Gender Balance in Computing

Evaluation of the i3 Belonging intervention

May 2022

Adam Jones, Clément Bisserbe and Neus Torres Blas

THE BEHAVIOURAL INSIGHTS TEAM

Contents

Executive summary

Overview of the project

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

The GBIC programme has been funded by the Department for Education (DfE), with the Behavioural Insights Team (BIT) and Women into Science and Engineering[3] (WISE) implementation teams serving as the primary delivery organisations, Raspberry Pi Foundation (RPF) overseeing the project, with BIT also acting as independent evaluators[4]. This report details the evaluation of an intervention in the Belonging strand of the programme, where the aim was to improve girls' attitudes towards computing by increasing girls' sense of Belonging within computing. Specifically, the Belonging approach consists of two different interventions aimed at increasing 9-10 year old girls' sense of membership and acceptance in the computing community via two different approaches.

The first intervention, called *"My Skills My Life"* and designed by WISE, aimed to increase self-identification by introducing young girls to women role models who had a career relating to computing. The second intervention, called "*Code Stars*", was designed by BIT as a 12-week programme involving a computing-themed, one-off lesson on artificial intelligence (AI) delivered by teachers, a homework task for pupils to complete with their parents or carers, and a follow-up lessons with teachers. Teachers received an action guide with evidence-based tips on how to engage girl pupils in the class. Throughout the programme, schools were asked to share weekly prompts with pupils' parents, which were designed to encourage conversation between pupils and their parents about computing in general and the AI lesson. Both interventions were delivered between September 2021 and February 2022.

---

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

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

[3] WISE is an independent community interest company whose aim is to achieve gender equality in STEM by driving diversity in sectors across the UK. Their programme *My Skills My Life* is one of many initiatives they have taken to increase diversity in STEM in the UK.

[4] BIT acted as both evaluator and implementer in this trial, but the two teams worked separately from one another.

Evaluation approach

The intervention was evaluated using a mixed methods approach. The quantitative impact evaluation focused on whether there was evidence that the intervention impacted a) girls' sense of belonging within computing, b) girls' attitudes towards computing, c) stated intention to study computing in future, and d) stated intention to study science, maths or technology subjects in the future. In parallel, an implementation and process evaluation (IPE) was conducted to explain the impact evaluation findings and explore implementation processes and possible mechanisms of change in targeted outcomes.

Impact evaluation

The evaluation design was a three-armed cluster randomised controlled trial that aimed to estimate the impact of the intervention, and was randomised at the school level with outcomes at the pupil level. The three arms of the trial were:

1. Control: Schools in the control arm taught usual computing lessons for a 12-week term in Autumn 2021[5].
2. *My Skills My Life* Intervention: Schools in this group received training materials and class plans for a 12 week term of computing lessons aimed to increase girls' sense of belonging within computing. In one of the lessons, a female role model working in computing attended the school.
3. *Code Stars* Treatment: Schools in this group received materials to deliver two stand-alone lessons and an action guide on gender-inclusive teaching to be used by teachers in the two lessons and normal computing classes. Schools were also given a series of prompts that they were asked to send to parents over a 12-week period. The messages were designed to encourage conversations between parents and pupils about the stand-alone lessons and computing in general, and teachers could send them via a number of different channels.

The primary outcome was measured using the Belonging sub-scale of the Student Computer Science Attitude Survey (SCSAS), a validated survey tool for assessing attitudes toward computing for school pupils. In total, 252 schools were originally recruited to participate in the trial, and 175 completed it (submitted endline survey data).

Implementation and process evaluation

Alongside the impact evaluation, a mixed-methods IPE was conducted to answer the following research questions:

1. What are the programs related to computing education (in addition to the interventions delivered) that are available to pupils in the schools?
2. Which components of the interventions were delivered and in what ways (both interventions)? *[both interventions]*
3. How did the parents engage with their child after receiving the text message prompts? *[Code Stars intervention]*

[5] This document outlines the key outcomes for the computing curriculum in the UK: https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study/national-curriculum-in-england-computing-programmes-of-study

4. What were teachers' views about and experiences with delivering the intervention lessons *[both interventions]*, as well as their views and experiences of having the role models in their class? *[My Skills My Life intervention]*
5. What were the girls' views of the interventions and in what ways did they engage with it? *[both interventions]*
6. What were the ways in which the intervention impacted girls' attitudes towards computing? *[both interventions]*

BIT researchers spoke with teachers at four schools where the interventions were delivered to understand teachers' experiences of delivering the intervention, alongside the feasibility of delivery, any programme adaptations and the perceived outcomes for girls. The final sample for the IPE consisted of two schools for each intervention. BIT researchers also conducted small group discussions with pupils and lesson observations in each of the four schools, to better understand how pupils experienced the intervention and how the intervention was being implemented in practice.

This was complemented by two interviews with role models who delivered part of the *My Skills My Life* intervention, and another two with parents who participated in the *Code Stars* programme (which was implemented by a different team than the evaluation team). Due to recruitment difficulties stemming from COVID-19 and low parental engagement with the evaluation, BIT could interview two parents from only one of the schools participating in *Code Stars*, which reduced the representativeness of the sample for the IPE.Finally, BIT carried out an online survey with teachers that took part in the *Code Stars* intervention at the end of the programme.

Findings

Evidence of impact

We found no conclusive evidence that the *My Skills My Life* or the *Code Stars* interventions had an impact on the following outcomes: a) girls' sense of belonging in computing, b) their attitudes towards computing, and c) their intention to study computing in the future. While the differences found between the primary outcomes in each of the treatment groups relative to the control group are positive, they are both small in magnitude and not statistically significant. Further, given the loss in precision and possible biases resulting from the observed attrition and baseline imbalances on the Belonging sub-scale survey score, we would advise against interpreting these differences as indicative evidence of a meaningful effect of the interventions. An additional result to note is that stated intention to study computing was 5.6 percentage points higher in the *Code Stars* intervention than in the control group, which was a significant difference before adjusting for multiple comparisons. While this could be interpreted as an indication of evidence of promise, it should be taken with caution given the differential attrition observed and the lack of impact detected on other outcomes.

While this did not translate into evidence of impact on the measured outcomes, the *My Skills My Life* intervention was positively received and seemed to address the barriers targeted

The IPE findings showed a very positive reception of the *My Skills My Life* intervention by teachers, pupils and role models. Disruptions due to COVID-19 delayed the start of the programme and forced some schools to change the format of lessons from in-person to online, where pupil engagement appeared lower in case study schools. Despite these disruptions, teachers at case study schools reported following the content provided by WISE to plan the lessons, and only making minor adjustments to shorten them given the large amount of content. Sessions with role models took place at all but two schools, where teachers took over the lesson with the help of additional content provided by WISE. These sessions were very popular with pupils, and role models adapted the lesson content depending on their individual preferences and STEM field.

We found some evidence of the intended mechanisms hypothesised to support an increase in pupils' sense of belonging and positive attitudes towards computing, including increased interest and motivation to learn more about computing careers, increased awareness of computing and STEM careers that could fit their skills and preferences, and increased confidence with computing.

The IPE findings suggest that the *Code Stars* lessons were enjoyed by pupils and that the intervention targets a relevant barrier, but the impact of the messaging component may have been limited by delivery and dosage issues

The IPE findings suggest that teachers delivered the two core lessons on AI and used the tips for inclusive teaching in the action guide as intended, and the lessons were described as fun and able to engage the pupils' interest in coding and the importance of computing. The delivery mechanism for the messages to parents component of the intervention was initially meant to be handled by BIT, but this had to be left to the teachers in response to the COVID-19 context, as school resources were being over-stressed due to the COVID-19 pandemic and could not deal with the administrative burden needed to comply with data privacy protocols.

Even though the conversation prompts provided by BIT were found easy to use by teachers, they struggled to follow the recommended weekly schedule when sending the messages to parents; COVID-19 disruptions and competing priorities for teacher and parent attention may have contributed to this challenge. Six out of 8 teachers in the teacher survey reported having sent the messages in some form, but 4 out of 8 teachers sent them with monthly frequencies or all at once, which may have limited their effectiveness in engaging parents. This limited teacher compliance in sending prompts and the low parental engagement reported by some teachers suggest that the intervention may not have increased discussions on computing at home as much as intended, though the parent interview sample limited our ability to investigate this mechanism.

There are a number of possible reasons for the lack of evidence of impact despite the positive reception to the interventions

There are multiple possible reasons related to the design of the *My Skills My Life* and *Code Stars* interventions, their delivery, and measurement challenges for the absence of conclusive evidence of impact despite these positive reported experiences:

- The COVID-19 challenges may have resulted in a lower dosage of the intervention than necessary to have an impact. Challenges included lower pupil and teacher attendance, changing some of the visits to schools by role models from being in-person to online, and lower school engagement.

- Given the need to rely on short-term and pupil-reported proxy indicators to measure impact, the outcome measures may not have been sensitive enough to capture differences in pupils' attitudes and intentions towards computing.
- The sample could have had particularly high baseline engagement with computing (due to its engaged teachers), leaving less scope for improvement as a result of the intervention.
- It is possible that the intervention may set off a chain of mechanisms which eventually contribute to intended outcomes, but that the changes in these outcomes were not yet apparent when measured immediately after the intervention.

There is not sufficient evidence to confidently determine the extent to which each of these factors may have influenced the evaluation results. The data collected as part of the evaluation also confirmed the relevance of interventions targeting gender balance in computing at Key Stage 2 level: in our sample, at baseline the difference between boys and girls was largest in their reported intention to study computing in the future (71% vs 56%).

Recommendations

The following steps could help to build on the positive teacher and pupil response to the *My Skills My Life* intervention:

1. **Facilitate teacher delivery of the intervention through reduced content or increased guidance**
   The feasibility of delivering the intervention could be improved by reducing the amount of content per lesson and time required to prepare for the lessons, or reducing the number of lessons, to reduce the disruption with the usual curricular computing lessons.

2. **Provide additional opportunities for pupil engagement with role models**
   Given the high pupil engagement with the session involving a role model and the potential need to increase the dosage of the intervention, providing more opportunities for direct engagement between pupils and female role models working in computing-related fields could reinforce any potential positive effects of the intervention while lessening the burden for teachers.

The following actions could improve the delivery and potential impact of the *Code Stars* intervention:

3. **Increase number of lessons to reinforce targeted mechanisms of impact**
   Teacher feedback on the action guide and the rest of resources provided for the AI-themed lessons, and pupil engagement in the lessons were both positive in the case study schools and indicative that the lessons were targeting the right barrier. Doing more lessons could help to increase girls' interest in computing more substantially and expose them to prosocial aspects of computing.

4. **Review the mechanisms for teachers to communicate messages to parents**
   The communication mechanism initially envisioned for this intervention, SMS messages, had to be revised based on feedback from schools relating to the

COVID-19 challenges they were facing at the time of implementation. The IPE findings suggest that the alternative mechanisms made available to teachers, such as emails and hard copies given to children as homework tasks, should be refined, and that providing clear guidance and additional reminders to teachers may be more effective than providing flexibility to choose the communication format and timing.

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

5. **Explore potential positive interactions between GBIC initiatives**
   For example, the IPE identified potential synergies between the Belonging interventions and the Informal Learning - Code Clubs intervention. This link was made in one of the *Code Stars* intervention messages to parents. Providing opportunities for pupils to access both types of interventions could enable girls to further explore their new interests sparked by initiatives like the Belonging interventions through extracurricular computing activities such as code clubs.

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

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

7. **Identify strategies to measure outcomes targeted by the intervention further into the future**
   While it requires planning, collaboration and resources, tracking relevant behavioural outcomes (in this case, subject choice from Year 10 onwards) multiple years after the intervention would also greatly enhance the ability to evaluate the impact of early interventions over a time horizon in line with the mechanisms and barriers hypothesised, and thus identify the most impactful ones. In this case, attempting to collect and analyse data on whether pupils in the evaluation sample select computer science as a GCSE subject once the choice arises would enable the estimation of the impact of the intervention on the long-term outcomes targeted, in addition to the short-term proxy indicators used in this evaluation.

8. **For any future adaptations or new interventions, consider additional small-scale piloting to refine delivery prior to a full-scale impact evaluation**
   Piloting interventions in school is complicated given the school staff involvement and coordination with schools it requires, particularly in the recent COVID-19 context. However, the possible improvements to the delivery of both interventions identified

through the IPE illustrate the value of small-scale piloting to inform improvements to the impact potential of any intervention before moving to a full-scale impact evaluation. Where possible, strategies to evaluate interventions at incremental scale and cost should be explored to maximise learning and resource efficiency.

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

1. Background

1.1 Gender Balance in Computing

Computing has a decades-old problem with gender imbalance with limited reliable evidence of what works in closing the gap[6]. Across England, only 6% of girls choose to continue studying Computer Science at GCSE level, and in 2020, only 21% of pupils taking GCSE computer science were female. The Gender Balance in Computing Project (GBIC) aims to tackle a number of known and well-researched barriers to girls engaging with computing[7], including a disconnect between extra-curricular computing activities and subject choice; a lack of encouragement to study computing; a lack of familial and other role models in computing and a perceived lack of relevance of computing to pupils. These barriers are addressed in the five intervention strands that comprise GBIC, with the common goal of increasing the number of girls who study GCSE and A Level Computer Science.

1.2 The 'Belonging' interventions

The gender imbalance in computing studies and computing careers means that girls don't see themselves represented in computing. Computing classes at GCSE and A-level, which historically have few or no girls taking the subject, may deter girls in lower years from selecting the subject. This can create a cycle of low participation because there is no 'critical mass' of girls participating, and being the only one or one of a few girls studying computing may be off-putting. Girls' sense of belonging in STEM is lower than boys' and strongly predicts their STEM interest and motivation[8] [9].

The need to belong - the need to form interpersonal attachments - is a fundamental human motive[10]. In this case, belonging to computing is defined as one's feelings of membership and acceptance in the computing domain. The goal of this suite of GBIC interventions was to address the low level of sense of belonging in computing and STEM careers of girls. It consisted of two different interventions aimed at increasing young girls' sense of membership and acceptance to computing via two different methods. The first intervention (*My Skills My Life*) was designed by WISE and aimed to increase self-identification by introducing young girls to women role models who had a career related to computing and using ten lessons on the importance of STEM and women in STEM. The second intervention (*Code Stars*) was

[6] Royal Society. (2017) After the reboot: computing education in UK schools. https://royalsociety.org/~/media/policy/projects/computing-education/computing-education-report.pdf
[7] Childs, K (2021) Factors that impact gender balance in computing. In Understanding computing education (Vol 1). Proceedings of the Raspberry Pi Foundation Research Seminar series
[8] Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of personality and social psychology, 102*(4), 700.
[9] Cassidy, R., Cattan, S., Crawford, C., & Dytham, S. (2018) How can we increase girls' uptake of maths and physics A-level? *The Institute for Fiscal Studies, R149.*
[10] Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117,* 497–529.

designed by BIT and aimed to increase the enthusiasm of girl pupils for the subject using two fun stand-alone lessons on AI and an action guide with tips for teachers on how to engage girl pupils on computing. It also aimed to encourage conversations between pupils and parents using prompts that were aligned with the lessons to increase girls' perception of the acceptance of computing at home. Both interventions were delivered between September 2021 and February 2022. Figure 1 illustrates the logic model for both interventions.

