes on the recommendation of the teachers involved. Pupils were provided with worksheets asking them about their lesson, and encouraging them to draw someone who works in computing, followed by a discussion about their responses (see Annex E and F for the full discussion guide and pupil worksheet). The latter exercise was designed to identify any initial evidence that lessons affected female pupils' views that computing was for people like them (a hypothesised mechanism in the intervention's logic model, see Annex G).

Teacher survey

After the intervention had been delivered, we asked all teachers in the treatment and control groups to complete a survey to share their experiences of the programme through a combination of likert-scale responses and free text.

Teachers in the control group were asked about their teaching confidence and business-as-usual computing activities in their school to help establish similarities and differences to the treatment group (see Annex D for the full survey).

Teachers in the treatment group were asked about their experience of training, how closely they adhered to the intervention, their thoughts on the impact of the intervention, and their overall experience of the programme and any suggested improvements (see Annex D for the full survey).

Note that no teacher survey was conducted before the intervention, as we did not want to overburden teachers with additional data collection requirements and risk increasing school attrition further.

Administrative data

We obtained data from RPF on the number of schools that completed the expression of interest form and completed the memorandum of understanding form. From the pupil survey, we also collected data on the number of pupil baseline and endline surveys completed for each school, as well as the number of pupils who were absent and who opted out.

Outcome measures (for quantitative evaluation)

To investigate the evidence of promise of the intervention, we used three outcome measures for the analysis of the quantitative data collected through the pupil survey:

1. The **primary outcome measure** is an overallSCACS score computed across 10 survey items from the pupil endline that ask pupils about their attitudes to computer programming. Based on the logic model, it is assumed that positive changes in these attitudes will translate into increases in the likelihood of choosing at GCSE or A Level computing in the future.
2. The **first secondary outcome measure** is pupils' stated intention to study computer programming as measured in the endline. It is assumed that this stated intention is positively correlated with actual choices of computing in formal education.[4]

[4] Previous research with Year 10/11 female pupils found that 73% of those that stated an intention to study computing at A Levels actually continued to do so. Of course, we are surveying pupils in Year 2,

3. The **second secondary outcome measure** is pupils' stated intention to study science and/or maths, as measured in the endline.

Our sample for the primary and secondary analysis consists of responses to the endline from pupils that are identified as girls. This sample of girls is defined as follows:

- The pupil's gender is female according to the classlist filled in by their teacher;
- If the pupil's gender was not filled in by their teacher (true for 128 out of 699 respondents to the endline survey), at least 95% of live births with their first name were recorded as girls instead of boys in England and Wales over 1996-2020[5].

Under this definition, there are 346 girls in our sample for the primary analysis. Note that changing the 95% threshold above has very little effect on the sample size - a 50% threshold would increase the sample size by 1, and a 100% threshold would reduce it by 4.

We correct for multiple comparisons within each type of outcome using the Benjamini-Hochberg step-up method[6]. This only applies to secondary outcomes (since there is one primary outcome).

We also analyse the primary and secondary outcomes for boys as exploratory analysis, where boys are identified in an analogous way. 326 boys are identified, meaning that 27 respondents could not be identified as being either girls or boys and are excluded from both sets of analyses.

---

so the time gap between intention and action is much larger here, likely making this figure an overestimate for the current context.

[5] Statistics compiled from ONS data: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/babynamesenglandandwales/2020#baby-names-data

[6] We do this correction to account for the fact that as the number of hypotheses tested simultaneously increases, the risk of false discovery (finding a significant result by chance) increases as well. The Benjamini-Hochberg step-up method involves taking the p-values from each comparison, arranging them in ascending order, and comparing them with a linearly increasing vector from 0.05/k (where k is the number of comparisons) to 0.05, instead of comparing p-values to a fixed significance threshold (usually 0.05). Once a comparison is found not significant, all remaining comparisons are also classified as non-significant.

Table 6: Outcome measures for the quantitative evaluation

<table>
<tr><th>Outcome Measures</th><th>Data to be collected</th><th>Point of collection</th></tr>
<tr><td>Primary: General attitudes towards computing</td><td>Overall score on the Student Computer Science Attitudes Survey (SCSAS) (score range: 10-30)[7] in the endline</td><td rowspan="3">Collected at the end (endline) of the 12-week evaluation period</td></tr>
<tr><td>Secondary: Stated intention to study computer programming in the future</td><td>Endline response to the question: Do you want to study computer programming in the future?

Possible responses were ‘Yes’ / ’No’ / ‘Maybe’ / ‘Don’t want to say’, and were binarised into:
● 1 = “Yes”
● 0 = “No”, “Maybe”
● Missing = “Don’t want to say”</td></tr>
<tr><td>Secondary: Stated intention to study science and/or maths in the future</td><td>Endline response to the question: Do you want to study [Science, Maths] in the future?

Possible responses were ‘Yes’ / ’No’ / ‘Maybe’ / ‘Don’t want to say’, and were binarised into:
● 1 = “Yes” for science and/or maths
● 0 = “No”, “Maybe”, for science and maths
● Missing = “Don’t want to say”</td></tr>
</table>

4. Analytical strategy

4.1. Analytical approach for quantitative data

Analytical model

The pupil outcome analysis involved the exploration of three outcome variables:

1. **Primary outcome**: General attitudes towards computing, measured as the overall SCACS score.
2. **Secondary outcome**: Stated intention to study computer programming in the future.
3. **Secondary outcome**: Stated intention to study science and/or maths in the future.

Intervention effects were estimated for primary and secondary outcomes to give an indication of potential effects for a future full-scale trial. However, non-statistically significant treatment effects should not be interpreted as evidence of intervention effectiveness.

[7] Heavily adapted from Haynie and Packman (2017). Available at: https://csedresearch.org/tool/?id=156.

As the primary outcome ([1] SCACS score) is continuous, we used a linear regression to assess the Intention-To-Treat (ITT) effect of the intervention on this outcome, with cluster-robust standard errors and clustering at the school level.

As the secondary outcomes (stated intention to study [2] computer programming, and [3] science and/or maths) are binary,we used a logistic regression to assess the ITT effect of the intervention on these outcomes, with cluster-robust standard errors and clustering at the school level.

Description of data

Tables 7 and 8 present the means and 95% confidence intervals reported for all baseline and endline outcome measures split by arm and gender.

Table 7: Pupil baseline survey data by treatment group and gender

Survey item	Values	Gender	Group	N (non-missing)	Mean (SD)	95% CI	P-value from chi-square test
Do you know what computer programming is?	1 = "Yes" 0 = "No", "Maybe" Missing = "Don't want to say"	Girls	Control	208	0.260 (0.439)	0.200, 0.320	0.2924
		Girls	Treatment	284	0.215 (0.411)	0.167, 0.263	
		Boys	Control	203	0.320 (0.468)	0.255, 0.385	0.0912
		Boys	Treatment	284	0.246 (0.432)	0.196, 0.297	
Have you done computer programming before?	1 = "Yes" 0 = "No", "Maybe" Missing = "Don't want to say"	Girls	Control	214	0.495 (0.501)	0.428, 0.563	0.1189
		Girls	Treatment	301	0.422 (0.495)	0.366, 0.478	
		Boys	Control	206	0.519 (0.501)	0.451, 0.588	0.5473
		Boys	Treatment	285	0.488 (0.501)	0.429, 0.546	
Do you like computer programming?	1 = "Yes" 0 = "No", "Maybe" Missing = "Don't want to say"	Girls	Control	204	0.588 (0.493)	0.520, 0.656	0.2216
		Girls	Treatment	292	0.527 (0.500)	0.470, 0.585	
		Boys	Control	195	0.631 (0.484)	0.562, 0.699	0.6651
		Boys	Treatment	272	0.607 (0.489)	0.548, 0.665	
Do you want to study computer programming in the future?	1 = "Yes" 0 = "No", "Maybe" Missing = "Don't want to say"	Girls	Control	211	0.536 (0.500)	0.468, 0.603	0.2814
		Girls	Treatment	288	0.483 (0.501)	0.425, 0.541	
		Boys	Control	204	0.583 (0.494)	0.515, 0.652	0.2236
		Boys	Treatment	279	0.523 (0.500)	0.464, 0.582	
Do you want to study science and/or maths in the future?	1 = "Yes" 0 = "No", "Maybe" Missing = "Don't want to say"	Girls	Control	213	0.831 (0.376)	0.780, 0.882	0.1251
		Girls	Treatment	293	0.771 (0.421)	0.723, 0.820	
		Boys	Control	200	0.860 (0.348)	0.811, 0.909	0.0446
		Boys	Treatment	277	0.783 (0.413)	0.735, 0.832	

Notes:
1. To measure imbalance for each item in the baseline survey, we conducted two-proportions tests on the response to each survey item.
2. Note that if a pupil gave two different answers for science and maths, we selected the most positive one (e.g. ,if they chose "Yes" for one subject and "Maybe" for another, this was coded as "Yes").

Table 7 shows that the control group scores higher at baseline than the treatment group in measures of awareness, experience and attitude toward computing, for both girls and boys. To quantify this imbalance, we ran a chi-square test on each survey item comparing the treatment and control groups, for boys and girls separately. We do not find any significant differences for girls at the 10% level, but we do observe two cases of imbalance for boys at the 10% level, one of which is also an imbalance at the 5% level. Some imbalances are expected given the low number of schools and high level of attrition. Imbalances at baseline mean that the estimated treatment effects obtained by just comparing the outcomes between control and treatment groups could be biased; however, controlling for baseline measures in the regression analysis does not affect the main takeaways.

Nevertheless, it is important to note that overall there was low knowledge, experience, and positive attitude towards computer programming. This is particularly observable in the fact that stated intention to study science and/or maths in the future is higher than stated intention to study computer programming in the future.