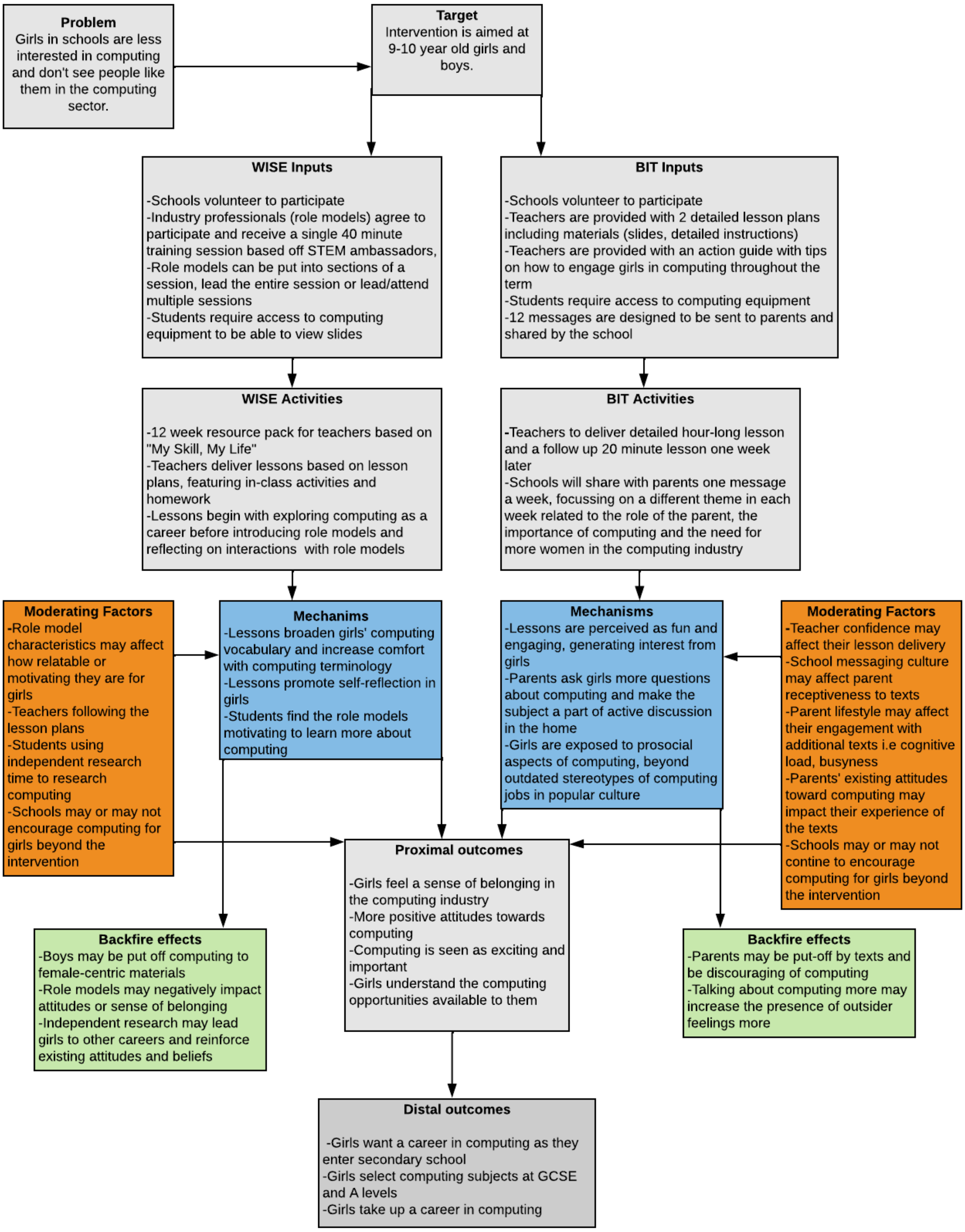

Figure 1: Logic model for the My Skills My Life and the Code Stars interventions

1.2.1 WISE *My Skills My Life* intervention

The intervention organised by WISE was an adaptation of their programme named *My Skills My Life*[11]. This programme encourages 11-19 year-old girls to find out about their preferred personality types using a quiz and matches them with real-life role models who have rewarding and successful careers in STEM. *My Skills My Life* is based in turn on the *STEM Ambassador* programme[12], which connects volunteering STEM professionals with secondary schools to promote STEM careers between young people.

WISE adapted the *My Skills My Life* programme for the GBIC project to focus more specifically on computing careers, rather than STEM in general, and tailored its content to primary school pupils. The resources were initially designed for pupils aged 11 and above; WISE adapted the quiz slightly to make its language more appropriate for 9-10 year olds and developed a Computing version of the resources featuring only role models working in the computing sector. The original programme *My Skills My Life* runs as a standalone session, whereas WISE developed additional lessons for this project.

The resulting intervention was a 12-week programme targeted at Year 5 pupils that included a) ten lessons on the importance of STEM and computing, role models in computing and finding a dream job, b) a quiz to be completed by pupils in one of the sessions on their strengths and personality types, in order to match them with job profiles tailored to their preferences, and c) the visit of a successful female role model working in a computing field during one of the lessons.

The logic model in Figure 1 describes the activities, mechanisms of change and outcomes targeted by the intervention. This intervention aimed to increase girls' positive attitudes towards computing through an increase in their sense of belonging in computing. One hypothesised mechanism for increasing belonging was exposure to female role models and increased motivation to learn more about computing careers. A second mechanism was that the lessons and the quiz activity would promote self-reflection in girls about their skills and preferences, and thereby guide them to identify fitting careers in computing and STEM. A third was that the lessons would broaden girls' computing vocabulary and increase their comfort with computing terminology.

Teacher-led lessons

Teachers were to deliver ten lessons each lasting 45-60 minutes during normal computing lessons, with homework to be completed between them. Two additional lessons were set aside at the beginning and end of the programme for pupils to complete a survey on attitudes towards computing. WISE provided the teachers with full lesson plans for each of the sessions, which included some homework suggestions (it was left to the teacher as to whether to do them or not). Table 1 provides an overview of the lessons given as part of the programme.

[11] https://www.wisecampaign.org.uk/what-we-do/expertise/welcome-to-my-skills-my-life/
[12] https://www.stem.org.uk/stem-ambassadors

Table 1: Summary of lessons

Lesson title	Objective
1. My Dream Job	To introduce a variety of STEM careers
2. A Day in the Life of You	Understanding how computing impacts all of our lives
3. Impact of Computing in the world around us	Being able to understand the working world and how computing is used in the wider world
4. Questions for Role Models	To collate a list of questions to ask the role models attending two sessions later
5. My Skills My Life quiz: Do they match me?	To discover what your own preferred way of working is, and then to explore job profiles of people who match your strengths
6. Real Life Role Models	To meet people like them who are happy and successful working in computing - **attended by a role model**
7-8. Create a presentation (two lessons)	To create a presentation for the class on a role model of your choice
9-10. Presentations (two lessons)	To showcase the research and learning over the lessons

The fifth lesson included the *My Skills My Life* quiz for the pupils on their preferred way of working, in order to match them with job profiles of role models with similar strengths according to their results. The last four lessons were reserved for pupils to create and deliver a presentation on a role model of their choice.

Session with a female role models

In the sixth lesson, a role model working in an industry related to computing spoke to pupils. The role models selected to deliver the 'Real Life Role Models' session were all female, and received training prior to meeting the pupils. They were matched locally with schools where possible, or based on their availability otherwise. The number of role models visiting schools varied with the size of the class, with the target of one role model to five pupils. In terms of recruitment, the role models all volunteered their time for the programme, and the criteria for volunteers was that they had to use computing in their day to day role.

The training consisted of one-hour webinar workshops provided by WISE to role models. They covered information about the project, the *My Skills My Life* resources, how to engage with primary school children and how to talk about their roles using words and in a way that they could understand and relate to. Not all role models were able to attend these sessions, thus WISE shared a recording.

During their lesson, WISE asked role models to speak about their personality types from the *My Skills My Life* quiz and how it fit with their jobs, what they did in their day to day role, how they got there, who inspired them and what they liked to do outside of work. WISE guidance was broad and further specifics were left to the role models' judgement. WISE provided them with further ideas, support and guidance where required.

1.2.2 BIT *Code Stars* intervention

The intervention implemented by BIT was a 12-week programme that involved: A computing-themed, one-off lesson for Year 5 pupils on artificial intelligence (AI) delivered by teachers at any time before the eighth week of the intervention; a homework task for pupils to complete with their parents or carers; and a follow-up lesson to be completed at some point later in the term. Teachers received an action guide with evidence-based tips they could use during their teaching to make sure girl pupils were engaged in the lessons (e.g. how to praise and encourage them). Throughout the programme, schools were asked to share weekly prompts with pupils' parents, which were designed to encourage conversation between pupils and their parents about computing in general and the AI lesson. Teachers also had to set time aside for pupils to complete baseline and endline surveys at the beginning and end of the programme, and continue teaching curricular computing lessons while they sent the messages to parents throughout the twelve weeks.

The logic model in Figure 1 describes the activities, mechanisms of change and outcomes targeted by the intervention. This intervention aimed to increase girls' positive attitudes towards computing through an increase in their sense of belonging in computing. One hypothesised mechanism was an increase in girls' interest and enjoyment of computing through two fun and engaging lessons and the encouragement tips from the action guide. A second mechanism was promoting active discussions on subjects related to computing at home between pupils and their parents. Another was exposure to prosocial aspects of computing, beyond stereotypes of 'geeky dark rooms', using both the lessons and messages shared with parents.

AI-themed *Code Stars* lessons and action guide

Teachers were asked to deliver a one-hour lesson to introduce pupils to the basic concepts behind AI and a short 30-minute follow-up lesson to revisit the concepts learnt in the first one. At the end of the second lesson, pupils received a certificate for their participation.

These computing resources were designed to assume no prior knowledge on AI on the part of pupils or teachers. The overall theme for the lessons was how AI can be used to clean up the ocean. In the first one-hour lesson, pupils were introduced to the theme and had an opportunity to use coding to train a robot to (virtually) clean the ocean and to distinguish between fish and floating rubbish. This lesson was intended to occur within the first six weeks of the intervention period. After the game pupils had to engage in discussion about AI and real life examples of it. After the lesson, pupils were set to complete a short home-learning task with their parents/carers before the follow-up.

Teachers were provided with the following resources via email to deliver the lessons:

- Detailed teacher instructions
- Power-point slides as a lesson plan
- Home-learning task instructions
- Certificate for the children
- Action guide with tips on inclusive teaching

Teachers were told to refer to the action guide while teaching the two *Code Stars* lessons and the usual planned computing lessons throughout the duration of the programme. The guide gave evidence-based tips to make the teaching more inclusive and help engage girl pupils during computing lessons. It gave ideas for teaching activities (e.g., using the Scratch coding platform, watch suggested videos, try pair programming, decorate the classroom with posters of real-life female role models in STEM) and quick teaching tips on how to communicate with pupils (e.g., how to praise and encourage them, how to address the subject of gender imbalance, how to talk about failure).

Conversation prompts for parents

Alongside the lessons, BIT developed nine weekly conversation prompts for parents/carers of pupils who were taking part in the *Code Stars* project to encourage them to speak about computing with their children.

The initial plan was for BIT to send parents the conversation prompts via text message every week for the twelve-week duration of the programme. Following feedback from schools, that option was deemed infeasible, as school resources were being over-stressed due to the COVID-19 pandemic and could not deal with the administrative burden needed to comply with data privacy protocols. In response to this challenge, BIT sent the messages to schools so that they could share the conversation prompts instead, either directly to parents or as a homework 'discussion task' for pupils.

Teachers could choose to share the prompts with parents via the following four options:

a) Text messages to parents/carers
b) Emails to parents/carers
c) Digital version of prompt as an online homework task for pupils (e.g., to put on a class website)
d) Paper version of prompt as physical homework task for pupils

The conversation prompts were designed and suggested to be sent out weekly during the 12-week duration of the programme, but schools and teachers were invited to send them out in the way they felt most appropriate, given the COVID-19 context at the time of implementation. BIT also included a suggested weekly schedule with specific dates alongside the messages and four different message versions depending on the format used to send them among the four options above. Table 2 below shows a summary of the message prompts shared with the parents.

Table 2: Summary of message prompts

Message content	Objective and content	Suggested sending week
1. Introduction + ‘Hidden Figures’ film	This message emphasised that the parental voice matters and that any conversations with children can make a difference. It also suggested to watch the film Hidden Figures with their child, an inspirational film about three female mathematicians in NASA's team of 'human computers'.	4th October (Week 5 of term)
2. Keeping safe online	This message included tips on e-safety for parents. This tried to build up goodwill between parents and the messaging service, as parents had provided feedback that e-safety (making sure their children are safe online) is important for them to learn about.	11th October (Week 6 of term)
3. Fun activity: ‘Build your own game’	The message included a tutorial and invited parents to create with their children a jumping game using the Scratch coding platform. It included a quote from Barack Obama.	18th October (Week 7 of term)
4. Sharing your learning	The message invited parents to talk with their children about what they have been learning in computing class and talk about two things they have learnt recently.	1st November (Week 8 of term)
5. What could you build?	The message highlighted a female role model and prompted parents to engage in conversation with their child on the importance of computing and how it can be used to build many things. It included a video from Karlie Kloss, a famous supermodel and part-time coder.	8th November (Week 9 of term)
6. Problem solving	Prompt to engage parents in conversation about a time their child didn't know how to do something on a computer, phone, or tablet but they figured it out.	15th November (Week 10 of term)
7. Where could it take you?	The message contained information on interdisciplinary connections between computing and other subjects like music, fashion, etc.	22nd November (Week 11 of term)
8. Using programming creatively	Link to the *Code Stars* lesson (using AI to clean oceans). The message asked parents to ask their children to show them the certificate from the lesson.	29th November (Week 12 of term)
9. Thank you + Activities to learn more	Closing and thank you message to parents. It invited them to encourage their children to join coding clubs and play educational games, and included a link to a game.	6th December (Week 13 of term)

1.3 GBIC partners

This project joins the National Centre for Computing Education, run by a consortium comprised of STEM Learning, the British Computer Society (BCS), and the RPF, with BIT and WISE, combining the extensive experience of organisations who have computing at the core of their mission with expertise in designing and evaluating interventions. The funding body for this programme as a whole is the Department for Education (DfE), and BIT fulfils the role of an independent and external evaluator for both interventions. The design and delivery of the *Code Stars* intervention and this evaluation report were carried out by different BIT teams to ensure the evaluation results were as objective as possible.

2. Methods

The evaluation used a mixed-methods approach. Quantitative data was collected via surveys completed by pupils in computing lessons pre- and post-intervention in both treatment and control schools as part of the impact evaluation. An implementation and process evaluation (IPE), which aimed to explore the mechanisms of change and to complement the quantitative survey findings, was also carried out. This section describes the methods used as well as the limitations of our approach.