Table 8 tabulates responses to the endline survey for the key outcomes, for girls and for boys. When calculating the “general attitudes towards computing” index (SCSAS score), we replace “Don’t want to say” answers with the same pupil’s average “score” across all other questions. We exclude one response from a boy who responded “Don’t want to say” to all relevant questions. We also present descriptive statistics for the SCSAS score for the sample of pupils who did not answer any questions with “Don’t want to say”. The analysis of the pupil endline survey data found that, for girls, the average computing attitudes score (i.e. SCSAS score) of 26.20 for the treatment group was lower than the average of 26.45 for the control group. However, the 0.26-point difference between the two groups is not statistically significant ($p = 0.703$). Given that the difference between the two groups is not statistically significant, the small magnitude of the difference, and the study limitations discussed in section 4.3, this result cannot be interpreted as the causal impact of the programme on girls’ outcome. If we only include complete responses (without any “don’t want to say” answers), the gap shrinks to 0.02 points, further caveating the interpretation of the difference observed between the two groups.

Table 8: Pupil endline survey data by treatment group and gender

Survey item	Values	Gender	Group	N (non-missing)	Mean (SD)	95% CI	p-value
General attitudes towards computing	Overall score on the SCSAS (score range: 10-30)	Girls	Control	161	26.45 (3.08)	25.97, 26.93	0.4521
		Girls	Treatment	185	26.20 (3.20)	25.73, 26.66	
		Boys	Control	159	26.59 (3.33)	26.07, 27.11	0.1591
		Boys	Treatment	166	26.04 (3.67)	25.48, 26.61	
General attitudes towards computing - complete responses only	Overall score on the SCSAS (score range: 10-30)	Girls	Control	126	26.65 (3.14)	26.10, 27.20	0.9488
		Girls	Treatment	145	26.63 (2.74)	26.18, 27.08	
		Boys	Control	131	27.24 (3.06)	26.71, 27.77	0.0045
		Boys	Treatment	132	26.02 (3.85)	25.35, 26.68	
Do you want to study computer programming in the future?	1 = “Yes” 0 = “No”, “Maybe” Missing = “Don’t want to say”	Girls	Control	155	0.432 (0.497)	0.353, 0.511	1.000
		Girls	Treatment	179	0.430 (0.496)	0.357, 0.503	
		Boys	Control	153	0.529 (0.501)	0.449, 0.609	0.4267
		Boys	Treatment	159	0.478 (0.501)	0.399, 0.556	
Do you want to study science and/or maths in the future?	1 = “Yes” 0 = “No”, “Maybe” Missing = “Don’t want to say”	Girls	Control	148	0.831 (0.376)	0.770, 0.892	0.0544
		Girls	Treatment	174	0.736 (0.442)	0.669, 0.802	
		Boys	Control	154	0.890 (0.314)	0.840, 0.940	0.0014
		Boys	Treatment	155	0.742 (0.439)	0.672, 0.812	

Notes:
1. To measure differences between the groups, we conducted a t-test for SCSAS score (primary outcome) and two-proportions tests for the other (secondary) outcomes.
2. Note that if a pupil gave two different answers for science and maths, we selected the most positive one (e.g. ,if they chose "Yes" for one subject and "Maybe" for another, this was coded as "Yes").

See Annex A for descriptive statistics on the 10 questions making up the SCSAS score (primary outcome).

4.2. Analytical approach for qualitative data

For the case studies, all interviews were audio-recorded and transcribed. Notes from the observations and pupil discussions were recorded by hand. All qualitative data were summarised and managed using the Framework Method[8]. Analysis involved working through the framework to identify the range of views and experiences, identify similarities within and between cases, and identify any explanatory patterns and themes against each of the relevant research questions.

Quantitative and qualitative findings were shared between researchers in order to identify areas of support or contradiction prior to the reporting stage.

4.3. Limitations

Pilot evaluation with limited sample size

This trial was proposed as a pilot study. Pilot studies allow implementation processes and data collection procedures to be tested, where the feasibility of running a large-scale impact evaluation (e.g., a full-scale randomised controlled trial) is unknown. By including a pilot stage, it is possible to determine whether an impact evaluation is viable, what approach is most suitable for conducting an impact evaluation, whether any adaptations are required prior to it taking place, and provide an indication of evidence of promise.

As such, this evaluation was not intended to estimate the impact of the intervention on the outcomes of interest, however the direction of any effect observed can provide some indication of promise of the intervention in improving female pupils' attitudes towards computing.

Attrition

Limitations related to sample size were accentuated by the high level of attrition observed throughout the study, to which the COVID-19 context likely contributed. RPF gathered feedback from schools who felt compelled to drop out due to pupil absences from self-isolation. The intended sample for this project was 60 schools across the treatment and control group. However, only nine of the 30 schools in the treatment group completed the baseline and endline survey, which served as a proxy for intervention completion. 10 of the 30 schools in the control group completed the baseline and endline surveys.

Further, in addition to reducing the sample size (and thereby statistical power) further, attrition is likely not to have been random, and school characteristics influencing the likelihood to complete all study activities could plausibly be correlated to pupil attitudes towards computing. For instance, schools with greater resources may be better able to minimise disruptions related to COVID-19, as well as have greater computing resources, which could in turn shape attitudes towards computing. While the size and direction of any bias introduced by non-random attrition are unclear, this risk is worth noting.

---

[8] Gale, N.K., Heath, G., Cameron, E. et al. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 13, 117.

Collection of survey data from KS1 pupils

The pupil survey tool used was cognitively tested with a sample of KS1 pupils prior to this trial, and adjusted based on lessons from this testing. Further, the results from the validation of the survey tool described in section 5.5 suggest that it was adequate to measure attitudes towards computing. That being said, the administration of the survey involved teachers assisting pupils to answer questions, and there is thus a risk that responses may be influenced by this involvement from teachers. Due to the gap in previous survey research of this topic with this age group, the assistance from teachers was considered a reasonable and necessary condition.

5. Findings

5.1. Evidence of promise

Evidence of promise was assessed using data from the pupil surveys, teacher surveys, teacher interviews and pupil discussions.

5.1.1. Quantitative evidence

Primary analysis: effect of the intervention on computing attitudes score

As this is a pilot trial, estimating treatment effects was not the primary objective of the analysis. However, we aimed to comment on possible evidence of promise of the intervention by estimating the direction and magnitude of any difference observed between the two groups (even if it is not statistically significant).

For boys, the average SCSAS score was 26.59 in the treatment group and 26.04 in the control group. This 0.55-point difference is not statistically significant ($p = 0.351$). If we include only complete responses, the difference increases to 1.23 points and is statistically significant at the 5% level ($p=0.022$). We think this is likely to be a spurious result. It is driven by more negative responses to question 16 (“knowing about computer programming will help me get a job”): excluding this question when forming the SCSAS score gives a p-value of 0.162 among complete responses.

Table 9 describes the output of the regression analysis conducted to assess the Intention-To-Treat (ITT) effect of the intervention on the computing attitudes score by pupil gender. As the outcome is continuous, we used a linear regression as per the study pre-specification. In the Evaluation plan, we specified that we would control for baseline measures related to computing if at least 50% of pupils know what computer programming is and have experience with it. These conditions do not hold: according to the baseline, 24.7% of pupils have knowledge of computer programming and 46.0% have experience with it. These figures are 23.4% and 45.2% respectively for girls only. Including baseline measures in the regressions has little effect on the estimated effects. We use cluster-robust standard errors clustered at the school level (which explains the difference in p-values in the regression tables relative to those in Table 8).

Column 1 presents the raw results for girls, including all respondents who answered at least one of the 10 questions used to form the SCSAS. Column 2 only includes girls who answered every question. Columns 3 and 4 are analogous to columns 1 and 2, but use boys as the sample.

Table 9: Estimated effects on computing attitudes score by gender

	(1)	(2)	(3)	(4)
Outcome: SCSAS	Girls - raw (incl. incomplete responses)	Girls - complete responses only	Boys - raw (incl. incomplete responses)	Boys - complete responses only
Control-group mean	26.45	26.65	26.59	27.24
Treatment-group mean	26.20	26.63	26.04	26.02
Estimated treatment effect (standard error)	-0.255 (0.656)	-0.023 (0.630)	-0.548 (0.573)	-1.229* (0.490)
N	346	271	325	263

Note: * $p < 0.05$, ** $p < 0.01$

Secondary analysis: effect of the intervention on 1) stated intention to study computing and 2) stated intention to study science and/or maths

Table 10 presents the results of the secondary analysis.

- Column 1 uses whether a pupil responded “Yes” to “Do you want to study computer programming in the future?” as the outcome.
- Column 2 uses whether they responded “Yes” to either “Do you want to study science in the future?” or “Do you want to study maths in the future?” as the outcome. Both specifications use a logistic model with no covariates.
- Columns 3 and 4 display results of the same analysis as columns 1 and 2 respectively, but with boys as the sample.

The estimated coefficients from logistic models are given as log-odds ratios. Since there are no covariates, we can also calculate the estimated treatment effect in terms of percentage points by subtracting the control group mean from the treatment group mean.

The estimated treatment effect on the stated intention to study computing is -0.2 percentage points for girls (from the control group mean of 43.2%). The estimated effect on the stated intention to study science and/or maths is -9.5pp (from the control group mean of 83.1%). This is a much larger point estimate but it is not statistically significant, a reflection of the small sample size.

For boys, we observe a negative treatment effect on the stated intention to study computing of -5.1 percentage points (from the control group mean of 52.9%). The estimated treatment effect on their stated intention to study science and/or maths is -14.8pp (from the control group mean of 89.0%). This is significant at the 5% level even after adjusting for multiple comparisons using the Benjamini-Hochberg procedure (unadjusted p-value = 0.014). While this result is statistically significant, given the study limitations described in section 4.3, it should be interpreted cautiously, and does not provide strong evidence of a negative effect of the programme on boys' intention to study science and/or maths.

Table 10: Estimated effects on stated intention to study computer science, and on stated intention to study science and/or maths in future, by gender (logit model)

	(1)	(2)	(3)	(4)
	Girls - intention to study computer science	Girls - intention to study science and/or maths	Boys - intention to study computer science	Boys - intention to study science and/or maths
Control-group proportion	0.432	0.831	0.529	0.890
Treatment-group proportion	0.430	0.736	0.478	0.742
Estimated treatment effect as log-odds ratio (standard error)	-0.009 (0.351)	-0.570 (0.424)	-0.206 (0.366)	-1.031* (0.421)
Estimated treatment effect as percentage points	-0.2pp	-9.5pp	-5.1pp	-14.8pp
N	334	322	312	309

Note: * $p < 0.05$, ** $p < 0.01$

5.1.2. Qualitative evidence

Teachers had positive perceptions on the impact of the intervention on female students

The endline survey asked teachers in the treatment group to assess what impact they thought the programme had on both their female and male students' attitudes toward computing, and we received 10 responses.

In general, teachers felt that the Storytelling lessons had a positive impact on female students' attitudes toward computing, particularly regarding their enjoyment, confidence, and belief that computing is for people like them.

Teachers identified some gendered effects of the intervention

When asked whether they had noticed any differences in the impact of the Storytelling lessons between male and female students, teachers commented that both genders had enjoyed their time and benefited from the intervention.

“...they all seemed to enjoy it. I don't think there's anyone [who has benefited more] in particular.” - Respondent S01

Although teachers liked the storytelling aspect of the intervention and thought that it was a good way to teach coding, there was little evidence to suggest that this had particularly appealed to female pupils compared to male pupils.

When teachers did note gender differences, this was in relation to male students’ experience of computing prior to the Storytelling lessons. One teacher highlighted the importance of the home environment in exposing pupils to computing through access to technology.

“I think boys just tend to pick things up quicker and they've got other skills that they get at home that seem to come in or, they'll just know how to do things on the iPad and the girls might not have touched it. I think they're more enriched at home.” - S02

Interviewed teachers also noted differences between which characters male and female pupils chose for their stories. The same teacher commented that female students chose more stereotypically female characters and male students chose “boy cars” rather than “girl cars”.

“I’ve really noticed how there’s already difference in views of, what’s a boy, what’s a girl, the boys are getting front of me, like, ‘I want a boy car, I don’t want a girl car’; Then we’ve got the other side where we’ve got fairy tales and princesses and, ‘Oh, I’m a bunny. Do you want to play with me?’” - S02

Teachers felt that the intervention improved pupils’ interest in computing across the board

Some teachers noted that male students were more positive at the beginning of the programme than female students, but felt that the lessons had engaged the interest of both male and female students. There was no evidence to suggest the intervention had a negative impact on male students as teachers indicated that interest was high among both genders.

“I can honestly say I don't think there was a difference between the girls and the boys [in terms of pupil engagement]… It was nice, lots of opportunities for them to make it personable to them, but I would say no, I've had an equal balance of girls and nobody, I think, ever thinks oh, I don't want to do this, they're always really excited to do it.” - S03

It was felt that the lessons may have helped to engage male and female students by allowing them to personalise their stories to their own interests; while teachers noted that male and female pupils often picked different characters, and in the pupil discussion groups both genders commented that they enjoyed being able to choose their own backgrounds and characters.

“I think the boys quite liked the wizard with making things disappear. The girls quite liked the bunny rabbits… I think they liked, when they bumped into them

> *something happened. The fighting games! They quite liked that one, the boys." - S02*

Teachers believed that the lessons improved both male and female students' computing skills and confidence

Overall, teachers noted that both male and female students had made good progress during the lessons, and felt that the programme enabled students to access the learning at their own level as they were able to choose which tools to use during the activities.

> *"[The programme] allows some children who might not be able to recall all [the different] skills… to still feel proud of their work, because they're making a choice in saying, 'Oh, no I just wanted my character to move along and talk', and they haven't made it more complicated than that. Whereas, others have got multiple characters and multiple backgrounds and all sorts of things. So it allows them to access it at their level, as well." - S01*

There is some preliminary evidence that the lessons may have helped female students to imagine that computing is for people like them

During the pupil discussions, students were asked to draw someone who works in computing. All of the male students drew a male figure, whereas some of the female students drew someone female, with several drawing their own teacher. In part this seemed to be because students thought that it would please their teacher, but some pupils also noted that their teacher was the main person they knew who worked on a computer. Without a comparison group we cannot infer if these findings were an impact of the intervention, of computing classes more generally or would have been the case even without any exposure to computing, but it is a positive indication that some female students were able to imagine women in computing.

5.2. Fidelity

Fidelity was assessed using data from the teachers survey, teacher interviews, pupils discussions and lesson observations.

Teachers delivered all of the lessons within 12 weeks

All teachers who completed the endline survey confirmed that they had delivered all 12 of the Storytelling lessons, and 5 of the 9 schools who were represented in the survey had been able to follow the one lesson per week guide. Some teachers delivered all of the lessons in less time by combining lessons or holding more than one in a given week, partly because pupils were already familiar with some of the content or progressed more quickly than expected, and sometimes because of simple timetabling preferences.

Other teachers took longer than 12 weeks to deliver all of the lessons, some due to delays caused by the COVID-19 pandemic, while others found that some of the lessons took longer than expected to cover all of the content. This did not seem to vary based on the teacher's level of confidence nor any school-level factors.

Teachers only made minor adjustments to lesson content

All survey respondents said that they had followed the lesson plan content either 'fairly' or 'very' closely. The adaptations that teachers mentioned did not appear to substantially alter the lesson content, and were instead simple changes in response to pupil needs such as slowing down or speeding up lessons, making use of different teaching methods (e.g. role play, live demonstrations of tasks on ScratchJR), or adapting terminology to make it easier for pupils to understand. One teacher discussed adaptations to terminology during their case study interview, highlighting the need to help pupils be confident using the technology rather than ensuring the terminology was exactly as intended.

> *"...though we might not necessarily call it a trigger block or a motion block… they knew and understood that this was a start block, this would get them to move, this would get them to talk. They might not necessarily use the right terminology, but they knew and understood what each of those blocks represented. I suppose it's, for me, getting that across to them and them being able to be confident using it." - S03*

5.3. Acceptability

Acceptability was assessed using data from the teachers survey, teacher interviews, pupils discussions and lesson observations.

The training was accessible but took too long

All survey respondents had completed all three of the online training modules. One teacher from a school with below average spend per pupil noted that the £200 contribution was particularly helpful to bring in supply teacher cover so that they could complete the training.

Some teachers felt that the training could have been shorter or was longer than advised, especially for those who were already quite familiar with ScratchJR, and in their interview a more senior teacher commented that they struggled to find good quality time in which to complete the training because they had a number of competing priorities.

> *"[The main challenge with the training was] [p]robably just the time really… - because you really want to be able to sit and give it that quality time to do it, and amongst everything else you're trying to do, it's just trying to find that time to do it." - S03*

Training prepared teachers well, but less so for the later lessons

Survey respondents from the treatment group universally seemed to have had good experiences of training and all felt it had prepared them 'fairly' or 'very' well to deliver the lessons, with one teacher citing in their interview that the introductory webinar had provided a good opportunity to learn from other teachers' questions.

> *"We had like two training sessions. There was like videos where it showed you what to do. They were really helpful… I watched the video afterwards, but I think you could go to a live Zoom presentation. There were teachers asking questions and I was like, oh, that's a good question." - S02*

Some teachers reported that the training and lesson materials were clearly explained and made the content easier to understand, and interviewees felt that the training gave them a good opportunity to familiarise themselves with the resources and lessons before teaching began, especially for those who had less experience with ScratchJR.

However, teachers with less computing experience reported that they would have benefitted from a chance to explore some of the more complex ScratchJR controls as these were unfamiliar to them when they were teaching the later lessons.

> *"those simpler [controls] that I could use were okay, but it was maybe those more complex ones, like the sending of the mail one… I [also] didn't know about changing... to the next scene. I only found that one out today, so maybe [the training could focus] on the more challenging ones that come towards the end [of the programme]." - S03*

Pupils generally enjoyed the lessons

When asked, pupils indicated that they had enjoyed the lessons and were frequently seen smiling and discussing their work with other pupils during our observations. In one observation, pupils expressed disappointment and let out a big collective sigh when told that the lesson that day would be their final Storytelling lesson.

In contrast to the views of some teachers - who felt that the lessons appealed well to all abilities - pupils identified by the teacher as "less able" appeared to enjoy lessons less as they expressed frustration when they were unable to work certain aspects of the programme e.g. colouring in their character properly. This was particularly challenging for pupils in one school as they had a less experienced teacher who was not always able to help them resolve the issues.

Teachers and pupils liked the storytelling aspect of the intervention

Teachers were positive about the use of storytelling to teach coding. In addition to the fact that pupils enjoyed personalising their stories, teachers felt that storytelling was a useful method for explaining the lesson content and that storytelling gave students something concrete to apply their skills to.

> *"I think [the storytelling aspect] gives them something real to work through, so it's not… abstract… I think through the storytelling, they're able to make it as funny or whatever they want, and it's also their own interest. [Female student name], she dotes on animals, so she's always having giraffes and all of that, so it's something*

> *that they can make their own connections too… Yes, I did really like the storytelling." - S03*

This was a theme that also emerged from the pupil discussions; students commented that they liked choosing different aspects of their story and making them 'real'. However, some teachers commented that the storytelling element of the programme was unclear, particularly for pupils in the earlier lessons. One teacher commented that students found the storytelling elements hard to follow because it did not match how the pupils had been taught about story writing in other lessons outside of computing.

> *"the way we teach story writing normally, is when they're in Year 1, they do a three-part story with a beginning, a middle and an end. Then in Year 2, we begin doing a five-part story, which is like introducing characters, build-up, a problem happens, they solve the problem, then there's an ending to the story. So they found it hard to see that as a story, and so then didn't really follow the lesson…" - S01*

Additionally, some teachers in the control group indicated that their own computing lessons involved storytelling to teach coding, with one teacher in the treatment group noting that the Storytelling lessons were "no different from how we would have taught coding on ScratchJR prior to the programme". This suggests that for some schools the intervention may not have been well distinguished from business-as-usual lessons, and in some cases, may have confused or contradicted the way in which the students are usually taught about storytelling. In this way, the intervention can be considered acceptable to teachers, but not necessary for all teachers.