2.1 Impact evaluation

The impact evaluation was designed as a randomised controlled trial (RCT), with three arms (one control, two treatment), and was randomised at the school level with outcomes at the pupil level.

2.1.1 Research questions and outcome measures

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

1. *Primary outcome*: Sense of belonging within computing, as measured by the belonging subscale of the SCSAS.
2. *Secondary outcome*: Attitudes towards computing as measured by the SCSAS.
3. *Secondary outcome*: Stated intention to study computing in the future.
4. *Exploratory outcome*: Stated intention to study science, technology or mathematics in the future.

The score for the Belonging subscale of the SCSAS was selected as the primary outcome as it more directly measures the outcome targeted by the two interventions, and reflects the hypothesised pathways to impact on girl pupils' choice to study computer science in Year 10 onwards, than the overall SCSAS score. Stated intention to study computing in the future was also selected as a secondary rather than the primary outcome, based on the assumption that pupils' reported prediction of whether they will study computing in multiple years is less reliable than attitudes measured through the SCSAS. The survey measures are further discussed in section 2.1.5.

Table 3 outlines the data that was collected to measure the outcomes listed above, and when it was collected.

Table 3: Impact evaluation outcome measures

Outcome measures	Data to be collected	Point of collection
Primary[13]: Sense of Belonging within computing	Score on the belonging subscale of the SCSAS (score 1-4)	All data were collected at the start (baseline) and end (endline) of the 12-week evaluation period[14]
Secondary: General attitudes towards computing	Overall score on SCSAS (five subscales weighted equally, score range: 1-4)[15]	
Secondary: Stated intention to continue studying computing	Survey of whether the pupil plans to continue studying computing with possible responses including "*Yes*", "*No*", or "*I don't know".* This was binarised into: ● 1 = "*Yes*" ● 0 = "*No*" or "*I don't know*"	
Exploratory outcome: Stated intention to continue studying Maths, Science or Technology	Survey of whether the pupil plans to continue studying Science, Technology or Maths with possible responses for each subject including "*Yes*", "*No*", or "*I don't know".* This was binarised into: ● 1 = "*Yes*" for at least one three subjects ● 0 = "*No*" or "*I don't know*" for all three subjects.	

The Student Computer Science Attitude Survey (SCSAS) was developed to measure attitudes towards computing (Haynie and Packman, 2017); the full survey and a summary of adaptations that took place between baseline and endline data collection are included in appendix 4. It contains 25 questions and has 5 subcategories (5 questions per subcategory): Confidence, interest, belonging, usefulness and encouragement. Within each subcategory, the 5 items are scored on a four-point Likert scale from 1 (strongly disagree) to 4 (strongly agree), and averaged to create subscores. Thus, each 5-item subscore has a potential range of 1-4. These subscores are averaged for a total score that has a potential range of 1-4, with 4 representing a very positive attitude towards computing.

[13] We used the belonging sub-scale as our primary outcome measure as we felt it was a more precise measure of the increase in the sense of belonging this intervention aimed to promote.
[14] Not all schools completed the intervention in the 12-week period allotted, as some schools completed the endline survey more than 12 weeks after completing the baseline survey (see sections 5 and 6).
[15] Adapted from Haynie and Packman (2017). Available at: https://csedresearch.org/tool/?id=156. See appendix 4 for details of adaptation.

2.1.2 Sampling and randomisation

The evaluation was designed as a three-armed clustered randomised trial that aimed to test the impact of both the *My Skills My Life* intervention and the *Code Stars* intervention, compared to usual school computing education (control). Both were designed to increase girls' sense of belonging within computing, and were randomised at the school level with outcomes at the pupil level. Randomisation was stratified on schools' Ofsted scores and the percentage of pupils with free school meal (FSM) status (above or below the median).

Pupils were blind to allocation during the programme and during outcome data collection, and therefore did not know that pupils at other schools received different lessons. Teachers were not blind to allocation, as they were responsible for delivering the materials, and as they had registered interest in the trial, would have been aware that there were control and treatment groups. We used school unique reference numbers (URNs) as unique identifiers. BIT conducted the randomisation.

Recruitment of schools was conducted by RPF. All primary schools in England were eligible for this trial, with the exception of schools participating in other GBIC trials where the potential overlap was deemed unacceptable[16]. Only the schools that could offer the full 12-week programme to a single cohort of pupils were able to enter the trial. Schools were also required to have female pupils (all-boys schools were excluded). Data was collected from both boys and girls, but only data from girls were analysed for primary and secondary analyses[17].

All schools that entered the sample did so voluntarily, which has implications for the external validity of the findings, as those schools that volunteer are likely to be more enthusiastic than the average school, and this may interact with the treatment effect to compound any effects. These schools may also be more likely to have above-average facilities for teaching computing, such as specialised classrooms or dedicated computing teachers, as schools without adequate facilities were unlikely to have signed up for a programme they couldn't implement due to a lack of facilities. As an example, if schools didn't have enough computers or tablets for the whole class to be taking the baseline and endline surveys at the same time, they may have been less likely to sign up. If these programs were only likely to be rolled out to schools who register interest, this difference between our sample and schools in general may not be a problem. However, these findings should be taken with caution if considering rolling out the program to all schools in England.

---

[16] Some of the schools in this strand of the trial also participated in the i5a Code Clubs strand of the GBIC trials. This was deemed 'acceptable overlap' in discussions between RPF and the BIT evaluation team, as the Code Clubs took place as an extracurricular activity, and contained a small number of pupils from each year group (the average size of the Code Clubs was 10-15 pupils). This was done to ensure we were able to recruit enough schools to be adequately powered across the different trials.

[17] We collected data from boys as well as girls for the following reasons: a) to ensure boys weren't being treated differently than girls in schools, b) it is possible boys may have also benefited from the programmes, and c) although not the primary objective of the trial, collecting data on boys was a low cost way to look at any potential differences between boys and girls in our outcomes. Boys data was looked at in an exploratory sense, and full tables detailing boys scores are in Appendix 2.

2.1.3 Attrition and final sample

Figure 2 describes school level attrition at the different stages between recruitment and the completion of the endline survey in each trial arm. High attrition rates were observed between randomisation and completion of baseline surveys, with proportionally more dropping out from the treatment groups than the control. At least part of this attrition is likely due to disruption of the coronavirus pandemic. In late summer and autumn 2021 and again in early January 2022, a wave of infections likely led to many teachers and pupils self-isolating at home. However, the differential attrition between experimental groups suggests a possible risk of bias in the analysis results. This risk is explored later in this report through balance checks on the baseline data received from these schools post-attrition.

To mitigate attrition to the extent possible, schools that had not yet completed baseline or endline surveys were contacted regularly with email reminders from the BIT and WISE implementation teams. The BIT implementation team were provided with daily to weekly updates on which schools had completed the surveys by the BIT evaluation team, and WISE were provided with weekly updates. Additionally, endline surveys were kept open until 31 January 2022, a full month after the initial deadline for completions.

Attrition was also observed between baseline and endline surveys, in terms of both schools failing to complete the endline and pupils within schools not completing the endline. The WISE arm of the trial observed greater attrition at this stage, which may be explained by the fact that the WISE endline surveys were completed slightly later than the BIT or control group surveys because of slightly varying intervention timelines. Due to this time discrepancy, the WISE surveys were completed when COVID cases were higher, which may have resulted in more pupils and teachers being out of school due to positive cases or self-isolation. This could also be explained by the fact that the BIT implementation team managed the baseline and endline survey process for both the control and BIT groups, including following up regularly with schools that had yet to complete the survey, and may have had more resources available for these follow-ups than were available for the WISE group.

Once all survey data was collected, data cleaning was conducted to remove any data points deemed potentially unreliable. All data was dropped for pupils who had answered in a straight pattern (e.g, a survey with the answer ‘Strongly disagree’ for every question of the SCSAS). This applied to 114 pupils from the total baseline sample (boys and girls) and 53 pupils from the endline sample (boys and girls). In cases where there were duplicate observations (the same pupil entering the survey twice), we kept only the first complete survey from the pupil. For example, if a pupil partially completed a survey first, and then fully completed another one at a later date, the first complete survey is the one we used. If a pupil never fully completed the survey, we retained their first partially complete entry.

The final data set consisted of (1) data from girls who had completed the endline survey matched to their baseline data and (2) data from girls who had completed only the endline survey. We used a multistep matching process to match as many baseline and endline surveys as possible. Responses were matched on a combination of school name, full name and date of birth; we also used survey completion dates when manually reviewing possible matches. To improve the number of matches we used both exact matching and ‘fuzzy matching[18]’ to account for the use of different name spellings or orders and data entry errors. We use an iterative process which involved loosening the matching criteria on different matching variables to identify possible matches at each stage, and manual review to confirm these possible matches. The final analytical sample for our primary sense of belonging measure consists of 2,599 girls[19].

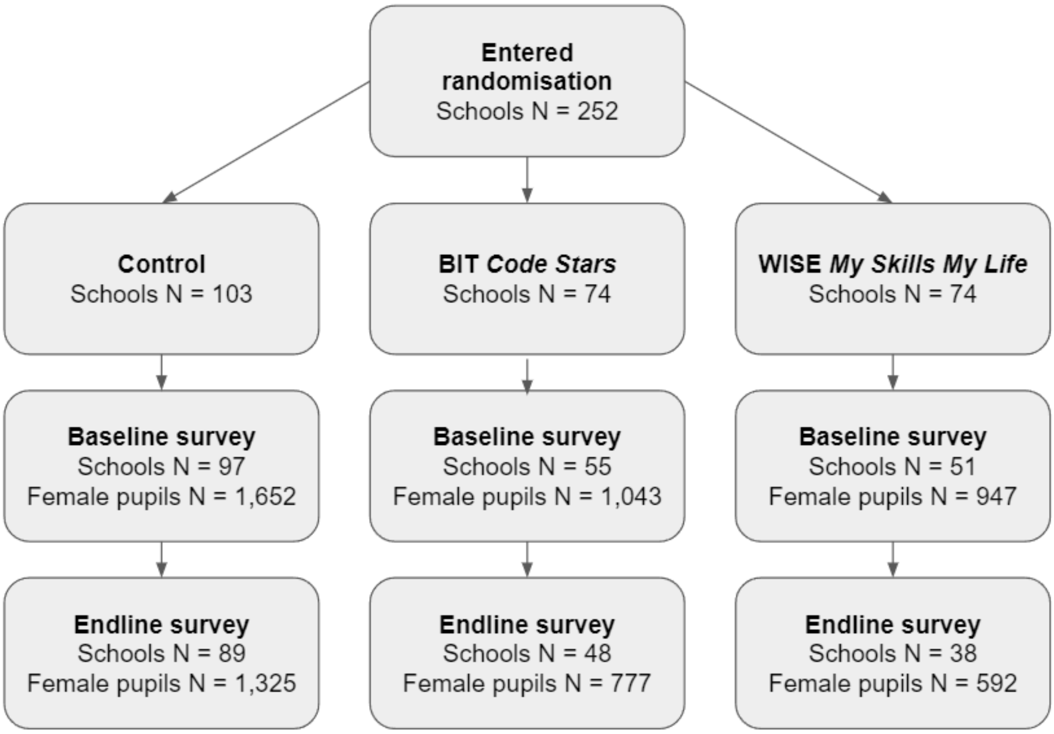

Figure 2: School level attrition

Baseline differences in outcome measures

Table 4 shows that at the point of baseline data collection, the groups were balanced in terms of general attitudes towards computing, intention to study computing in the future, or intention to study science, maths or technology. However, the schools in the WISE intervention were significantly lower in terms of feelings of belonging in computing. We believe this is likely a result of the attrition between the randomisation and baseline survey.

It was also the case that the baseline data collection was extended so that some schools may have completed the baseline survey after the intervention had begun, which may have moved their baseline outcomes from their ‘true’ baseline pre-intervention.

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

[19] Our final sample varies between outcome measures, as we chose to keep pupils who only partially completed the survey but whose data we could use in our models, even if the surveys weren’t complete. For example, some pupils answered the intention to study computing questions, but not the full SCSAS survey.

Table 4. Balance checks for all baseline data

Covariate	Percentage (or mean) per arm			Balanced?[20]
	Control	BIT	WISE	
Gender				
Boys	47%	47%	49%	Yes
Girls	51%	52%	49%	
Non-binary/Other	2%	1%	1%	
Baseline scores (Girls only)				
Belonging score	3.02	3.02	2.95**	No
SCSAS score	2.89	2.89	2.86	Yes
Intention to study computing	56%	58%	55%	Yes
Intention to study science, maths or technology	77%	78%	76%	Yes
** p < 0.01, * p < 0.05, + p < 0.1, denotes statistically significant differences from the control arm				

Baseline differences in outcome measures for pupils that completed endline surveys

Table 5 (below) shows that at the point of baseline data collection, the groups were balanced in terms of general attitudes towards computing, intention to study computing in the future, or intention to study science, maths or technology. However, the schools in the WISE intervention were significantly lower in terms of feelings of belonging in computing. It is worth no

[20] Balance checks were conducted by regressing the dependent variable (e.g. gender) on our treatment variable. The regressions were either linear or logistic, dependent on the variable.

Table 5. Balance checks for baseline data of pupils that completed the endline survey

Covariate	Percentage (or mean) per arm			Balanced?
	Control	BIT	WISE	
Gender				
Boys	45%	48%	48%	No
Girls	53%	50%	51%	
Non-binary/Other	2%	2%	1%	
Baseline scores (Girls only)				
Belonging Score	3.02	3.00	2.94**	No
SCSAS score	2.88	2.87	2.86	Yes
Intention to study computing	55%	55%	55%	Yes
Intention to study science maths or technology	77%	78%	76%	Yes
** $p < 0.01$, * $p < 0.05$, + $p < 0.1$, denotes statistically significant differences from the control arm				

Implications for final analysis

The high attrition rates and baseline imbalances described above carry implications for our analysis. Firstly, the differential attrition across groups, with proportionally more control schools completing the trial, has the potential to bias our estimates of the impact of this intervention. The direction of this bias will depend on how the schools that dropped out of the control group differed from those that remained:

1. If it were the case that those that dropped out were less enthusiastic about teaching computing at baseline (and thus likely to score lower on attitudes and intentions surrounding studying computing on a pupil-level), this might bias our estimates upwards, as a lower proportion of the control group would be more enthusiastic about computing relative to the treatment groups.
2. If it were schools that were higher in factors positively correlated with our outcome measures that dropped out in the treatment groups (e.g., it is possible they struggled to find time and resources to complete the programmes), it would bias our estimate downwards.

Though pupils in the WISE group had a lower belonging score than the control group and the difference was statistically significant, the fact that there was no statistically significant difference on the other baseline indicators in table 5 suggests that this imbalance may have been due to chance, and makes it difficult to make any conclusions about the direction of a possible bias in the impact estimated resulting from this imbalance.

2.1.4 Analytical strategy

The models analysing the belonging and total SCSAS scores used a linear OLS regression, while the intention to study computing and intention to study science, maths, or technology used logistic regression (full model specifications are outlined in appendix 2). All analyses were Intention to Treat (ITT) meaning that outcomes were analysed on the basis of the groups that teachers and pupils were randomly allocated to, regardless of their compliance with the intervention. The covariates (baseline SCSAS score, school Ofsted rating, school proportion of pupils with free school meal eligibility) were chosen as they could potentially influence the outcomes, thus controlling for these variables could increase the precision of the impact estimates.

All planned covariates were checked for missing data pre-analysis. For some schools in the sample, we were unable to obtain an Ofsted rating due to there not being one publicly available. For these schools, we elected to assign them to an extra value of the categorical variable of Ofsted rating[21].

Given that the endline data would likely include some pupils who did not complete the baseline dataset, we specified pre-trial decision rules for dealing with missing data as baseline scores on the SCSAS were to be used as a covariate in the analysis. In the final sample, approximately 20% of pupils were missing baseline SCSAS scores (above the threshold of 5% for listwise deletion), and multiple imputation[22] was performed, whereby predicted values were substituted where data was missing.

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

1. Multiple imputation model
2. Missingness indicator model[23]
3. Complete case analysis

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

---

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

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

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

2.1.5 Limitations

Attrition

Differential attrition across experimental groups can lead to bias in treatment effect estimation. While baseline imbalance in outcome measures between groups can be partially addressed through using baseline SCSAS as a covariate in the analysis, we cannot be confident that there are not unobserved variables driving that baseline difference that we do not control for in analysis. The consequence of this would be that these unobserved variables potentially interact with attitudes and intentions, and could bias the results. The implications for the analysis of the differential attrition observed are described in section 2.1.3.

While differential attrition was observed, we were still powered to detect our target effect size. 252 schools were recruited with a 40% attrition buffer to detect an estimated Cohen's *d* of 0.16 (or a 0.08 point increase) along the 1-4 belonging subscale, and a Cohen's *d* of 0.20 (or a 0.10 point increase) for the full SCSAS 1-4 scale. This meant approximately 64 control schools and 44 schools per treatment group were required to detect our desired effect size, meaning the final analytical sample was sufficient to detect the originally targeted effect size.

Pupil survey outcome measures

Given the nature and objectives of the intervention, defining and measuring outcome indicators were challenges inherent to the evaluation. The intervention aims to reduce gender gaps in school subject choices from Year 10 onwards by intervening in earlier years, in Year 5. While this early intervention approach may offer important benefits in terms of reducing barriers that may arise or increase in later years of education, it also creates a need to rely on short-term 'proxy' indicators that can be measured within the evaluation period (in this case directly after the intervention completion), yet could predict school subject choices in Year 10. This is particularly challenging as some of the barriers to girls choosing computing as a subject in later years (and that the intervention aims to prevent) may arise after Year 5 and before Year 10; this would imply a risk that the effect of these barriers are not captured in the data collected while the pupils are in Year 5.

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

To address this dual challenge, the evaluation approach focused on attitudes towards computing as the primary outcome, and hypothesised that these could be measured and predict future subject choice. The survey tool used, the SCSAS, was cognitively tested to increase its reliability in measuring these attitudes with a small group of Key Stage 2 pupils from schools outside of this intervention. While these efforts should help, they are unlikely to fully address these challenges, and some issues were raised in testing in terms of different children interpreting some of the questions in different ways. The IPE also identified that completing the pupil baseline and endline SCSAS surveys could take an excessive amount of time according to some teachers. The possible implications of these measurement challenges for the results are discussed in section 7.1.

2.2 Implementation and process evaluation

Alongside the impact evaluation, an IPE was conducted to examine the mechanisms of change and the diversity of implementation and programme delivery for the *My Skills My Life* intervention and the *Code Stars* intervention.

2.2.1 Research questions

The IPE aimed to address the following research questions:

1. What are the programs related to computing education (in addition to the interventions delivered) that are available to pupils in the schools?
2. Which components of the interventions were delivered and in what ways (both interventions)? *[both interventions]*
3. How did the parents engage with their child after receiving the text message prompts? *[Code Stars intervention]*
4. What were teachers' views about and experiences with delivering the intervention lessons *[both interventions]*, as well as their views and experiences of having the role models in their class? *[My Skills My Life intervention]*
5. What were the girls' views of the interventions and in what ways did they engage with it? *[both interventions]*
6. What were the ways in which the intervention impacts girls' attitudes towards computing? *[both interventions]*

2.2.2 IPE research design

A mixed-methods design was used to address the research questions above; this included case studies of the intervention schools and an online teacher survey for the *Code Stars* intervention.

Each case study included an interview with a teacher who delivered the programme, a class observation and focus group with pupils. In the case of the *My Skills My Life* intervention, the role model who attended the school was also interviewed. For the *Code Stars* intervention, two parents were also interviewed in one of the case study schools.

The case studies helped to capture experiences at a few schools in depth. In the case of the *Code Stars* intervention, they were complemented with an online survey for teachers at the end of the programme in order to capture the breadth of experiences of classroom teachers, including those who might not have been able to implement one or more intervention components. The approach to sampling, participant recruitment and data collection for the case studies and *Code Stars* teacher survey is described below.

2.2.3 Sampling

Data for the IPE was collected from four case study schools (two for each intervention) and a teacher survey for the *Code Stars* intervention. A purposive approach to sampling was used to capture a range and diversity of views and experiences by taking into account multiple key characteristics of schools and participants (teachers, parents, role models, pupils). There were two stages of sampling: (i) sampling of schools and (ii) sampling of pupils, teachers, parents and role models for each case study. The following characteristics were considered when sampling in each stage:

1. **Schools:** The sample included diversity in Ofsted rating and proportion of pupils eligible for free school meals (FSM) in line with the stratification of the schools for the impact evaluation, as well as geographical location. The proportion of FSM eligible pupils at each school was retrieved from the DfE national information about schools[24].

2. **Participants:**
    - *Teachers:* The teacher delivering the lesson in the case study school.
    - *Role models (for My Skills My Life intervention)*: The role model delivering the relevant lesson in the case study school.
    - *Parents (for Code Stars intervention)*: A parent of a pupil who received the prompts to be shared by schools as part of the intervention.
    - *Pupils*: Those that participated in the *Code Stars* intervention's one-hour lesson, or participated in lesson 6 of the *My Skills My Life* intervention with the role model, given that the focus groups took place after this lesson.

Due to the impact of COVID-19 on teacher workload and visiting policies, school engagement with the intervention and teacher availability had to be taken into account alongside school characteristics for the sampling. BIT had to rely on the teachers to recruit parents for the interviews, as the teachers had the information on what parents had opted in to receive the conversation prompts and whether they had received them and could therefore comment on them. Due to issues of low parental engagement throughout the intervention, the teachers in the two sampled schools struggled to identify parents that fulfilled the criteria and were willing to participate in the interview. As a result, BIT was unable to interview parents in one of the schools and had to interview two parents from the other school that agreed to take part in the IPE. Both of them reported not having received the message prompts, so the IPE results about the conversation prompts are limited as a result. Table 6 details the achieved sample of case study schools and the data collection activities carried out in each of them.

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

Table 6: Achieved case study sample

School	School profile	Sampled individuals	Data collection activities
S01	-Located in the North East -Above average FSM - Ofsted rating: Good	- Classroom teacher (with curriculum responsibilities for computing) (*Teacher 1*) - 6 Year 5 (9-10 years old) pupils - Role model (*role model 1)*	- Teacher interview (online) - Lesson observation (in person) - Pupil focus group discussion (in person) - Role model interview (online)
S02	-Located in the Midlands -Below average FSM -Ofsted rating: Good	-Classroom teacher (with curriculum responsibilities for computing) (*Teacher 2*) - 6 Year 5 (9-10 years old) pupils - Role model (*role model 2)*	- Teacher interview - Lesson observation - Pupil focus group discussion - Role model interview - All activities were online
S03	-Located in the East of England -Below average FSM -Ofsted rating: Good	-Classroom teacher (with curriculum responsibilities for computing) (*Teacher 3*) - 5 Year 5 (9-10 years old) pupils	-Teacher interview (online) -Lesson observation (in person)
S04	-Located in the North West -Above average FSM -Ofsted rating: Outstanding	-Computing specialist teacher (*Teacher 4)* - 5 Year 5 (9-10 years old) pupils - Parent of a boy pupil (*Parent 1)* - Parent of a girl pupil *(Parent 2)*	-Teacher interview (online) -Lesson observation (in person) -Parent interviews (online)

2.2.4 Data collection methods

BIT researchers conducted the following data collection activities:

- *My Skills My Life* intervention*:*
  - 2 teacher interviews
  - 2 lesson observations
  - 2 focus groups with 5 pupils
  - 2 interviews with a role model (who delivered lesson 6 and 7 to the pupils)

- *Code Stars* intervention*:*
  - 2 teacher interviews
  - 2 lesson observations
  - 2 focus groups with 5 pupils
  - 2 interviews with a parent
  - Online teacher survey

We describe each type of data collection activity in turn below.

Interviews

In each case study, semi-structured 30-45 minute interviews were conducted with two teachers, two role models (*My Skills My Life*) and two parents (*Code Stars*). These interviews were used to answer the research questions pertaining to the delivery and factors influencing delivery of the intervention pupil engagement, and mechanisms of impact.

The interview topic guides were developed by a qualitative researcher with feedback from the RPF team; they were informed by the theory of change and the intervention materials shared by the implementation teams. These interviews were conducted using an online platform (e.g, Google Hangouts) or over the phone, depending on the preference of the participants. They were audio recorded and transcribed to aid with analysis. The interview guides can be found in appendices 6 to 8.

Lesson observations

A BIT researcher observed a lesson in each case study school to better understand how the interventions were being delivered and assess pupil engagement, lesson fidelity and facilitators and barriers to lesson delivery. The researcher recorded field notes following a structured proforma (see appendix 4 for the full observation pro-forma).
For the *My Skills My Life* intervention, a BIT researcher observed the fifth lesson ("*My Skills My Life* quiz: Do they match me?") in one case study school and the sixth lesson, the one attended by the role model, in the other case study. For *Code Stars*, we observed the 1h-long lesson on AI in both case studies. These lessons were perceived to be core components of the intervention as described by the implementation partners.

Pupil focus groups

25-30 minute focus groups were conducted with pupils after the lesson observations of the 1-hour long lesson delivered by the teacher for the *Code Stars* intervention, and after observation of the fifth lesson (led by a teacher) and the sixth lesson (led by a role model) for the *My Skills My Life* intervention. The aim of these focus groups was to explore pupils' views and experiences of taking part in the intervention, as well as to get a better understanding of the possible mechanisms of impact. These focus groups provided pupils with opportunities for discussion, and intended to support pupils to reflect on and share their thoughts on their experience.

BIT qualitative researchers developed pupil discussion sheets with feedback from the RPF team to guide the discussion. Due to changes in school attendance because of COVID-19, these focus groups had to be restricted to 5 pupils per group and were led by a BIT researcher. The analysis of the focus groups relied on the pupils' answers in the discussion sheets and notes taken by the BIT facilitator.

Online teacher survey

Following completion of the *Code Stars* intervention, the BIT evaluation team designed and sent out a short feedback survey to teachers at involved schools, in order to help us to:

- Understand whether teachers administered the lessons as intended, and whether there were any specific challenges teachers faced with the lessons.
- Understand whether, how, and how often the conversation prompts were shared with parents, and whether there were any specific challenges associated with this.
- Collect any general feedback from teachers about the programme.

The rationale behind designing and sending out this survey was that, in contrast with the *My Skills My Life* intervention, we had less information on how the intervention was implemented. The teacher survey was seen as a low-cost way to attempt to obtain information while placing a low burden on teachers. As we did not want to put any additional burden on teachers post-intervention, the survey was sent out alongside the end-of-project email, and we did not follow up with teachers. The survey received a total of 9 responses out of 55 schools in the sample. The teacher survey is included in appendix 5.

2.2.5 Analysis

Interview transcripts and field notes were managed using the Framework Approach[25]. This involved summarising transcripts and notes into a matrix organised by themes and sub-themes (columns) as well as by individual cases (rows). The managed data was then interpreted with the aim of identifying and categorising the range of phenomena present in each of the sampling groups. We conducted case and theme analysis to focus on providing rich descriptions of participant experience whilst looking for explanations, linkages and typologies within and across participant groups. The responses in the open questions of the surveys were also included in the qualitative analysis of interview transcripts and observation notes.

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

1. The case study approach means that findings should not be generalised across all participants, but rather understood as conveying some of the range and diversity of participant experiences.

2. Due to COVID-19, we were unable to fulfil our intended sample of schools and parents, instead adopting a convenience sampling approach, focusing on schools and parents that were able to take part in the research activities during the pandemic. This caused our IPE sample to be less diverse in terms of school characteristics and experiences and likely to show a limited range of implementation approaches.

3. Due to low parental engagement and time constraints from COVID-19 disruptions, the BIT team only had access to two parents from the same case study school that had given their consent to participate in the *Code Stars* intervention. Neither of them reported having received the message prompts and could therefore provide information about their effectiveness. Furthermore, one of them was a parent to a male pupil, which also limited the information we could gather about girls' experiences with the programme in general. Our findings on that feature of the *Code Stars* intervention are thus limited by the data we were able to collect.

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

3. Implementation context

3.1 Computing education and resources in schools

This section presents the findings for the first research question of the IPE and describes the programmes related to computing education that are available to pupils in schools, in addition to the interventions delivered. Interest in the subject and interaction with computing beyond the classroom was hypothesised as a mechanism of change during the intervention design, thus our evaluation looked into the curricular and extracurricular activities related to computing that were available to girls in the school context.

Starting with the curricular activities, even though all schools in the case studies followed the national curricula, there was a degree of variability across the schools that did the programme in teaching methods that depended on the resources and teaching schemes for computing chosen by the school, as well as on the engagement and computing skills of the teacher delivering it.

Three of the four visited schools had opted to rely on online third party platforms to teach the usual computing curriculum to the Year 5 pupils. The ones identified in our case studies were the Purple Mash computing scheme, Knowlsey scheme and NCCE Teach Computing programme, which are designed to be self-contained teaching programmes that follow the national computing curriculum. These schemes contained games, interactive lessons and included coding activities. The fourth school, by contrast, had their computing programme designed and taught by the PPA teacher[26], who was in charge of computing and music class at the school, and used an advanced, interactive and varied curriculum which also included content creation and design, coding, and general computer use, with a focus on real life applications.

Teacher interviews revealed a lot of variability in the computing knowledge and skills of the teachers delivering the computing lessons. The teachers who were less experienced with computing were relying on the self-contained schemes and did not feel confident in teaching the curriculum without having that guidance.

Regarding extracurricular computing activities, the options in the case study schools were coding, gaming or computing clubs. The only teacher who gave further information on the registered pupils in one of the clubs said that the coding club members were gender balanced. Apart from the after-school clubs and the curricular computing lessons, no other activities to encourage computing specifically for girls were mentioned by the teachers.