Pupils enjoyed the chance to work with others and display their work

In the discussions, pupils commented that they liked working with others during the lessons and enjoyed the chance to display their work. Pupils were proud of demonstrating what they'd learned and one teacher commented that there had been clear interest from some parents.

> *"...they want to use Scratch, then they do it at home. The little one that came in later, she's been isolating so she's back today, then she's just done a quick one then she just showed me - I said, 'Oh, that's really good.' 'Yes, Dad helped me with that.' There has been that engagement at home as well." - S03*

This highlights the acceptable elements of the intervention beyond the storytelling itself, including collaboration and a chance to present work. This suggests that RPF should consider all elements of intervention delivery and to include these mechanisms in future iterations.

Teachers' own confidence and computing skills improved during the programme

Teachers noted in the interviews that the intervention was helpful for building their own computing experience and confidence in teaching computing. Teachers found the lesson materials simple and easy to pick up, and one teacher noted that they had learned a lot from their pupils as they progressed through the lessons with them.

> *"…when I told them that we were going to be doing these lessons they were really excited to tell me the things they already knew and the things that they'd done before, and all that sort of thing. So it's quite nice to have a couple of expert children in the class. You can always go, 'Oh, how did you do that?' and so I've used them a bit. Yes, I was fairly confident going into it, but definitely my knowledge has grown through doing it." - S01*

5.4 Feasibility

Feasibility was assessed using data from the teacher interviews and lesson observations.

Lessons were feasible for teachers to deliver

Teachers we interviewed generally commented that the lesson that day had gone well and this was confirmed by researchers' observations; pupils listened attentively and appeared to progress through the lesson material without major challenges. Despite their varying levels of experience, teachers demonstrated a good ability to manage pupil behaviour which helped to minimise disruption during the lessons, with one teacher utilising techniques such as calling "1, 2, 3, eyes on me" to refocus the pupils' attention at the end of an independent activity.

Teachers' impressions of lesson appropriateness varied

In interviews, teachers commented that the lessons were pitched at exactly the right level and differentiated well for varying abilities. They liked that pupils had a choice about which tools they could use, and felt that the programme was easy to use and allowed pupils to gradually build their skills so that they could all achieve something, including those with special needs.

> *"We've got some children with special needs. They'll be able to do something whether it's just create their own character and draw images on it and just make it move in a simple way. [Whereas] you've [also] got some children that can add repeat functions and loops and wait functions in. There's a lot to it... I liked the way that it was… differentiated really effectively." - S02*

During observations, pupils demonstrated competence at logging in to the programme and progressing with their work independently.

More confident teachers - including one from a school that offered extra-curricular computing activities - separately commented (one in an interview, the other in our survey) that the earlier lessons might be appropriate for year 1 pupils, with the later, more complex lessons reserved for those in year 2.

Good lesson materials made delivery easier but the slides were viewed as too crowded

Teachers commented that the Storytelling slides and lesson plans had a good level of detail and were well-explained, which made it easier to pick up and teach to the pupils. However, a less experienced teacher felt that some of the slides lacked some useful information, such as which parts of the slide to draw pupils' attention to. At the same time, they felt that the "overview" slides contained far too much detail, such that the teacher physically covered parts of the slide to allow pupils to focus on the most relevant information.

> *"Then I found the - I don't know what it was called - like a little overview that I'd often leave up whilst they were getting on with their work, was often packed, and it was like a whole screen, like a list of instructions. Whereas, I think maybe spreading that bit out a bit over a couple of slides, and guiding them through it… would have been more useful."* - S01

School resources affected how smoothly the lessons ran

Some teachers mentioned that they did not have enough tablets to have one per child during the lessons. Teachers mitigated resourcing issues by borrowing tablets from other classes, while other teachers simply asked students to work in pairs. Working in pairs appeared to aid discussion, but in some cases more confident pupils dominated use of the tablet, meaning less confident students had less time to work on the programme. There were no observed gender-based trends in terms of which pupils dominated time on the tablet, but in general teachers commented that male pupils were often more confident coders.

In our observation of a less-wealthy school, the teacher appeared to struggle for time during the lesson. Older tablets caused delays as some were not working or took a long time to load, and in the absence of a teaching assistant pupils were forced to queue while waiting to ask the teacher a question. By comparison, in one of the wealthier schools, the teaching assistant supported the less confident children to ensure they progressed through the lesson.

> ***Interviewer:*** *"...Obviously, there's the iPad's, but is there anything else that you need for each lesson?"*
> ***Teacher:*** *"A TA?! I suppose just making sure everything's charged up and sometimes the ScratchJr on the Chromebooks takes a long time to load. It takes just a bit of time."* - S02

5.5. Readiness for trial

Readiness for trial was assessed using data from the pupil surveys, teacher surveys and teacher interviews.

5.5.1. Quantitative evidence

Recruitment and attrition

As described in previous sections, high attrition was observed in this trial. While this may indicate a risk that recruiting and retaining a sufficiently large sample of schools for a full-scale trial of this intervention may be challenging, it is difficult to determine the extent to which the COVID-19 context contributed to the attrition observed, and the extent to which similar attrition levels would be observed in a different context.

We observed a particularly large drop-off in participation between the randomisation and baseline stages with schools in the control group, suggesting that greater focus on combating attrition amongst schools that do not receive the intervention at the baseline stage may help bolster participation. Given the response to COVID-19 has developed, we expect attrition for a non-pilot full version of this trial to be lower and more in line with attrition levels seen in the ongoing GBIC trials. This suggests that although this trial's recruitment is insufficient to meet requirements for a well-powered trial, school-level attrition was high, and the surveys were not completed and returned in sufficient numbers to meet requirements for a well-powered trial, the likelihood is that in a non-COVID circumstance this criteria may have more easily been met.

Three schools were recruited for the qualitative case studies. We used a purposive sampling approach to select a diverse sample based on teaching confidence, free school meal rates, school spend per pupil, and Ofsted rating. We emailed three schools that met the criteria to invite them to take part as a case study. Only one school accepted this invitation - the other two either cited difficulties hosting researchers during the COVID-19 pandemic or did not respond - so we switched to a convenience sampling approach to recruit the two remaining schools. While the final sample contained a diverse range in terms of most school-level factors, we were unable to recruit any schools with an Ofsted rating of 'requires improvement'.

We emailed all teachers in the treatment and control groups a teacher survey at the end of the 12 week intervention period. 19 schools were sent the treatment survey, and 22 schools were sent the control survey. 12 teachers from nine schools responded to the treatment survey[9], a response rate of 47% at the teacher level. Six teachers from five schools responded to the control survey, a response rate of 23%.

Pupil computing attitudes survey validation (exploratory multiple-factor-analysis)

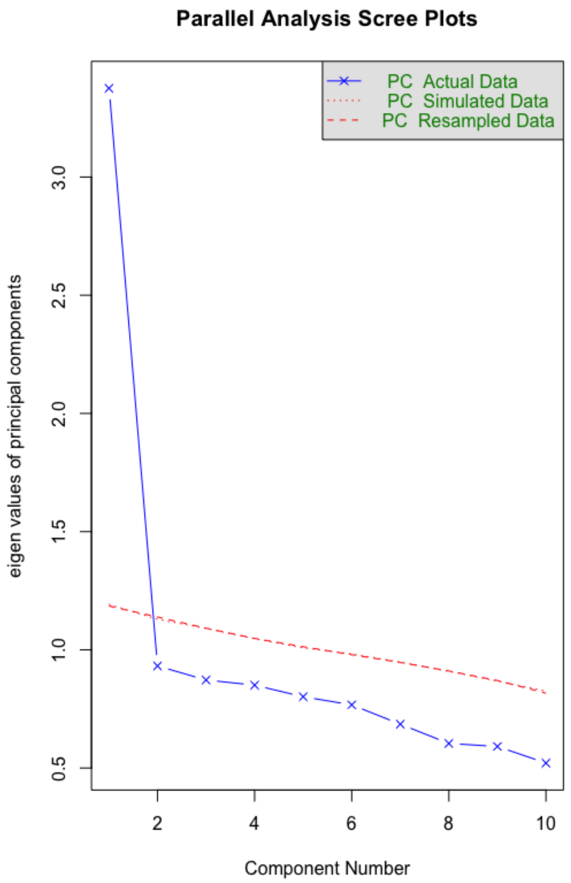
The current design of the computing attitudes survey is a novel contribution of this project given the absence of validated surveys assessing these attitudes for KS1 (Year 2) pupils. To

[9] Some schools used multiple teachers to deliver the Storytelling lessons to their computing class

establish basic face and content validity, the survey has already undergone review by a Year 2 computing teacher specialist, and cognitive testing with a small group of Year 2 pupils. To support this previous work we conducted multiple-factor-analysis to estimate the number of meaningful constructs underlying the survey data. This test is intended to determine whether the survey measures only this one construct, or multiple ones. We want the survey to measure only one “construct”, attitudes towards computing, rather than multiple constructs at once (e.g., attitudes towards computing, but also previous experience with computing and/or computing ability).

The scree plot of the survey variables, presented in Figure 2, represents how much variation each factor (survey item) captures from the data. By comparing each factor’s eigenvalues (where an eigenvalue describes how much variance each factor explains in the data) obtained from the observed dataset and those from a simulated dataset of random values,he scree plot identifies one construct to keep as the first factor has an eigenvalue in the observed dataset that is greater than the eigenvalue of the simulated random dataset. In other words, there are not multiple constructs that are needed to capture the variation across the survey items.

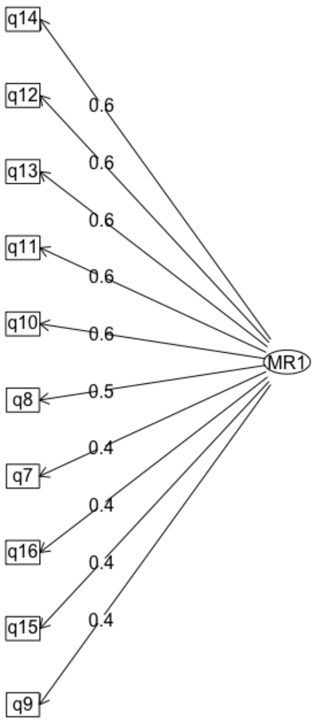
Figure 2: Scree plot of SCSAS scores



The path diagram plot in Figure 3 visualises the relationship between factors and the underlying construct, together with each survey item’s loadings on these factors (i.e., how much the item is influenced by that factor).