[26] PPA stands for "Planning, Preparation and Assessment". PPA teachers fill in for the main class teachers so those can have their PPA time. They are specialised teachers who may teach one or more different subjects to multiple year groups.

All schools observed in the case studies had sufficient electronic resources for pupils and could provide all of them with laptops or iPads for their computing activities. Depending on the school resources, pupils had to share their device with another pupil during the class.

3.2 Attitudes towards computing

Girls had quite positive attitudes towards computing at the beginning of the programme. Baseline scores on attitudes towards computing were quite high: 56% of female pupils in the baseline sample said they intended to study computing in the future. However, statistically significant differences between boys and girls in attitudes toward computing illustrated the potential for the intervention. The difference between boys and girls was largest in their intention to study computing in the future (71% vs 56%). Table 7 compares baseline scores for all outcome measures for boys and girls, also breaking down the SCSAS survey into its five subscales.

Table 7: Pupil baseline survey data by gender (outcome indicators emboldened)

Outcome	Values	Gender	N (non-missing)	Mean (SD)
SCSAS: Belonging subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,390	3.00 (0.52)
		Boys	3,171	3.08 (0.56)
Total SCSAS score	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,390	2.88 (0.48)
		Boys	3,171	3.00 (0.51)
SCSAS: Confidence subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,390	2.82 (0.61)
		Boys	3,171	2.99 (0.65)
SCSAS: Interest subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,390	2.95 (0.65)
		Boys	3,171	3.11 (0.69)
SCSAS: Usefulness subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,360	2.99 (0.56)
		Boys	3,138	3.09 (0.59)
SCSAS: Encouragement subscale	Mean score of likert scale questions (Strongly disagree - strongly agree) with a range of 1-4	Girls	3,360	2.65 (0.66)
		Boys	3,138	2.73 (0.71)
Intention to study computing in future	1 = “Yes” 0 = “No”, “Don’t know”	Girls	3,423	57% (0.50)
		Boys	3,200	70% (0.46)
Intention to study science, technology or maths in the future	1 = “Yes” 0 = “No”, “Don’t know”	Girls	3,423	77% (0.42)
		Boys	3,200	86% (0.34)

None of the four interviewed teachers reported having noticed any differences in engagement, interest or ability between boys and girls in their usual computing class, although one teacher noted that online gaming was more popular with the boys in their class, which they usually saw together with interest in computing in older pupils. Three of the teachers observed that differences in interest tended to appear amongst older pupils, but not in Year 5 pupils (the target group for the intervention).

Some of the interviewed teachers also observed that pupils with more positive attitudes and better skills at computing activities were also higher attainment pupils in their class or tended to be more skilled at maths, logic and problem-solving. Another teacher remarked that some of the girls that were more interested in computing or better at it, and more familiar with and positive towards computing careers, had models at home: Parents who were scientists, programmers or used computers in general. It should be noted however that the IPE was conducted with a small number of schools, and therefore these observations cannot be generalised across all participants.

4. Impact evaluation findings

4.1 Key findings

We did not find evidence of a statistically significant impact of either the *My Skills My Life* or the *Code Stars* intervention on the following outcomes, relative to the control group and by the end of the evaluation period (directly after the completion of the intervention):

1. Girls' sense of belonging in computing (SCSAS belonging subscale score).
2. Girls' attitudes towards computing (total SCSAS score).
3. Intention to study computing in the future.

Figure 3 shows the raw control mean and estimated treatment effect of the both interventions using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention. The 95% confidence interval of this treatment effect is also shown on the bar of the intervention group. It should however be noted that we are not making any comparisons between treatments, and that we are comparing each treatment arm individually to the control.

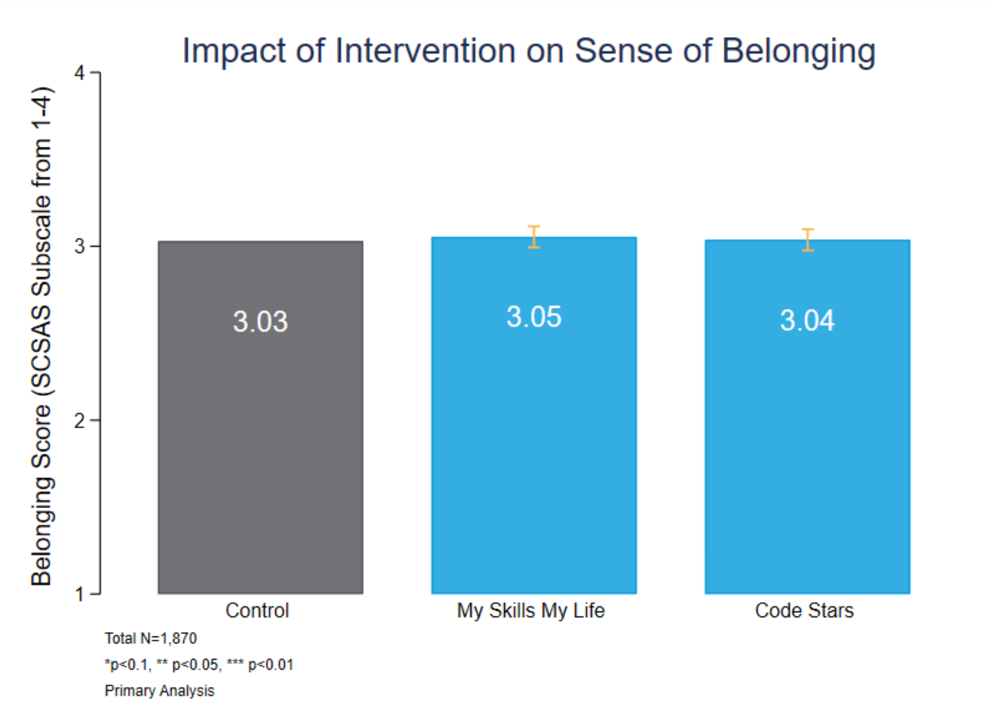

Figure 3: Model-adjusted sense of belonging scores, control arm, My Skills My Life treatment, and BIT Code Stars treatment (SCSAS scale ranges from 1-4)

We also found no evidence in our exploratory analysis that either intervention had an impact on girl pupils' intention to study maths, science or technology in the future.

Looking across all three arms (control, *My Skills My Life*, and *Code Stars*), one thing to note is that girls' sense of Belonging is generally quite high. The mean score on the belonging subscale was 3.02 (SD=0.53), suggesting that broadly speaking, girls do feel like they belong in computing at the KS2 level.

4.2 Findings for the *My Skills My Life* intervention

Table 8 provides the results of our main regression specification, using imputed values, across all outcomes for the *My Skills My Life* intervention. Appendix 3 contains full regressions with all three model specifications described in section 2.1.2 (multiple imputation model, missingness indicator model, and complete case analysis).

The control mean for the belonging subscale was 3.03 (SD=0.52) out of 4, which was higher than the average SCSAS score which was 2.89 (SD=0.47) out of 4, suggesting that belonging may not be one of the main barriers to girls' general attitudes towards computing. Baseline intention to study computing was also high at 50% (SD=0.50) for the control group, as was girls' intention to study science, maths or technology in the future, which was 76% (SD=0.43).

The difference in the primary belonging outcome after controlling for covariates was 0.024 (p=0.457) points on the belonging subscale, and as table 8 shows, none of the outcome measures were significantly impacted by the *My Skills My Life* intervention. This finding was consistent across all three model specifications.

Table 8: My Skills My Life impact evaluation results (all outcome measures)

Outcome	(1) SCSAS belonging subscale score (Primary)	(2) Total SCSAS score (Secondary)	(3) Intention to study computing (Secondary)	(4) Intention to study science, maths or technology (Exploratory)
Control group mean	3.03	2.89	49.81%	76.24%
Treatment group mean	3.00	2.85	49.55%	75.96%
Estimated treatment effect	0.024	-0.006	0.16pp	-0.14pp
N	2,599	2,599	2,600	2,614

While the treatment effect estimate found for the primary outcome in the *My Skills My Life* group relative to the control group is positive, it is small in magnitude and not statistically significant. Further, given the loss in precision and possible biases resulting from the observed attrition and baseline imbalances on the Belonging sub-scale survey score, we would advise against interpreting these differences as indicative evidence of a meaningful effect of the intervention.

Figure 4 shows the raw control mean and estimated treatment effect of the intervention using the pre-specified model, i.e. the estimated change that would be seen in the control group had those pupils received the intervention. The 95% confidence interval of this treatment effect is also shown on the bar of the intervention group.

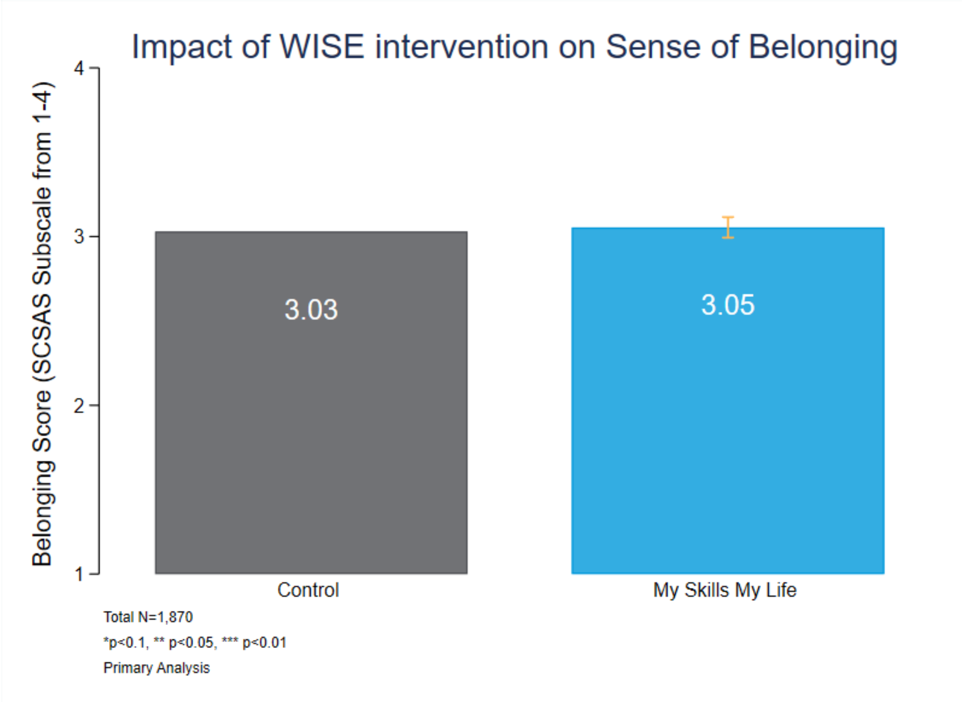

Figure 4: Model-adjusted sense of belonging scores, control arm vs My Skills My Life treatment (SCSAS scale ranges from 1-4)

4.3 Findings for the *Code Stars* intervention

Table 9 provides the results of our main regression specification, using imputed values, across all outcomes for the *Code Stars* intervention. Appendix 3 contains full regressions with all three model specifications described in section 2.1.2 (multiple imputation model, missingness indicator model, and complete case analysis).

The control mean for the belonging subscale was 3.03 (SD=0.52) out of 4, which was higher than the average SCSAS score which was 2.89 (SD=0.47) out of 4, suggesting that belonging may not be one of the main barriers to girls' general attitudes towards computing. Baseline intention to study computing was high at 50% (SD=0.50) for the control group, as was girls' intention to study science, maths or technology in the future, which was 76% (SD=0.43).

The difference in the primary belonging outcome after controlling for covariates for was 0.007 (p=0.819) points on the belonging subscale, and as table 9 shows, none of the outcome measures were significantly impacted by the *Code Stars* intervention. This finding was consistent across all three model specifications. For the secondary outcome of stated intention to study computing in the future, after controlling for our covariates the mean was 5.6 percentage points higher in the treatment group than in the control group, though this difference was not statistically significant after adjusting for multiple comparisons. While this could be interpreted as an indication of evidence of promise, it should be taken with caution given the differential attrition observed and the lack of impact detected on other outcomes.

Table 9: BIT Code Stars impact evaluation results (all outcome measures)

Outcome	(1) SCSAS belonging subscale score (Primary)	(2) Total SCSAS score (Secondary)	(3) Intention to study computing (Secondary)	(4) Intention to study science, maths or technology (Exploratory)
Control group mean	3.03	2.89	49.81%	76.24%
Treatment group mean	3.01	2.88	54.67%	75.73%
Estimated treatment effect	0.007	0.020	5.6pp[27]	0.43pp
N	2,599	2,599	2,600	2,614

While the differences found between the primary and secondary outcomes in the *Code Stars* group relative to the control group are positive, they are both small in magnitude and not statistically significant. Further, given the loss in precision and possible biases resulting from the observed attrition and baseline imbalances on the belonging sub-scale survey score, we would advise against interpreting these differences as indicative evidence of a meaningful effect of the intervention.

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

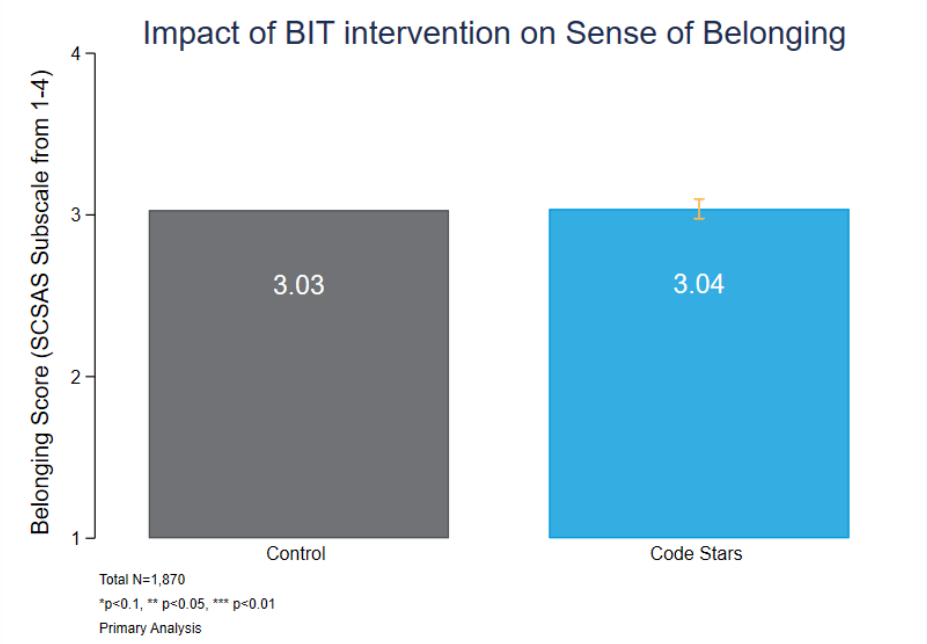

Figure 5: Model-adjusted sense of belonging scores, control arm vs BIT Code Stars treatment (SCSAS scale ranges from 1-4)

[27] While this effect was significant (p<0.10) before correcting for multiple comparisons using the Benjamini-Hochberg step-up procedure, it lost its significance after performing this correction. All other effects were insignificant before correcting for multiple comparisons, and therefore we did not perform this correction on our other p-values.

5. IPE findings on the WISE *My Skills My Life* intervention

This section presents the findings for the IPE on the *My Skills My Life* intervention and is divided into four sections:

1. Fidelity and approaches to delivery, which explores whether the intervention was delivered as intended;
2. Intervention appropriateness, which explores whether the intervention design was fit for purpose and presents our findings on its feasibility and acceptability;
3. Responsiveness and engagement from pupils to the intervention;
4. Mechanisms of change through which the intervention could have affected the target outcomes in the logic model.

5.1 Fidelity and approaches to delivery

Key findings

- Teachers delivered all the lessons planned in the programme at different paces.
- Disruptions due to COVID-19 delayed the start of the programme and forced some schools to change the format of lessons from in-person to online.
- Teachers mostly followed the content provided by WISE to plan the lesson and made minor adjustments to shorten the lessons.
- In spite of COVID-19 disruptions, sessions with role models took place at most schools with only two exceptions where teachers took over the lesson with the help of additional content provided by WISE.
- Role models opted for varied approaches in session content and activities depending on their individual preferences and STEM field.

The following section examines the degree of fidelity with which the teachers and the role models delivered their lessons. This was assessed using data from teacher and role model interviews and lesson observations.

5.1.1 Teacher-led lessons

All teachers either booked a role model session or gave a replacement lesson themselves. The completion dates of the pupil surveys suggest that some teachers completed the programme in less than 12 weeks, presumably by completing more than one session per week, and that others some took slightly longer. If the programme was run according to schedule, there would have been exactly 11 in-school weeks between baseline and endline survey completion. While it is difficult to know exactly how many 'in-school' weeks the intervention took due to variation term schedules and start dates, there was one half-term break during the intervention period, and about half the schools completed endline surveys before the two-week christmas break. This suggests that if schools completed the intervention according to schedule, they would have taken either 12 or 14 weeks to complete it, depending on when they started. The mean number of weeks taken to complete the intervention, inclusive of school holidays, was 12.85 (SD 3.63), suggesting that most schools took between nine and sixteen weeks to complete the survey. The minimum amount of time taken was 7 weeks, and the maximum was 21 weeks.

Disruptions due to COVID-19 delayed the start of the programme and forced some schools to change the format of lessons from in-person to online. This was communicated to BIT by the WISE team in charge of programme delivery and was later confirmed in the lesson observations and teacher interviews.

Interviewed teachers reported following the content provided by WISE to plan the lesson and only making minor adjustments to shorten the lessons. Any adaptations mentioned in the interviews did not appear to substantially alter the lesson content and were aimed at shortening its length in order to fit it in the allotted hour for computing class. One of the interview teachers opted to skip minor parts of the lesson content or deliver those parts more quickly and with less depth than planned in the lesson guides.

5.1.2 Role model-led lessons

Two out of 40 schools in the treatment sample could not complete the 'Real Life Role Model' session with the female role model as intended due to COVID-19 concerns, and the teachers had to take over the lesson, as those schools reported to the WISE delivery team**.** WISE sent them role model booklets with profiles and information about 72 different female role models, and teachers administered a lesson to pupils based on this booklet. One of the teachers confirmed this in the survey. This alternative was prepared by WISE prior to the intervention start in case of any COVID-related concerns the teachers or role models may have had. Meeting the role model was a core part of the intervention design, as it was intended to help pupils develop their sense of belonging through relationship building. Missing this element of the intervention may thus have limited the impact of the intervention in those two schools, compared to the rest of schools in the *My Skills My Life* group[28].

COVID-19 delays also affected the in-person lessons with the role models and caused some schools to deliver them online. An interviewed role model reported having postponed the two sessions she was scheduled to deliver at different schools, and doing them online due to COVID-19 restrictions. The role model stated that in one of the online sessions, she struggled to hear the pupils so the teacher had to be the one relaying their questions to her and the role model could not interact directly with the pupils, which made it more difficult for her to engage with the pupils.

Role models opted for varied approaches in session content and pupil activities depending on their preferences and career fields. WISE provided the role models with an online training, including a framework on how to deliver the sessions, the general structure to

[28] Re-running the quantitative models in the impact evaluation for all outcomes without these two schools did not produce different results.

follow and the resources they could use. Role models adapted the session content depending on their preferences and their professional field. For example, during the interviews one of the role models reported explaining their normal day-to-day at their job, while the other one had preferred to talk about their professional field (cybersecurity) more generally, which she thought the pupils would find more exciting. Both of them reported speaking about their personal background, why they chose their particular field and their path to their current role during their session, following the broad instructions given by WISE.

The role models that we interviewed reported using various strategies to engage pupils, like showing a fun video about cybersecurity, or doing interactive activities that simulated their jobs. One of the role models in the case study was a business analyst and made the pupils solve a business case that involved problem-solving and creative thinking and required them to design a mobile phone application. The other role model worked in cybersecurity and chose to challenge pupils with an activity of decoding an encrypted message.

5.2 Intervention appropriateness: Feasibility and acceptability

Key findings

- The programme had a very positive reception by pupils, role models and teachers.
- Role models felt well supported by the schools, class teachers and the WISE delivery team and maintained reliable and constant communication with them throughout the COVID-19 disruptions.
- Teachers thought the quality of the resources prepared for them was high and easily accessible.
- The length of lesson content and number of lessons were perceived as a risk to the feasibility of the intervention.

Intervention feasibility and its reception by the delivery actors (teachers and role models) were assessed using data from teacher and role model interviews and lesson observations.

5.2.1 Intervention reception

The intervention was described as a positive experience by teachers, role models and children.

Both role models stated in their interviews that they would do it again and recommend the experience to their colleagues and friends in their field. They also reported having positive experiences with the recruitment process, and described the programme website and recruitment process as ‘easy to use’.

Participating teachers who were interviewed also reported having very positive experiences with the intervention. In one of the interviews, one teacher stated that they would recommend the programme, especially for the opportunity it gave to less socio-economic advantaged pupils of any gender to learn about careers they are less exposed to.

"It has helped children aspire to bigger careers - most of them want to work in factories or just not work at all and go on benefits because that's what their parents do. This programme has opened their eyes to other things." (Teacher 1)

5.2.2 Support for role models

The sixth lesson in the programme was designed to be delivered by a female role model working in a STEM field and was one of the core components of the intervention. This section describes the findings on the adequacy of the resources and support that were available for the role models, and the facilitators and barriers they faced during the programme.

Role models identified the following facilitators to delivering their sessions: a) reliable and constant communication and support from the school and from WISE; b) support from the teacher before and during the session; c) having pupils prepare questions for the role model before the sessions; and d) the training provided by WISE.

All role models interviewed in the evaluation were grateful for the clear and constant communication with WISE and the schools, who kept the volunteers aware of any developments and updates and facilitated the reschedule of sessions after the disruptions caused by COVID-19. A role model commented she had felt very supported by WISE and the schools:

"*I think this initiative is great. The way it is being run and managed is also great. I felt quite supported, informed throughout the process."* (Role Model 1)

Both role models reported they were grateful for the help of the class teacher to prepare and deliver their lesson. Both mentioned they had been helped by the teacher to prepare their session and had checked with them beforehand that the content of the lesson was adequate for the pupils.

One of the two role models interviewed by BIT was not able to attend the training provided by WISE before delivering the session, though she communicated with the class teacher to get her approval of the session content before delivering the session.

5.2.3 Training for role models

The training provided by WISE covered information about the project, and gave a framework for using the *My Skills My Life* resources and how to talk about their roles in a way that primary school pupils could understand and relate to.

Role models considered the training and resources helpful to prepare their sessions and decide on their content. One of the role models interviewed reported finding the guidance on how to communicate and engage with the pupils very helpful, whereas the other reported she would have liked more guidance on how to adapt her talk to the pupils' age. She described struggling to know what level of difficulty was appropriate for that age group when explaining some concepts related to her STEM job and when designing the activities to make them interesting for the pupils.

5.2.4 Teacher-delivered lessons

The seven lessons delivered by the teachers were another core component of the *My Skills My Life* intervention. These lessons were designed and scripted by WISE to be delivered without requiring a lot of preparation by the teacher. This section outlines the findings on the appropriateness of the lesson content and the teachers' experiences delivering the lessons.

There was variation in the teachers' feedback on the lesson resources and how much time they had spent preparing the lessons.

The lessons were age-appropriate and enjoyed by both pupils and teachers. According to teacher comments and the pupils' behaviour during the lesson observations, the content was age-appropriate for Year 5 children. Teachers described the content as "enjoyable" and "interesting". An issue raised by one of the teachers was the lack of content tailored to lower attainment pupils or pupils with difficulties with vocabulary.

Teachers appreciated the level of detail in the lesson plans and how easy it was to access the resources. The two teachers we spoke with said that the lessons had a good pace and that they appreciated having clear lesson objectives, success criteria, detailed activities and lists of key questions. They said accessing the resources provided by WISE for the lessons had been straightforward.

The extensive amount of resources and lesson content resulted in longer preparation times for teachers, particularly for those with less computing knowledge. Even though the structure and content of the lessons was already planned for them, one of the teachers reported they had needed more time to prepare than they would usually spend preparing for computing lessons in order to go over the extensive amount of resources. Pre-existing knowledge in computing and STEM fields influenced whether teachers needed more time to prepare for the lesson; some teachers reported having to do additional research on computing and STEM jobs, which increased the time spent on preparing for the lessons.

The full length of lesson content was difficult to cover during the time allocated to computing class and also presented practical difficulties to prepare the handouts. This challenge came across in both teacher interviews. One of the teachers had to leave some content out or go over it with less detail than what was expected in the lesson guides. The amount of content also resulted in long print-outs that had to be distributed to pupils.

The number of lessons was found by another teacher to displace too much of the time needed to teach the rest of the computing curriculum. In one of the interviews, the teacher suggested shortening the programme and reducing the number of lessons to make it feasible in the long term.

5.3 Responsiveness and engagement from pupils

Key findings

- The session with the role model was very popular with pupils.
- Pupil engagement was lower in online lessons relative to in-person ones.
- Teacher experience and ability were important moderators for pupil engagement.
- Pupils least interested in the subject were the hardest to engage.

The pupils' responsiveness and engagement with the lessons and materials was assessed using teacher and role model interviews, field observations of the fifth lesson (led by a teacher) and the sixth lesson (attended by a role model) and pupil focus groups conducted after those observations.

The *My Skills My Life* programme was met with positive reception from the pupils. Some words used by pupils to describe the lessons in the focus group discussion sheets were "interesting", "fun" and "cool". There was no negative feedback from pupils in their comments.

5.3.1 Facilitators to pupils' engagement

The intervention features described below were found to support pupil engagement during the teacher and the role model lessons.

The session with the role model was the most popular part of the intervention. Pupils actively participated and were enthusiastic when they could interact with the female role models and ask the questions they had prepared for them in the previous lesson. That activity was successful in engaging most of the pupils in the class, as reported by role models and teachers of different schools, even in those cases where the lesson had to be conducted online, which made interaction and rapport between the role models and pupils more difficult. In the lesson we observed, pupils asked the role model many questions about her personal interests, such as her hobbies, whether she liked their job, or whether she was into gaming. The other role model interviewed reported spending time answering the same type of questions. She also reported inviting an additional person working in a similar field to provide support during the activities and help answer the pupils' questions.

Both interviewed teachers thought that meeting the role model had been the favourite part of the programme for the pupils. This was reiterated by some pupils in the focus groups.

"They've certainly enjoyed meeting the role model and doing that lesson of - as you know, one child was like, completely changed his career path, what he wants to do. There's definitely impact there." (Teacher 1)

"I think the really important thing was our live role model." (Teacher 2)

*"I loved learning about [*role model name]*'s job during the day. it was so cool"* (Pupil 1)

"My favourite bit was when the lady talked about how she got her job" (Pupil 2)

Problem-solving and creative activities planned by the role models were met with high levels of engagement during their sessions. Role models planned activities that simulated parts of their job such as coding, decoding an encrypted message or designing a mobile app to solve the business case, which were very successful in keeping the pupils' interest, according to the interviewed role models and teachers and observation notes.

The quiz planned for session 5 and other activities that promoted self-reflection by the pupils received very positive feedback from pupils and kept them engaged, according to teacher comments. The quiz asked the pupils about their skills and preferences and matched them with recommended job roles. Pupils were also engaged with other activities like watching videos during the lessons and the role models session, and discussing real life examples and the applications of computing in the children's daily lives.

Pupils liked to use electronic devices like laptops or iPads. Getting to use them was reported as one of the things pupils looked most forward to during computing class, both in the usual classes and also during the teacher-led lessons designed by WISE, even though one teacher commented that the lesson content was more paper-based than their usual class.

Teacher experience and ability were found to be an important moderator for pupil engagement in the observed lessons. Role models reported being helped by teachers in managing the pupils' speaking turns and keeping the pupils engaged during the session. This was particularly the case for the role model who had to do her session online, who noted having to rely on the teacher for leading the discussion and moderating pupil participation.

5.3.2 Barriers to pupil engagement

The barriers to pupil engagement described below were identified through the IPE.

Having the lesson with the role model online (as opposed to in-person) hindered the interactions with the pupils, which resulted in lower pupil engagement in the role model session observed by BIT, as mentioned in the previous section.

Lack of pupil interest in the subject (either computing or the role model career field[29]) was a common determinant of low engagement across the interviews and observations. One role model reported that the uninterested pupils in the class were mostly girls and that they kept quiet for most of the class. This is consistent with the slightly lower baseline SCSAS scores observed for girls relative to boys, described in section 3.2.

Difficulties with language and vocabulary were a barrier for low attainment and non-English speaking pupils. This was mentioned by an experienced teacher in their interview, who said that the issue became a problem in lessons that required a lot of reading. This barrier was also observed during lesson observations. Experienced teachers who could

[29] The two role models interviewed for the IPE worked in Business analysis and Cybersecurity.

identify these difficulties were able to partially offset this barrier by providing extra support and attention to those pupils.

5.4 Mechanisms of change

Key findings

- There was some evidence that the intended mechanisms were addressed to support an increase in pupils' sense of belonging and positive attitudes towards computing.
- These mechanisms included increased interest and motivation to learn more about computing careers, increased awareness of computing and STEM careers that could fit their skills and preferences, and increased confidence with computing.
- The evaluation did not identify any backfire effects for male pupils.

This section describes how the intervention targeted the mechanisms intended to contribute to increase girl pupils' sense of belonging and positive attitudes towards computing hypothesised in the logic model (Figure 1). These mechanisms included 1) increased interest and motivation to learn more about computing careers, 2) increased awareness of computing and STEM careers that could fit their skills and preferences, and 3) increased confidence with computing and computing terminology.

These findings are based on teacher and role model interviews and pupil focus groups.