This analysis suggests that the items did not capture multiple distinct constructs,suggesting that this survey does not require adaptations prior to being used in a full-scale survey.

Figure 3: Path diagram for multiple-factor-analysis



Intracluster correlation coefficients (ICC)

Table 11 reports the ICC observed for each outcome measure. The purpose of highlighting the intra-cluster correlation (ICC, i.e. within schools) for the primary and secondary outcomes of this trial is to facilitate the preparation of a future full scale trial. For example, the ICCs can be used to develop power calculations to better estimate the required level of recruitment needed to conduct a properly powered trial.

Table 11: Intracluster correlation coefficients observed for primary and secondary outcomes

Outcome	ICC
Primary outcome: general attitudes towards computing (SCSAS score)	0.18
Secondary outcome 1: stated intention to study computer programming in the future	0.11
Secondary outcome 2: stated intention to study science and/or maths in the future	0.10

Note that the ICC value observed for the primary outcome is higher than the range of 0.05-0.10 generally observed in other GBIC trials conducted by BIT for interventions targeting KS2 and KS3 pupils. This is consistent with the tendency for ICC to be larger at lower grades reported in Hedges & Hedberg (2007).

5.5.2. Qualitative evidence

Teachers found the pupil surveys difficult to administer

Teachers commented that pupils were unable to complete the surveys on their own because their reading and comprehension skills were not advanced enough. This meant that the surveys were extremely time consuming to administer as teachers had to provide extensive support to explain and help answer the questions. Teachers also raised concerns that this might lead to unreliable data, as pupils' answers might be influenced by the teacher or other pupils.

6. Conclusions and recommendations

Discussion of findings

The intervention is ready for a full scale impact evaluation, though we recommend minor adjustments to adapt the intervention and evaluation strategy based on the pilot findings. We did not observe any evidence of promise of the effectiveness of the intervention in the quantitative findings, however this was more likely due to methodological constraints than to the intervention itself. While this pilot was never intended to demonstrate programme impact, the high attrition from both treatment and control groups meant that even observations around magnitude and direction of effects cannot be interpreted with confidence. Unfortunately, the impact of COVID-19 on schools' ability to participate in the pilot meant that overall, quantitative evidence was limited.

Evidence from the qualitative data may also help to provide suggestions for improvements to the intervention itself, as opposed to the evaluation design. However due to the limited scope of the qualitative research, these can only be considered partial and not definitive reasons for why there was no observable difference between treatment and control groups on the three outcome measures. Teachers mentioned that they already used different forms of storytelling in order to teach computing to this age group. In this way, it is possible that the intervention was not distinguished enough from business-as-usual computing to notice a difference between treatment and control groups.

Furthermore, as pupils are very young, their previous experience with computer programming is likely to be limited, so potentially mere exposure to the subject was enough to generate more positive attitudes toward computing. This may help explain the low knowledge, experience, and positive attitude towards computer programming at the baseline levels, but also why any exposure to computing, for both treatment and control, may have improved attitudes in both groups. Teachers also highlighted the engagement of both male and female pupils, stating that the intervention was appealing for all pupils, which may go some way to explain why there was no observable difference between genders.

Overall, the intervention was acceptable to teachers and pupils and feasible for teachers to deliver. The training mostly prepared teachers for programme delivery, despite some feedback that greater clarity around more complex coding would have been useful. By contrast, confident teachers actually requested to attend less of the training as it covered topics of great familiarity for them. Both teachers and pupils enjoyed the lessons, and pupils were able to engage with the computing content through their stories. Based on this feedback, the training and content only require minor adjustments before proceeding to trial, and these are listed in the intervention recommendations below.

It seems that the high attrition observed in this pilot may have been due to the COVID-19 context, and thus, we expect attrition for a non-pilot full version of this trial to be lower and more in line with attrition levels seen in the ongoing GBIC trials. The higher attrition rates in the control group between randomisation and baseline surveys suggest that in addition to running the trial without the limitations that COVID-19 brings, extra care may be needed to support schools in the control group to remain in the programme.

Below are adaptations to the intervention and the evaluation we believe will improve the feasibility of a trial and the ability to observe any impact the intervention may have.

Recommended adaptations to the intervention

1. **Find ways to shorten the training or allow teachers to skip content they have covered.** Teachers already familiar with ScratchJR found it difficult to justify devoting their time to the entire training. Instead, it may be more suitable to break the training into modules and allow teachers to attend the necessary sessions. This may allow teachers to strike a balance between preparation for the intervention and their existing workloads.
2. **Ensure the training introduces teachers to all of the content covered in the lessons.** Teachers with less experience teaching computing commented that they did not feel as prepared to handle the more complex topics included in the later lessons. While this was not an issue for experienced teachers, the above recommendation of splitting the training content into modules based on level of experience may also benefit those with less experience to feel comfortable asking for support in an environment with others with similar experience.
3. **Consider allowing teachers to decide which lessons to deliver based on the ability of their class**. Teachers commented on the appropriateness of the lessons for different year levels, and overall, seemed to find the lessons appropriate for a variety of students. There was feedback, however, that the first half of the lessons seemed more appropriate for Year 1 pupils and the more complicated second half was deemed more suitable for Year 2 pupils. It may be helpful in future iterations of the programme to let the teachers make a decision based on the levels and ability of their class, and potentially to also provide some guidance on the complexity upfront.

Recommended adaptations to the evaluation

4. **Review attrition strategies, focusing particularly on schools that are randomised into the control group.** The dropoff in control schools may suggest that schools may not be incentivised to take part in the evaluation if they are not able to receive the programme. It may be helpful to guarantee that they receive all materials at the end of the evaluation, as well as training materials to help teachers navigate the programme. Future implementation should also consolidate data collection into the pre and post surveys, and avoid additional measures such as prefilled classlists.
5. **Conduct further cognitive testing for the survey.** Although initial cognitive testing was conducted (with a limited sample), teachers did provide feedback that further simplification of language would be helpful. To this end, we would suggest conducting more cognitive testing to assess pupil comprehension once further tweaks to the survey are made.
6. **Conduct further validity testing for the survey.** Any analysis estimating treatment effects using this survey should be strongly caveated until more research is done to test its validity. Even after the alterations proposed in recommendation5, and with good teacher support for pupils completing the survey, it is likely that a substantial amount of measurement error will remain.

Annexes

Annex A: Pupil endline surveys and responses

Table 12: Pupil endline survey - individual questions used to construct primary outcome (survey questions 7-16)

Computing attitudes survey					
7	I think I am good at computer programming				Don’t want to say
8	I think it is easy to answer questions in computer programming				Don’t want to say
9	When I have to solve a problem in computer programming, I feel...				Don’t want to say
10	I like computer programming				Don’t want to say
11	When I write a computer program, I feel...				Don’t want to say
12	I wish I could do more computer programming lessons				Don’t want to say
13	In computing programming class, I feel...				Don’t want to say
14	Computer programming is for people like me				Don’t want to say
15	I get on with the people in my computer programming class				Don’t want to say
16	Knowing about computer programming will help me get a job				Don’t want to say

Table 13: Pupil endline survey responses on attitudes towards computing by treatment group and gender

Survey item	Gender	Group	N (non-missing)	Percentage with sad face	Percentage with neutral face	Percentage with happy face
7. I think I am good at computer programming	Girls	Control	157	3.8%	29.3%	66.9%
	Girls	Treatment	179	3.4%	20.7%	76.0%
	Boys	Control	159	1.9%	28.9%	69.2%
	Boys	Treatment	167	5.6%	27.3%	67.1%
8. I think it is easy to answer questions in computer programming	Girls	Control	156	7.1%	42.9%	50.0%
	Girls	Treatment	175	8.6%	43.4%	48.0%
	Boys	Control	153	7.8%	34.0%	58.2%
	Boys	Treatment	161	14.3%	34.8%	50.9%
9. When I have to solve a problem in computer programming, I feel...	Girls	Control	151	6.6%	33.1%	60.3%
	Girls	Treatment	179	9.5%	37.4%	53.1%
	Boys	Control	156	6.4%	31.4%	62.2%
	Boys	Treatment	158	8.9%	31.6%	59.5%
10. I like computer programming	Girls	Control	156	1.9%	15.4%	82.7%
	Girls	Treatment	180	2.8%	8.3%	88.9%
	Boys	Control	154	5.2%	8.4%	86.4%
	Boys	Treatment	164	4.9%	13.4%	81.7%
11. When I write a computer program, I feel...	Girls	Control	154	2.6%	26.6%	70.8%
	Girls	Treatment	176	2.3%	18.8%	79.0%
	Boys	Control	151	3.3%	25.8%	70.9%
	Boys	Treatment	157	8.3%	17.2%	74.5%
12. I wish I could do more computer programming lessons	Girls	Control	154	6.5%	15.6%	77.9%
	Girls	Treatment	181	5.5%	17.7%	76.8%
	Boys	Control	153	7.8%	11.8%	80.4%
	Boys	Treatment	160	10.0%	13.8%	76.3%
13. In computing programming class, I feel...	Girls	Control	154	1.9%	20.1%	77.9%
	Girls	Treatment	179	1.7%	14.0%	84.4%
	Boys	Control	153	5.2%	16.3%	78.4%
	Boys	Treatment	161	7.5%	14.3%	78.3%

14. Computer programming is for people like me	Girls	Control	153	6.5%	23.5%	69.9%
	Girls	Treatment	177	11.3%	22.0%	66.7%
	Boys	Control	146	10.3%	13.0%	76.7%
	Boys	Treatment	160	11.9%	16.9%	71.3%
15. I get on with the people in my computer programming class	Girls	Control	154	1.9%	20.1%	77.9%
	Girls	Treatment	178	6.2%	13.5%	80.3%
	Boys	Control	149	4.0%	13.4%	82.6%
	Boys	Treatment	158	7.0%	8.9%	84.2%
16. Knowing about computer programming will help me get a job	Girls	Control	149	4.7%	34.9%	60.4%
	Girls	Treatment	173	20.8%	32.4%	46.8%
	Boys	Control	155	9.0%	16.7%	69.7%
	Boys	Treatment	150	21.3%	30.7%	52.7%