5.4.1 Increased interest and motivation to learn about computing-related careers

Pupils found the role models' jobs exciting and wanted to know more about them. The role models showed their day-to-day lives and explained their jobs using fun activities, like the business case or the decryption activity, which seemed successful at generating pupil excitement at both of the observed schools. In the focus groups, pupils expressed excitement about the role models' jobs:

"I loved learning about Hannah's job during the day. it was so cool" (Pupil 1)

"my favourite bit was when the lady talked about how she got her job" (Pupil 2)

"It's definitely had an impact on the children, which I've liked to see. One of my children, he's completely obsessed now with a business analyst" (Teacher 1)

Teachers reported noticing that both boy and girl pupils were more interested in computing after the role model and teacher-led lessons. One of the teachers was surprised about the increase in interest:

"There's been a few children, that I wouldn't have expected to be as interested, that have shown more of an interest in computing." (Teacher 1)

The higher interest in computing extended to coding and motivated pupils to engage with computing beyond the classroom and learn more about it. In one of the case studies, a teacher reported that some female pupils had joined the school Code Club after the programme[30].

Class observations and teacher interviews could not identify any differences in attitudes towards computing or their engagement with the subject in girls over boys, though there were descriptive differences between boys' and girl's scores on our main outcomes, as described in section 3.2.

5.4.2 Increased awareness of computing and STEM career opportunities

Certain activities specifically targeted the mechanism of increasing pupil awareness of computing and STEM careers that could fit their skills and preferences. For example, the fifth teacher-led lesson made the pupils take a quiz to discover what their preferred way of working was and then made the pupils explore job profiles in STEM that better matched their skills.

In the two case studies, the lessons and the quiz promoted self-reflection on the pupil's own skills, their likes and dislikes and their thoughts on potential future jobs. In the worlds of a pupil:

"[Learning about computing makes me feel] *good because it helps me think more about what we want to be / what I want to be*" (Pupil 3)

The intervention widened the understanding of the variety of computing-related careers available to them and made pupils aware of their importance. In the two focus groups, pupils described computing as important to get good jobs in the future and to improve their skills. Interviewed teachers also pointed out that thanks to the programme, pupils got to know about job roles they would not have been exposed to otherwise, because of a lack of similar role models in their close environment.

"*It has helped children aspire to bigger careers - most of them want to work in factories or just not work at all and go on benefits because that's what their parents do. This programme has opened their eyes to other things*." (Teacher 1)

"I just kept saying to the children, 'I didn't know these jobs existed.' Especially with us being rural and where we are, and so that's why it was an eye-opener for the children as well." (Teacher 2)

5.4.3 Increased confidence with computing

After the lesson with the role model, girls reported being more confident in their computing abilities and felt equally capable and valid as boys for STEM jobs. The message that girls were equally valid and competent in computing as boys resonated a lot

[30] This school was also participating in the GBIC Code Clubs trial implemented with guidance from RPF and evaluated by BIT.

with female pupils, according to interviewed teachers and role models, and appeared several times during the pupils’ discussion repeated by both girl and boy pupils:

"*Anyone can be anything they want"* (Female pupil)

"If a boy can do it, a girl can do it - no one’s better than the other" (Male pupil)

Girls also had more confidence in their computing and coding skills and became more comfortable with them after the lessons that included coding activities, as observed by both of the interviewed teachers and the two focus groups with pupils.

One interviewed teacher reported believing that exposing female pupils to female role models in STEM during the lessons increased their awareness of the female presence in the STEM career fields and their sense that they could go on to have a career in these fields:

"*When it was showing all of the females in the jobs, nobody went 'Oh, I didn't know that a female could do that' but I think they were amazed by the role of jobs and the fact it was all females doing it*" (Teacher 2)

By interacting and asking questions to the role model about her likes and dislikes and her background, female pupils could see the female role model as relatable and a possible career path for themselves.

“It was the meeting the role model that had the biggest impact on them. [...] It was just that opportunity to meet somebody who actually does these roles that they don’t see in normal life, and actually see the fact that, ‘I can actually do computing when I’m older, it’s not just something that I learn at school’” (Teacher 1)

We did not find evidence of increased familiarity with computing vocabulary playing a significant role in increasing confidence with computing in general, though this was one of the aspects envisioned for this mechanism in the logic model.

5.4.4 Backfire for boys’ attitudes towards computing

We did not find any evidence of negative impact or backfire effect on boys. The female-centric role model session or job role examples with only women did not lead to a lower engagement during the session. The impact evaluation also did not report lower numbers in the sense of belonging in computing of boy pupils after the intervention (see appendix 2).

One of the role models reported in her interview that most boys participated in the class discussion while a few girls uninterested in the subject were very quiet during the lesson. In that case the boys were already more interested in the subject and more confident in their computing skills, and they ended up taking a more prominent role during the discussion with the role models and the activities. She also reported that having one-on-one interactions with them was effective in engaging the less involved girls.

6. IPE findings on the BIT *Code Stars* intervention

The following section presents the findings for the IPE on the *Code Stars* intervention and is divided into four sections:

1. Fidelity and approaches to delivery, which explores whether the intervention was delivered as intended;
2. Intervention appropriateness, which explores whether the intervention design was fit for purpose and presents our findings on its feasibility and acceptability;
3. Responsiveness and engagement from pupils to the intervention;
4. Mechanisms through which the intervention addressed the target outcomes in the logic model.

6.1 Fidelity and approaches to delivery

Key findings

- Observed and interviewed teachers delivered the two planned lessons and used the provided resources as intended.
- Teachers incorporated most of the tips in the action guides in their teaching, though some were reluctant to use some of the optional tips on inclusive teaching.
- Teachers used all four options provided to send messages to parents (text message, email, online homework task and hard copy homework task). Using hard copies was slightly preferred over the other options.
- Lack of parental engagement and interest were cited by surveyed and interviewed teachers as reasons for not sending all or some messages.
- Teachers struggled to follow the recommended weekly schedule and some sent the messages with monthly frequencies or all at once, which likely limited their effectiveness in engaging parents.

This section describes the degree of fidelity with which the teachers taught the lessons and sent the 9 message prompts. This was examined through teacher and parent interviews, and lesson observations, as well as an online survey emailed to teachers after the intervention.

Code Stars stand-alone lessons

Interviewed teachers and those who answered the survey delivered the two planned lessons and followed the lesson guides. Seven out of 8 teachers who answered the online teacher survey reported delivering both lessons (the one-hour long main lesson and the 20-minute long follow up), while the remaining teacher delivered only one of them. Teachers reported following the detailed lesson guides and using all the content provided to plan their lessons in the interviews and the survey. No notable deviations on lesson content and length were seen during the lesson observations or flagged in teacher interviews.

All surveyed and interviewed teachers relied on the action guides to different degrees, but some teachers were reluctant to use some of the teaching tips on inclusive teaching. The use of action guides with tips and activities for inclusive teaching was optional but recommended. According to survey responses, their use was widespread but teachers relied on it to different degrees (1 teacher in the survey used all of them, 4 used most of them and 3 used some of them). An experienced teacher reported that they had found the encouragement phrases for girls pupils awkward and struggled to integrate them in their teaching in an organic manner. A survey respondent also reported struggling to use the encouragement phrases during the *Code Stars* lessons, but that they had found them easier to use during a coding lesson with year 4 pupils.

Pupils used the required laptops or tablets, but sometimes had to share them depending on the school resources. Pupils in the case study schools either had an individual device for themselves or had to share it with another classmate, depending on available resources at the school.

Conversation prompts for parents

Several instances of non-compliance and partial compliance with important deviations were observed during the evaluation.

Teachers used one or more of the four messaging options provided and did not deviate from those. There was a slight preference for using hard copies reported in the teacher survey. Among teachers who answered the survey, one used emails, one used phone messages, three opted for giving the hard copies to the pupils and and one used the online teaching platform in their school (which was also the option preferred by one of the interviewed teachers). They used each approach by itself or a combination of two (for example, giving hard copies to pupils and also posting them online). One of them also reported using a different option and gave the hard copies directly to parents. The choice of format depended on the teacher's judgement but also on the school messaging culture and the system usually used by the school to send messages to parents.

Teachers struggled to follow the recommended weekly schedule and sent the messages with other frequencies. The weekly calendar was recommended by BIT, as the conversation prompts were meant to be sent throughout the programme to promote conversation at home more or less regularly during the 12-week duration of the intervention. Only 2 out of 8 survey respondents sent them weekly in a regular schedule, and two others spaced them more and sent them monthly. Two teachers sent the messages all at once, which may have reduced the chances of engaging parents during the programme.

One of the teachers who complied with the recommended schedule reported that it had been hard to remember to do it. An interviewed teacher who did send them regularly reported having pre-scheduled their sending.

Despite these challenges with compliance to the message schedule, 6 out of 8 teacher respondents of the online survey sent the messages to parents. Two out of 8 teachers who answered the survey did not send the message to parents, and the two interviewed parents (from one school) said they had not received them. One of the reasons cited for non-compliance was that teachers were sceptical about the effectiveness of sending the prompts to parents.

Lack of parental engagement and interest were repeatedly cited by teachers as barriers to sending the messages. In the online survey, teachers mentioned they were discouraged from sending the messages due to low parental engagement, low parent interest and confidence about computing, and parents not seeing computing as a priority compared to other things like the weekly maths and English homework.

"Some parents did not want to support their children due to their own lack of confidence. Others did not see it as important as doing the weekly maths and English homework." (Teacher survey respondent)

An interviewed teacher who did not send the messages reported that they chose to do the homework reflection and discussion tasks in the classroom instead. They justified that choice saying they had doubts that the pupils and parents would do it, and thus opted to do it themselves in the classroom.

6.2 Intervention appropriateness: Feasibility and acceptability

Key findings

- Teachers found the quality of resources satisfactory, their access straightforward and the lessons easy to teach. Interviewed teachers suggested making the resources downloadable or available offline to avoid disruptions caused by problems with Internet connection.
- The focus on AI was a challenge for teachers unfamiliar with the topic.
- Teachers thought the coding activity was too easy for Year 5 pupils.
- Administrative tasks like sending the messages to parents and overseeing the completion of the pupil survey were very time consuming in some cases.
- Teachers found the content very broad and would have liked to have more specific learning objectives.
- The suggested weekly schedule for the prompts added an administrative burden on the teachers every week. Spacing them more or sending regular reminders to teachers could improve their feasibility.

Intervention feasibility and its reception by teachers and parents was assessed using data from the teacher survey, teacher and parent interviews and lesson observations.

6.2.1 Design of lessons and resources for teachers

The *Code Stars* one-hour lesson was fully planned by BIT and the resources were designed to be ready to deliver: teachers were provided with power-points for both lessons, separate teaching instructions with time guides, an action guide and links to the online resources for the teaching activities.

Teachers found the quality of resources satisfactory and appreciated the level of detail of the teaching instructions, which included time guides that set the pace of the lesson. The contents and resources were available in one place and teachers found them easy to find. Both interviewees and several survey respondents stated the teacher instructions were helpful and described the lessons as “easy to teach” and “easy to plan”.

However, one of the interviewed teachers considered the resources were too detailed and numerous and reported spending more time than expected to prepare the class and go over all of the content provided by BIT.

The content of the AI-themed lesson was found to be too broad and teachers wished for more guidance on specific learning objectives and takeaways for the pupils. One of the interviewed teachers felt the lesson aimed to cover too much and that a more targeted approach would have been more effective. The other one stated that the *Code Stars* content was harder to structure than their normal computing classes. The teacher did not know what the expected outcome should be and how the children should be reacting to it, and did not feel confident in their delivery.

Both interviewed teachers described the rubbish picking game and the AI examples as too easy and simple for Year 5 pupils. In the teacher surveys, one respondent added that sometimes the children had prior knowledge so the lessons finished earlier than they had planned. In the rubbish picking game, pupils had to help a robot identify what was a fish and what was floating rubbish in the sea, and this was designed to give them an idea of how algorithms work. Afterwards pupils had a discussion about real-life examples of AI.

The AI-themed content was a challenge for teachers unfamiliar with the topic. Although the resources were designed assuming no previous knowledge on AI from the teacher, one teacher unfamiliar with the topic reported they had to take extra time of preparation to research it before the lesson, and struggled to explain the concept with confidence given their lack of related knowledge. This challenge was also reported by a teacher in the online survey.

> *“The content was a challenge as I am not very knowledgeable about AI and I did not know what to teach the children and where to stop with the information.”* (Teacher survey respondent 7)

This finding implies that the lesson content was found easy by the children but hard for some of the teachers. This is possible because the game was designed as a simple labelling exercise similar to many other games played by children at that age, and the AI examples were easy to understand for pupils as they related to real life. However, the difficulty of the content for the teachers was much higher as they had to explain the concept of AI to pupils, which may be challenging to do confidently for teachers without computing knowledge. This was the case for the interviewed teacher who mentioned being uncomfortable with the topic, while the other teacher BIT interviewed was specialised in teaching computing and found the subject easy to explain.

The resources for the learning activities relied on a stable Internet connection as some of the lesson activities needed access to a website. In one of the observed classes, slow Internet connection generated some technical problems and caused a long disruption that slowed the pace of the class. Teachers in interviews and surveys suggested making the resources downloadable or available offline.

6.2.2 Conversation prompts

The resources for the conversation prompts were described as easy to find and use. No issues about the content of the messages or difficulties with the sending formats offered by BIT were raised during the evaluation.

Teachers struggled to follow the recommended weekly schedule. One respondent of the teacher survey admitted they had a hard time remembering to send them. The weekly schedule may have placed an excessive administrative burden on top of the rest of teaching responsibilities, as only few of the survey respondents managed to send the messages every week and two opted to send them all at once. The amount of time spent on administrative tasks like sending the prompts or printing the resources was also brought up in one of the interviews as a concern for the feasibility of repeating the implementation.

The most commonly reported barrier to sending the prompts was lack of parental engagement and interest and the difficulty of checking if the parents were using them. As mentioned in Section 6.1, teachers were sceptical about the effectiveness of the prompts to engage parents and cited these factors as a main reason for discouragement from sending the prompts.

6.3 Responsiveness and engagement

Key findings

- Teachers perceived parental engagement to be low throughout the programme. Possible reasons mentioned by teachers were parents' perception of computing as less important than other subjects, lack of familiarity with computing, and low confidence.
- Similarly, low computing skills and knowledge, as well as language barriers, made it hard for an interviewed parent to engage in conversation about computing at home.
- Pupils found the one-hour AI lesson fun and engaged with the rubbish picking game and the open discussion.
- Pupil confidence, computing skills, academic ability and baseline interest in computing were important determinants of pupil engagement.
- Lack of access to technology at home prevented some pupils from doing homework.

This section presents the findings on how the programme recipients (parents and pupils) responded to and engaged with the intervention. This research question was answered using teacher and parent interviews, observation notes from observing the one-hour lesson in the two case study schools, and the pupil discussion sheets gathered during the focus groups after the observed lessons. For parent interviews, it is worth noting that the parent sample in the IPE was reduced to two parents from the same school due to recruitment challenges, as has been noted in section 2.2, and it is not likely to be representative of the average rate of compliance for the programme.