Annex B: Teacher interview guide

GBIC 1b Pilot: RP Teacher Interview

The interviews should last no more than 45 minutes. The timings given for each section are a guide - you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee says, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introduction	Explains the purpose and 'ground rules' of the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some **background** to the school and the **context** in which the story telling programme is being delivered.	10 mins
3. Delivery experience	This section will focus on understanding the **perceived quality of the intervention,** as well as experiences of the programme's **delivery**, including **barriers and facilitators** to delivery.	10 mins
4. Training	To understand the teacher's experience of the **training** they received to deliver the programme	5 mins
5. Pupil engagement and programme mechanisms	To explore the teacher's perception of **pupils' engagement in the sessions** and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified.	5 mins
6. Research - data collection	To learn about their experience of administering the pupil surveys and views of the survey (including how to improve it)	5 mins
7. Close	Thank you and close.	2 mins
Important: - Remember to probe throughout about gender differences		

1. Introduction	3 mins
Introduction: ● Introduce yourself	Orientates respondent

<table>
<tr><td>

- Introduce BIT and RPF – explain that we are independently evaluating the The GBIC programme, which is funded by the Department for Education, involves the delivery of five interventions by the Coalition for a World-Leading Computing Education (a consortium of STEM Learning, the British Computer Society, and the Raspberry Pi Foundation).

Aims of this interview:

We are here to learn more about how the story telling intervention has worked in your class. We're interested in what involvement you have had with the programme, what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in your class.

This interview:

- Should take no more than 45 minutes
- We want to understand your point of view. No answers are right or wrong

Anonymity and privacy:

- All information gathered will be in strict confidence, unless there are concerns about safeguarding.
- Will not use your name in any reports we write up; we might use quotes, but we will de-identify them first
- If you feel uncomfortable answering a question we can just skip it
- Just say at any point if you want to stop altogether - this is not a problem and you do not have to give a reason for stopping
- Check if they have any questions before starting

Recording:

- Explain you would like to audio record the conversation, to help us capture everything you said correctly, in your own words; will have transcript
- Gain verbal consent (not recorded) and start the audio record

</td><td>

and gets them prepared to take part in the discussion.

Outlines the 'rules' of the interview.

</td></tr>
<tr><td>

2. Background context

</td><td>

10 mins

</td></tr>
<tr><td>

Before we start discussing the storytelling lessons, I'd like to learn a bit more about you and your experience with computing.

How long have you been teaching?

- How long have you been working at this school?

How would you describe your experience with computing?

- Probe around personal experience and teaching computing knowledge and confidence

What are your views about computing teaching in primary schools?

- Which pupils do you think are more interested in computing? Probe about boys vs girls, which age group + why; confidence and knowledge
- What did you think of the GBIC programme when you first heard of it?
  - Why do you think there is a gender gap in those who pursue computing?
  - In what ways do you think pupils experience in primary school …
    - Contribute to this gender gap?
    - Try to reduce the gender gap?

</td><td>

Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the story telling programme is being delivered.

</td></tr>
</table>

<table>
<tr><td>What computing lessons or activities do pupils usually do?
- How would you say these compare to the GBIC storytelling programme?</td><td></td></tr>
<tr><td>3. Delivery experience</td><td>15 mins</td></tr>
<tr><td>We're interested in understanding how you found delivering the story telling sessions.

Let's start with discussing the lesson you did today. From your perspective, how did the lesson go today?

- Probe around
    - Resources available to deliver the sessions
    - Confidence with delivering the content
- Probe about any specific details of what you observed in the lesson regarding
    - Approach to teaching
    - Tools and materials used
    - Pupils engagement in the lesson
- How would you say the lesson today differed from the other lessons?
    - Probe around approach to teaching it, pupil engagement in the lesson etc

Let's talk about the overall programme - how have you found the story telling lessons?

● What do you like about the story telling programme?
● What has been challenging?
    ○ What do you think could have helped to overcome these?
● What recommendations do you have about how the programme could be improved?

How many lessons have you delivered so far?

● Probe around whether delivered all of them or some and why (and intention to deliver the rest of them)
● What preparation did you need to do to deliver the lessons?
● What changes did you make to the…
    ○ Content of the lessons?
    ○ Approach to delivering the lessons?

How have the story telling sessions been organised and timetabled in your curriculum?

● Probe around
    ○ Which day and time they take place + whether same day/time every week
    ○ How long lessons last + any variation in this?
    ○ Where do they take place + any variation in this?
● In what ways have you changed your approach to organising and timetabling the lessons since you first started delivering the lessons?

[If you have time + might come up earlier on in the interview] What resources are needed for you to deliver these lessons?

● What did you think of the lesson plans available?
● What did you think of the Scratch Jr software?
● Any other resources?
● Probe about what works well / could be improved</td><td>This section will focus on understanding the perceived quality of the intervention, as well as experiences of the programme's delivery, including barriers and facilitators to delivery.</td></tr>
</table>

<table>
<tr><td>

[Possible, might not have time] **Let's think about any future teachers that might be delivering the Storytelling lessons. Based on your experience, what do you think might stop a teacher from being able to adequately employ these sessions in their classroom?**

- **Probe around <u>why and what can be done to overcome these challenges</u>**

Prompts:

CAPABILITY

- Understanding what is expected of me as a teacher of this programme
  - Knowledge of what to teach in these session
  - Knowledge of how to teach these session
- Confidence with the content and computing generally
- Computing capabilities
- Didn't have the right skills
- Didn't have the right support from my schools
- Didn't have time

OPPORTUNITY

- Time management
- Too busy / competing teaching requirements

MOTIVATION

- Don't see the value

</td><td></td></tr>
<tr><td>4. Training</td><td>5 mins</td></tr>
<tr><td>

What did you think about the training delivered as part of the GBIC Storytelling programme?

- What worked well about the training sessions?
- What did you find challenging about the training sessions?
- How did the training sessions support you to deliver the storytelling programme?
- If you could change anything about the training, what would it be?

</td><td></td></tr>
<tr><td>5. Pupil Engagement and Programme Mechanisms</td><td>5 mins</td></tr>
</table>

How have the pupils in your class responded to the story telling sessions? - Probe about who engaged more and who engaged less + why - How would you compare the engagement with the lessons amongst girls and boys in your class? - What could be done to improve the engagement of those who engaged less? **What changes have you seen in your pupils as a result of taking part in the story telling programme?** ● Prompts: ○ Knowledge ○ Confidence ○ Interest in computing and learning more about it ○ Attitudes towards computing ○ Possibility of having a career in computing ● How would you compare the difference between the impact of the lessons on girls vs boys? ● Probe around **differences between girls vs boys** in outcomes + why ● What is it about the programme that you think has helped them to develop? ● Can you describe any negative impacts of the programme for pupils? **Who do you think benefited most from the lessons?** ● Who do you think did not benefit from the lessons? ● Beyond your classroom and in the wider school, how do you think this teaching approach would be beneficial for these groups of pupils? **What effect do you think taking part in the story telling programme has had for you?** ● What was it about the programme that you think led to these changes? ● Probe around knowledge and confidence in teaching computing and teaching generally ● In what ways do you think the storytelling approach can be used outside of teaching computing?	To explore the teacher's perception of pupils' engagement in the sessions and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified..
5. Research - data collection process	5 mins
Researcher note: *relevant RQ we want to answer is what are teachers' views on the feasibility of data collection and data handling procedures?* **We would like to know how we can improve the pupil surveys and approach to administering them. What was your experience like with doing the surveys with the pupils?** ● Probe how long it took, when did they do it (all in one go or spread one throughout the week?), how pupils responded to it ● Prompts - what did you think of the … ■ Content of the surveys ■ Online survey platform (ease of navigation etc) ● How challenging was it for you to collect the information from pupils? ■ What worked well and could be improved about our processes ● What would make it easier for you to do the pupil surveys?	5 mins

6. Close	2 mins
Overall, what would you recommend about the storytelling programme to other teachers and schools? ● Why/Why not? **Was there anything else that you were hoping to discuss that we haven't yet had a chance to talk about?** Thank you for your time!	Thank you and close.

Annex C: Classroom observation proforma

GBIC 1b Observation proforma

Important background details

- **Task:** Observing the teacher delivering one of the computing lessons

- **Aim:** to understand how teachers approach delivery and the ways in which children engage in the lessons
    - Researchers may approach pupils as they are working (where parental consent has been provided), to ask questions about their understanding and engagement with the tasks
- **Recording notes:** Suggested approach is to take short bullet points in a notebook during the observations (and quotes if possible) and expand on them further asap after the session
    - **Keep in mind: try to observe and compare how girls vs boys interact with the lesson materials, each other and the teacher**

- **IMPORTANT - Before the observation:**
    - Review the content for the specific lesson that you'll be observing (see here) so you know what to look out for and can spot anything that is missing / extra (make note of this in the reflective notes column)

TIPS for writing observation notes
• **Golden rule:** Record what you see & hear (setting, people, behaviours, words, activities) • **Two types of notes (please distinguish which is which when taking notes):** 1. **Descriptive:** Use descriptive words to document what you observe. Being descriptive means supplying yourself with enough factual evidence that you don't end up making assumptions about what you meant when you write the final report. - Ex: instead of noting that the pupil is "engaged," state that the pupil is talking to other pupils and the teacher; they are asking and responding to all questions. 2. **Reflective notes:** Note ideas, impressions, thoughts, any questions and judgements (what works well / not well) you have about what you observed.

Date and time: **Name of school and class:** **Number of pupils present:** **Name of observer / researcher:** **Other adults present:**

Topic	Observation notes	Reflective notes
Description of the space (classroom & school) **-**chairs, tables, windows etc -student formation (how seated; small/ large groups; visible to teacher)		
Description of resources used (eg tech, how many/ pupil)		

Describe the people (how many, any other info you learnt about them eg ages, gender etc) - Staff & pupils		
Describe lesson content and activities -which lesson is it - topics covered -activities they need to do -Aims / expectations / pupils Qs or concerns		

Teacher approach to delivery: -presentation, then Qs and activities? Any games etc? -confidence [note down behavioural indicators of why you think they look confident or not]		
Teachers approach to supporting pupils -how they respond to pupil questions		

and concerns -any accommodations / changes made for those with different needs?		
Describe pupil engagement & behaviour: **How are pupils responding to the content and activities?** **-**note any conversation topics or questions asked; 1:1 or group work etc -any difficulties with specific content / activity -behaviour and body language (looking at teacher, laughing, side conversations etc)		

-differences btw girls and boys		
Other		

Annex D: Teacher survey

GBIC i1b - Treatment Survey

Background and Consent

Section 1: Screening ____

You are invited to share your experiences with teaching computing in this 15-minute survey which is conducted as part of the Gender Balance in Computing (GBIC) programme evaluation.

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

- Yes, I am a classroom teacher who has delivered the GBIC Storytelling project lessons in my school
- No, I am NOT a classroom teacher who has delivered the GBIC Storytelling project lessons in my school
    - [If this is selected, an error message pops up]: "*Thank you for your interest in filling out this survey. Unfortunately, this survey focuses on teacher experiences of delivering lessons and so needs to be completed by teachers who are participating in the GBIC programme. Please email storytelling@bi.team to make the Behavioural Insights Team aware that this does not apply to you / let them know which teacher they should share this link with instead.*"

[next page] ____

The purpose of the survey is to understand your views and experience with delivering the GBIC Storytelling lessons to your KS1 class. **It should take around 10-15 minutes to complete.**

How your data will be used
The survey is being carried out by the Behavioural Insights Team (BIT) and the Raspberry Pi Foundation. This work is done on behalf of the Department for Education who is the data controller for all information collected as part of this programme of work. We will collect your responses to the survey only for the purposes of the GBIC programme evaluation.

All data will be stored on a secure, encrypted server and access will be limited to the immediate research team. The data will be aggregated and analysed by the research team and included in an evaluation report. This report will not include any information that can identify you (i.e., your name, name of the school). All data will be destroyed six months after the submission and approval of the final report by the Department for Education, which is expected to be in late 2022.

You can change your mind about taking part at any time and withdraw your data from the evaluation up until we begin analysis on July 30, 2021. You don't need to give a reason to withdraw your consent, just let us know by emailing storytelling@bi.team.

For more information about how we collect and process your personal data, please refer to the Department for Education's Privacy Notice.

More Information

Thank you for reading this. If you have any questions or would like more information about this research, please contact storytelling@bi.team. You can also contact BIT's Data Protection Officer at dpo@bi.team if you have any questions regarding your personal data.

Our lawful basis for processing your personal data is legitimate interests (as per Article 6 (1) (f) of the GDPR) and we have considered that your interests and fundamental rights do not override those legitimate interests. It is necessary in BIT's 'legitimate interests' to process the personal data identified above in order to conduct the GBIC programme evaluation, which has been commissioned by the Department for Education. The research project fulfils BIT's core business aims including undertaking research, evaluation and information activities in sectors that will deliver social impact.

Consent

You are free to decide whether you'd like to take part in the survey. If you would like to participate, please provide your name and the name of your school below:

First name: ___________________________

Last name: ____________________________

Name of the school you work in: ____________________________

Optional - We will be conducting 15-20 minute follow-up interviews with some teachers to learn more about their experience of delivering the programme. If you are happy for a researcher to contact you about a potential interview, please provide your contact details below (this does not commit you to being interviewed):

Email: ____________________________

Phone number: ___________________

Click 'next' to proceed to the survey on the next page.

[next page] ___

1. **How confident are you in teaching computing to KS1 pupils?**

- Likert scale: 1 (not confident at all) - 5 (Very confident)

Computing in your school

2. **What types of computing-related activities are available to KS1 pupils in your school?**
    - Lessons in computing (not including the GBIC Storytelling lessons)
    - After-school computing clubs
    - Lunchtime computing clubs
    - Other: [free text]

[next page] ___

Your experience with the GBIC Storytelling training

3. **Did you take part in the online training for the GBIC Storytelling programme?** Please tick one box.

- Yes, I completed <u>all three</u> online sessions (Session 1: introduction, Session 2: getting started, and Session 3: investigating resources). *If selected 'yes', jump to Q5*
- Yes, I completed <u>one or two</u> of the online sessions*. If selected, continue with Q4 below.*
- No, I did <u>not</u> complete any of the online training sessions. *If selected, continue with Q4 below. Skip Q5 and q6.*

4. **What was the reason(s) why you were not able to complete all of the training sessions?** Check all that apply.

- I did not have time to complete all of the sessions/I was too busy during the school day
- I was not supported to take time out of the school day to complete the sessions
- I did not feel like the sessions would be useful
- I did not know about the training sessions
- I could not access the training sessions
- Other:

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

- Likert scale 1 (not well at all - very well)

6. **How do you think the training could have been improved to better prepare you for delivering the lessons?**

[Free text box]

Your experience with delivering the GBIC Storytelling lessons

7. **How many of the 12 GBIC Storytelling lessons were you able to deliver between April and July 2021?** [select all that apply]

- All of them
- Some of them (select all that apply)
    - Unit 1, lesson 1 - What's ScratchJr?
    - Unit 1, lesson 2 - Once upon a time…
    - Unit 1, lesson 3 - Repetition, repetition, repetition
    - Unit 1, lesson 4 - Timing is everything
    - Unit 1, lesson 5 - Two's company
    - Unit 1, lesson 6 - Let's get moving

    - Unit 2, lesson 1 - Speech bubbles
    - Unit 2, lesson 2 - Messages
    - Unit 2, lesson 3 - I'm going to bump into you
    - Unit 2, lesson 4 - I'm going to tap you
    - Unit 2, lesson 5 - Drawing ideas
    - Unit 2, lesson 6 - Let me tell you a story
- None of them (skip questions 9-16)

8. **(if some or none) What were the main challenges you experienced in delivering all of the lessons between April and July 2021?** Select all that apply

- Pupils did not enjoy the lessons
- I did not feel confident delivering some of the lessons
- I was too busy to deliver the lessons
- It was hard to prioritise this programme over other lessons
- I did not always have access to the right equipment/facilities to deliver the lessons
- I did not have the teaching materials
- I did not always have the right support staff to deliver the lessons
- Other: [free text]

9. **How many school weeks (i.e. not including school holidays or half-term) did it take you to deliver the lessons?**
    - Less than 12 weeks
    - 12 weeks exactly
    - More than 12 weeks

10. **(if less or more) What were the main reasons it took less/more time than 12 weeks?**
    (select all that apply)

- (less than 12 weeks)
- I did not deliver all of the lessons
- I was sometimes able to combine lessons into individual sessions because they were shorter to deliver than expected
- I was sometimes able to deliver more than one lesson in a single week
- Pupils advanced through the lessons more quickly than expected
- Other: (free text)

(More than 12 weeks)
- I did not always have time to deliver one lesson every week
- I sometimes had to split lessons over two/multiple sessions because it took longer than expected to go through the content
- Pupils did not advance through the lessons as quickly as expected
- Other school plans sometimes took priority over the storytelling lessons (e.g. events, school trips, other lessons)
- Other (free text)

For each lesson, 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 plan, and what adaptations you might have needed to make.

11. **How closely do you think you followed the content set out in the lesson plan for each lesson?**
    Likert scale - 1 Not at all closely - 5 Very closely

12. **Did you need to change or adapt the lesson content in any way?**
    Yes
    No

13. **(if yes) Which lessons did you need to adapt the lesson content for?**
- All of them
- Some of them (please select)
- Unit 1, lesson 1 - What's ScratchJr?
- Unit 1, lesson 2 - Once upon a time…
- Unit 1, lesson 3 - Repetition, repetition, repetition
- Unit 1, lesson 4 - Timing is everything
- Unit 1, lesson 5 - Two's company
- Unit 1, lesson 6 - Let's get moving

- Unit 2, lesson 1 - Speech bubbles
- Unit 2, lesson 2 - Messages
- Unit 2, lesson 3 - I'm going to bump into you
- Unit 2, lesson 4 - I'm going to tap you
- Unit 2, lesson 5 - Drawing ideas
- Unit 2, lesson 6 - Let me tell you a story

-

14. **(if yes) Please provide details about how and why you adapted the content for some lessons**
(free text)

The impact of the programme

15. Thinking about girls in your class, what kind of impact do you think the GBIC Storytelling programme has had on their _____?

	Very positive impact	Somewhat positive impact	No impact	Somewhat negative impact	Very negative impact
Computing skills	❏	❏	❏	❏	❏
Confidence in computing	❏	❏	❏	❏	❏
Enjoyment of computing	❏	❏	❏	❏	❏
Interest in doing more computing	❏	❏	❏	❏	❏
Belief that computing is for people like them	❏	❏	❏	❏	❏

If you have selected 'somewhat negative' or 'very negative' for any answers, please provide further details below: [free text]

16. Now thinking about boys in your class, what kind of impact do you think the GBIC Storytelling programme has had on their ______?

	Very positive impact	Somewhat positive impact	No impact	Somewhat negative impact	Very negative impact
Computing skills	❏	❏	❏	❏	❏
Confidence in computing	❏	❏	❏	❏	❏
Enjoyment of computing	❏	❏	❏	❏	❏

Interest in doing more computing	❏	❏	❏	❏	❏
Belief that computing is for people like them	❏	❏	❏	❏	❏

If you have selected 'somewhat negative' or 'very negative' for any answers, please provide further details below: [free text]

Your experience with the GBIC storytelling programme

17. **Overall, how would you rate the quality of the GBIC Storytelling programme?**

- Poor
- Fair
- Good
- Very good
- Excellent

18. **Overall, would you prefer to run the GBIC storytelling lessons in place of your normal computing lessons with this age group?**

    -Yes

    -No

    -Maybe

Please briefly explain your answer: [free text]

19. **How do you think the GBIC Storytelling programme could be improved?**

[open text]

20. **Any other comments:** [open text]

Thank you for completing this survey

GBIC i1b - Control Survey

Background and Consent

Section 1: Screening

You are invited to share your experiences with teaching computing in this 5-minute survey which is conducted as part of the Gender Balance in Computing (GBIC) programme evaluation.

This survey needs to be completed by the **teacher who is participating in the GBIC programme (which includes the control group who have not been delivering any GBIC Storytelling lessons) - please confirm whether this applies to you** [check box]:

- Yes, I am the classroom teacher who is participating the GBIC programme
- No, I am **NOT** the classroom teacher who is participating in the GBIC programme
    - [If this is selected, an error message pops up: *Thank you for your interest in filling out this survey. Unfortunately, this survey focuses on teacher experiences of delivering lessons and so needs to be completed by teachers who are participating in the GBIC programme. Please email storytelling@bi.team to make the Behavioural Insights Team aware that this does not apply to you / let them know which teacher they should share this link with instead.*

[next page] ____

The purpose of the survey is to understand your experience with computing and the computing-related activities available to pupils in your school. **It should take around 5 minutes to complete.**

How your data will be used

The survey is being carried out by the Behavioural Insights Team (BIT) and the Raspberry Pi Foundation. This work is done on behalf of the Department for Education who is the data controller for all information collected as part of this programme of work. We will collect your responses to the survey only for the purposes of the GBIC programme evaluation.

All data will be stored on a secure, encrypted server and access will be limited to the immediate research team. The data will be aggregated and analysed by the research team and included in an evaluation report. This report will not include any information that can identify you (i.e., your name, name of the school). All data will be destroyed six months after the submission and approval of the final report by the Department for Education, which is expected to be in late 2022.

You can change your mind about taking part at any time and withdraw your data from the evaluation up until we begin analysis on July 30, 2021. You don't need to give a reason to withdraw your consent, just let us know by emailing storytelling@bi.team.

For more information about how we collect and process your personal data, please refer to the Department for Education's Privacy Notice.

More Information

Thank you for reading this. If you have any questions or would like more information about this research, please contact storytelling@bi.team. You can also contact BIT's Data Protection Officer at dpo@bi.team if you have any questions regarding your personal data.

Our lawful basis for processing your personal data is legitimate interests (as per Article 6 (1) (f) of the GDPR) and we have considered that your interests and fundamental rights do not override those legitimate interests. It is necessary in BIT's 'legitimate interests' to process the personal data identified above in order to conduct the GBIC programme evaluation, which has been commissioned by the Department for Education. The research project fulfils BIT's core business aims including undertaking research, evaluation and information activities in sectors that will deliver social impact.

Consent

You are free to decide whether you'd like to take part in the survey. If you would like to participate, please provide your name and the name of your school below:

First name: ___________________________
Last name: ____________________________

Name of the school you work in: ___________________________

Click 'next' to proceed to the survey on the next page.

[next page] ____

About your experience with computing

1. **How confident are you in teaching computing to KS1 pupils?**

- Likert scale: 1 (not confident at all) - 5 (Very confident)

Computing in your school

2. **What types of computing-related activities are available to KS1 pupils in your school?**

- Lessons in computing
- After-school computing clubs
- Lunchtime computing clubs
- Other: [free text]

3. **Thinking about your KS1 class, since April 2021 how often have they received a computing lesson (on average)?**

- More than once a week
- Once a week
- Every couple of weeks
- Once a month
- Less than once a month
- Never (skip to q 6 if selected)
- Other:

4. **Thinking about computing lessons your class has received since April 2021, were any of the following types of content covered in those lessons?** (Please select all that apply)
    - Programming an object or character to move
    - Sequencing computing instructions
    - Using computing/programming to tell a story
    - Repetition and the use of loops in computing
    - Changing the speed of a computing animation
    - Programming characters to speak to one another (e.g. through speech bubbles)
    - Programming different objects or characters to physically interact with one another
    - None of the above
    - Other: [free text]

5. **Which of the following have you done when teaching computing to your class this academic year (2020-2021)?**

- Used externally-developed step-by-step computing lesson plans
- Used externally-developed teaching software such as ScratchJr
- Provided opportunities for children to engage independently with a computing task
- Other:

Your experience of computing and teaching computing

6. **Since March 2021, have you received any training or Continuing Professional Development (CPD) focused on developing <u>your own knowledge or skills in computing</u>?**
    - Yes
    - No
    - Other:

7. **Since March 2021, have you received any training or Continuing Professional Development (CPD) focused on <u>teaching computing</u>?**
    - Yes
    - No
    - Other:

Your experience with the GBIC programme

8. **What types of computing-related training, support or materials have you received as part of the GBIC programme?** Select all that apply.

- I have not received any computing related training, support or materials
- I attended an introductory webinar

- I attended a training session(s)
- I received computing teaching materials (e.g. lesson plans)
- I received guidance from an experienced computing teacher

9. **Any other comments:** [open text]

Thank you for completing this survey

Annex E: Pupil discussion guide

GBIC 1b Pupil Discussion Guide

Important background details

- **Aim of the discussion:**
    - **BAU & background:** their views of computing and what computing-related activities they do in school
    - **Engagement:** How pupils have found taking part in the intervention lessons, including what they liked and disliked
    - **Impact:** What they have learnt during the lessons
- **Participants:** Year 2 (aged 5-7)
- **Method:**
    - Will take approx 15-30mins
    - In-person after the lesson observation
    - Approach will need to be tailored depending on what's agreed with the teacher (eg speaking to a group of pupils or full class or 1:1 - please see Sampling and Recruitment guide for further details;
- **Session will NOT be audio recorded; time should be allocated asap after the discussion to write down what is said (and take some notes during the session if possible) + take photos of any output (eg sticky notes)**
- Before + during fieldwork:
    - Check if there are any pupils in the class for whom parents opted out / withdrew consent (note down names);
    - Ask the teacher to point out which pupils they are and don't take any notes of what they said;

Preparation

- **Materials:** Print the emoji's and bring sticky notes and extra markers / pens

General best practice

- Keep a good pace when moving through discussion and activities to maintain engagement
- Listen and respond to the contributions of all pupils - even those seemingly off-topic or less relevant. This will ensure they feel comfortable contributing further.
- Use simple and accessible language in discussions
- **Options for getting pupils to share their views:**
    - Go around in a circle and ask all to share (but they can say no if they don't want to)
    - Teacher or researcher chooses which pupils
    - Ask them to raise their hand
    - Lollipop sticks (sticks have childrens names on them and are drawn out of the cup at random)

Topic	Discussion points
Intro	• Say hello again + who you are

- Here to hear from you about what you've been doing during the computing lessons, what you like and don't like
- Describe activity you're going to do depending on what was agreed
- Emphasize that:
    - They don't have to participate if they don't want to
    - There are no right or wrong answers - I just want to hear about what you think
- As we talk, I might take a few notes about what you say so I can remember and share it with my team too
- Ask if they have questions

[Teacher and researcher should participate too]

NB. These are a list of activities you might want to do - pick 1-2 depending on time available and set-up of the session (e.g., full class, small group, etc)

Thanks for having me join your lesson today.

I'd like to hear from you about what you thought of it. Let's all write down our answers on a sticky note and then we can go around in a circle and share what we wrote down.

- *Prep:* Print this and put it on the board in the classroom (label it 1-5 in case some just want to write down the number)

  I

- *Important:* When pupils are sharing what they put down - probe to learn more about why they feel that way

1. **How did you feel about the lesson today?** Draw one of the emoji's or write the number down on the sticky note.

***Activity:* [Thumbs up / down]**

Let's do an activity where we close our eyes and put our head down on the desk. I will ask a question and you can answer by showing a thumbs up to say 'yes' or thumbs down to say 'no'. Then we open our eyes, look around and chat about our answers. For example, if I say my favorite subject is PE; I would show a thumbs up.

- *Take note of anyone who changed their thumbs up / down*

- I like computing
- I am better at coding
- Using computers is fun
- I want to have a job in computing when I grow up
- Girls are good at computing
- Boys are good computing
- Computing is fun
- Computing is creative

[Sentence completion activity]

I'd like you to finish some sentences: Let me give you an example. So if I was asked to complete the following sentence 'My favorite part about going to school is …' I would finish it with 'Doing PE'.

Now your turn…

1. My favourite part about going to school is …

Now let's think about the **computing lesson today** specifically.

1. My favorite part about the computing lesson today is …
2. The part I didn't like the computing lesson today is …
3. Today I learned …
4. Today I was confused about …

Let's think about all the **computing lessons you did in the last few weeks.**

1. Learning about computing …
2. Learning about computing makes me feel …
3. Using ScratchJ is …
4. I would like computing more if …
5. Computing is fun when …
6. Computing is boring when …
7. I would find learning computing easier if …

Instructions: **[Drawing activity and/or word association]** Options presented below - tailor depending on resources available

- Might want to ask them to write it down or draw it and then share
- **Probe about details, to further describe their chosen word and why they picked it (if possible)**

Let's play a game where we say the first thing that comes to mind about a specific word. For example, if I say 'what comes to mind when I say animal' and I would respond with 'cat'.

Now your turn …

- What comes to mind when I say
  - Animal
  - Computer
  - Coding
  - School
  - Scratch Jr
  - Computing teacher
  - Computing job

Let's do some drawing. Imagine someone who works in computing. Can you draw this person?

- When they share, probe more about gender, age, how they look (what they wear etc)

Wrap up	- Ask if there's anything else they'd like to tell you about computing

	- Thank the pupils for their time and collect their sticky notes - Take notes asap after the session

Annex F: Pupil worksheet

Storytelling project!

Look at the faces below. Put a tick under the face that shows how you felt about today's lesson!

In the box below, tell us why you chose to tick this face

Imagine someone who works in computing

Can you draw this person?

Annex G: Programme logic model