6.3.1 Parental engagement

Parental engagement was low throughout the programme. This was reported by the two interviewed teachers and teacher survey respondents, and illustrated by the difficulties faced by BIT researchers during parent recruitment for the interviews described in section 2.2. Low parent exposure to the intervention and/or engagement led to difficulties to identify parents who 1) had agreed to receive the messages and 2) had received them and were willing to participate in the interview. As a result, BIT had to interview two parents from one school who agreed to participate in the programme but did not receive the message prompts. These difficulties identifying parents who had engaged with the programme despite support from teachers suggests uneven or low parent engagement, and is consistent with the findings described in section 6.1 that lack of compliance by some teachers meant that many parents likely never received the message prompts.

Teacher survey responses suggested that parents did not perceive computing as important as other school subjects and prioritised other school activities like helping their children with their weekly maths and English homework. Lack of confidence was also cited in the survey as a reason for some parents not wanting to support their children with computing.

School culture on communicating with and engaging parents in school activities appeared to be a moderating factor for the parents' engagement in the teacher interviews**.** An experienced teacher, who had sent the prompts, was sceptical about their effectiveness to engage parents and get them in high-level discussions on the importance of computing just by sending them the messages. They commented in their interview that engaging parents was something the school usually struggled with.

> *"I don't really know how much parental engagement we would have had off the back of the prompts in the videos. It's something that we always struggle with anyway. I don't think sending home the prompts and asking the children to discuss it with their parents is a very realistic expectation for us, without a lot more pushing."* (Teacher 4)

They also added that the parents who were more responsive to the prompts were those already more engaged with the school activities and their children's education, while the teacher had struggled to engage less involved parents. The same teacher suggested that giving pupils physical prompts like certificates (like the one given at the end of the follow-up lesson), notes or stamp cards was usually more effective in getting the pupils talking to their parents and doing activities outside of the classroom, in their experience.

Language barriers may have made it hard for some parents to engage in conversation with their children about computing at home. A non-English speaking parent reported in their interview that they struggled to talk to their children about the computing activities they were doing at school (among those, the *Code Stars* lesson) in detail due to limited vocabulary in English related to computing and coding. This language barrier could also have extended to using the conversation prompts, as they could be hard to understand to non-English speaking parents, though we were not able to investigate this as both interviewed parents had not received the prompts.

Previous knowledge and familiarity with computing was an important moderator for parent engagement. Parents' awareness of the importance of computing and STEM careers and their familiarity with the subject before the intervention emerged as an important determinant of their engagement in parent and teacher interviews. One of the interviewed parents worked in a STEM-related career and also had relatives working in computing, and was thus aware of its importance and reported actively encouraging their child to engage with computing at home.

Parent knowledge and computing skills also affected the degree to which parents engaged in conversations about the activities that children had done during the *Code Stars* lesson. An interviewed teacher observed that in their class, several pupils were more skilled and familiar with computing and electronics than their parents. A parent reported that they struggled to understand what her child was doing in class, which limited the kind of conversations they were able to have on the subject.

These findings suggest that the intervention design was effective for parents who were already more engaged than average with their children's education in computing and familiar with the topic, and less so to engage other kinds of parents that may need a more intensive intervention to engage with their children's computing education. The evidence on parental engagement with the prompts, having discussions at home about computing and encouraging their children to learn and engage with it, came from parents who were already a) more responsive to communications from the school and the teachers, b) more familiar with computing and aware of its importance because of their background or social circles, and/or c) engaging in conversations with their children about their future and their preferences. Meanwhile, the intervention struggled to target the barriers preventing engagement from parents who did not have those characteristics. The generation gap in knowledge and skills for computing between parents and children was also a common concern in the parent interviews and a barrier to them engaging in those conversations that may not have been addressed by the intervention design.

6.3.2 Pupil engagement

The pupils' responsiveness and engagement with the lessons and materials were assessed using teacher and parent interviews, lesson observations and pupils' discussion sheets from the focus groups.

Pupil engagement in the lessons was very good in the two observed cases; this was also reported in teacher interviews and surveys. No notable differences in engagement between boy and girl pupils were observed in the lessons. The teacher who taught the lesson online commented that pupil engagement was similar to how they behaved in their usual computing classes. Several surveyed and interviewed teachers commented that pupils enjoyed the lessons and tasks and found them interesting. One of the interviewed teachers said that the activities had been able to engage a pupil in their class who did not like computing in general.

The lesson was generally enjoyed by the pupils in the two focus groups, although some pupils reported not liking it and others not finding it challenging enough, which matched what their teacher had said during the interview about the difficulty being too low for that year group. Another interviewed teacher commented that the low difficulty of the programme helped less advanced pupils to keep up with the lesson.

Differences in engagement were related to baseline interest in computing and coding. In one of the case study schools, the pupils who engaged the most with the intervention lesson and activities were high performing boy pupils who were already more interested in computing and coding and more skilled in both, according to the statements of their teacher, while some girls and less computer savvy pupils in general ended up relying on their more advantaged classmates to do the activities. The teacher reported that the intervention seemed to give new perspectives on computing and fostered curiosity on the subject, but that this was mostly the case for boy pupils.

Facilitators to pupil engagement

The intervention features and activities described below were found to facilitate pupil engagement with the lesson.

Pupils found the AI theme interesting and enjoyed the discussion around it, especially around the real-life examples. One of the teachers reported that pupils were very engaged with high level discussions on the ethics of AI. Pupils were observed actively participating in the discussion in both of the observed lessons, especially when the discussion involved examples of AI present in the children's lives. A teacher from the case studies pointed out that the more open ended discussion task, which did not have a right or wrong answer, was easier and more engaging compared to the robot coding activity for those pupils who were not confident about their skills.

Pupils found the rubbish picking game fun and showed high engagement with it, as evidenced by lesson observations and pupils' comments in the focus groups. One of the interviewed teachers said the task prompted children to use their imagination and problem-solving skills. Coding was also the favourite part of computing lessons for several pupils in the discussion sheets.

The creative activity also gave a reason to pupils to 'show off' what they had achieved in their Code Stars class to their parents and friends after school, according to the statement of a teacher.

> *"The interactive aspects worked well as the children were engaged and we were able to speak to them more fully about what they were doing and the opportunities computing can lead to, whereas some lessons often involve more troubleshooting which enables less discussion."* (Teacher survey respondent)

Teacher characteristics and experience were also important for keeping the pupils engaged and interested during the lesson, especially when the lesson focused on more complicated topics. Explaining complicated concepts in an accessible way and relating them to the pupils' lives by using real life examples of computing and AI were some of the techniques observed during the lesson observations to keep pupils engaged, especially those with learning or language difficulties.

Barrier to pupil engagement

Pupil confidence, computing and coding skills, and pre-existing interest were important determinants of engagement during the coding activity and the discussion. A teacher reported that the less confident pupils in their class were reluctant to answer questions or work in problem solving exercises and deferred to more skilled classmates to solve the task for them.

Some of the more advanced concepts could be an impediment for pupils with learning or language difficulties, who struggled to understand them. Teacher experience was an important moderator because it determined their ability to explain complex concepts to pupils, promote discussion in class and involve the pupils with learning difficulties.

Lack of access to technology at home was a barrier for some pupils to do their homework and continue to engage with computing after school. One of the case study schools had provided their pupils who had no access to technology with an iPad so they could do their homework.

6.4 Mechanisms of change

Key findings

- There was some evidence that one of the intended mechanisms was partially addressed to support an increase in pupils' sense of belonging and positive attitudes towards computing.
- Pupil engagement and interest in the two lessons were high, suggesting that the intervention generated interest in computing, though it is unclear that this generated substantively more interest than usual computing classes.
- Low teacher compliance in sending prompts and low parental engagement suggest that the intervention did not effectively increase discussions on computing at home, though the parent interview sample limited our ability to investigate this mechanism.
- Pupils saw computing as important after the lesson, however there was no evidence that they related this perceived importance to prosocial aspects of computing.

This section describes how the intervention targeted the mechanisms intended to contribute to increase girl pupils' sense of belonging and positive attitudes towards computing hypothesised in the logic model (Figure 1). These mechanisms included 1) increased interest and enjoyment of computing (mainly via the two lessons), 2) encouraging active discussions around computing between parents and pupils (mainly via the prompts to parents), and 3) exposure to prosocial aspects of computing (via both the lessons and prompts).

These findings are based on teacher and parent interviews, lesson observations, pupil focus groups and online teacher surveys.

6.4.1 Increased interest and enjoyment of computing

Pupil engagement and interest in the two sessions were both high, as described in section 6.3.2. Pupils enjoyed both the AI theme and the examples covered in the lessons, suggesting that the intervention did generate interest in computing.

One teacher reported that computing and programmer careers were more popular after the intervention: "coder/programmer" was a more popular career path in a survey the pupils did afterwards, as opposed to other traditionally popular options among the children like footballers, popstars, etc. The film "Hidden Figures" which was featured in one of the message prompts also raised the pupil's interest in a coding career, as said by the same teacher.

"There's been a few children that I wouldn't have expected to be as interested, that have shown more of an interest in computing". (Teacher 3)

However, there was no clear evidence in the IPE that the intervention generated further interest than normal computing classes, or that this increase was substantive, at least in the short term. None of the interviewed teachers noticed any increase in engagement during normal computing class from girl pupils after the intervention, although they also said it was too early to tell as none of them had had enough normal classes after the intervention. They also did not report observing any notable differences from the usually high engagement normally observed by them during curricular computing classes.

6.4.2 Promoting discussions on computing between parents and pupils

Teachers' perceptions were that parents did not engage with the conversation prompts generally, or would not if they sent them; thus the mechanism of promoting active conversations about computing appeared not to be targeted successfully. These perceptions were raised during the two teacher interviews and in the online teacher survey by most of the respondents.

It is important to note that the IPE could not gather data directly from parents about their experiences with the conversation prompts on whether the prompts were able to engage parents in active discussions on computing with their children after school, as both of the parents interviewed reported not having received them. The IPE was hence limited in its ability to investigate the effectiveness of the prompts to achieve the intended mechanism (generate conversations about computing at home) from the perspective of parents.

However, the parents who were interviewed said that their children had enjoyed the *Code Stars* lesson and had talked about it with them enthusiastically, although that was not prompted by the conversation prompts. Those interactions had come about during usual conversations with their children about what they had learnt and done in school that day.

6.4.3 Exposure to prosocial aspects of computing

Computing was seen by pupils as more important after the lesson, as shown by pupil comments and the comments of an experienced teacher. The lesson on the real life uses of AI and how programming a robot works helped pupils to link what they were learning about in computing class with the real world uses of computing, which is a connection they usually do not make in such an early age, according to the teacher's view. This teacher said in their interview that the intervention showed pupils the utility of coding and the role of programmers, and made them more aware that it was a valid career option. Several pupils commented during the focus groups that they saw computing as important for their future careers and for developing their skills.

However, we found no evidence that pupils related this perceived importance to prosocial aspects of computing, rather than importance to their personal skills development or future careers.

7. Conclusions and recommendations

7.1 Summary and interpretation of findings

Evidence of impact

We found no conclusive evidence that the *My Skills My Life* or the *Code Stars* interventions had an impact on the following outcomes: a) girls' sense of belonging in computing, b) their attitudes towards computing, and c) their intention to study computing in the future. While the differences found between the primary outcomes in each of the treatment groups relative to the control group are positive, they are both small in magnitude and not statistically significant. Further, given the loss in precision and possible biases resulting from the observed attrition and baseline imbalances on the Belonging sub-scale survey score, we would advise against interpreting these differences as indicative evidence of a meaningful effect of the interventions.

For the secondary outcome of stated intention to study computing in the future, after controlling for our covariates the mean was 5.6 percentage points higher in the *Code Stars* group than in the control group, though this difference was not statistically significant after adjusting for multiple comparisons. While this could be interpreted as an indication of evidence of promise, it should be taken with caution given the differential attrition observed, the lack of impact detected on other outcomes, and the limitations of this indicator in predicting that pupils will go on to select GCSE Computer Science (covered in section 2.1.5).

While this did not translate into evidence of impact on the measured outcomes, the *My Skills My Life* intervention was positively received and seemed to address the barriers targeted

The IPE findings showed a very positive reception of the *My Skills My Life* intervention by teachers, pupils and role models. Disruptions due to COVID-19 delayed the start of the programme and forced some schools to change the format of lessons from in-person to online, where pupil engagement appeared lower in case study schools. Despite these disruptions, teachers at case study schools reported following the content provided by WISE to plan the lessons, and only making minor adjustments to shorten them given the large amount of content. Sessions with role models took place at all but two schools, where teachers took over the lesson with the help of additional content provided by WISE. These sessions were very popular with pupils, and role models adapted the lesson content depending on their individual preferences and STEM field.

We found some evidence of the intended mechanisms hypothesised to support an increase in pupils' sense of belonging and positive attitudes towards computing, including increased interest and motivation to learn more about computing careers, increased awareness of computing and STEM careers that could fit their skills and preferences, and increased confidence with computing.

The IPE findings suggest that the *Code Stars* lessons were enjoyed by pupils and that the intervention targets a relevant barrier, but the impact of the messaging component may have been limited by delivery and dosage issues

The IPE findings suggest that teachers delivered the main component of the intervention (the two AI-themed lessons) and used the tips for inclusive teaching included in the action guide as planned. In the two focus groups, the lessons were described as fun and able to engage the pupils' interest in coding and the importance of computing. The delivery mechanism for the messages to parents component of the intervention had to be adjusted in response to the COVID-19 context and the resulting stress on school resources. Even though the conversation prompts provided by BIT were found easy to use by teachers, they struggled to follow the recommended weekly schedule when sending the messages to parents; COVID-19 disruptions and competing priorities for teacher and parent attention may have contributed to this challenge. Six out of 8 teachers in the teacher survey reported having sent the messages in some form, but 4 out of 8 teachers sent them with monthly frequencies or all at once, which may have limited their effectiveness in engaging parents.

There are a number of possible reasons for the lack of evidence of impact despite the positive reception to the interventions

There are multiple possible reasons related to the design of the My Skills My Life and Code Stars interventions, their delivery, and measurement challenges for the absence of conclusive evidence of impact despite these positive reported experiences:

- **Dosage:** Delivery challenges such as needing to deliver sessions online due to COVID-19 restrictions or insufficient time to cover the full lesson content could have prevented each intervention from being fully implemented at all treatment schools.

- **Sensitivity of outcome measures:** As discussed in section 2.1.5, the nature and objectives of the intervention (which involved influencing future career choices of 9-10 year-old pupils) provided challenges in reliably measuring the intended outcomes, which created a need to rely on short-term and pupil-reported proxy indicators that may not capture the full and longer-term impact of the intervention in pupil choice of study subject in the future. While strategies were implemented as part of the evaluation to mitigate these challenges, they are unlikely to fully address them, and some teachers reported pupil challenges in understanding and completing the SCSAS. This may have led to some measurement error and limited the ability to measure the constructs precisely enough to identify variation between the treatment and control group.

- **High engagement with computing at baseline:** Attitudes towards computing were found to be high in all groups in the sample at baseline. This may be partially due to the characteristics of the sample (schools with at least one teacher interested in taking part in a gender balance in computing project and higher buy-in from school senior leads), which could differ from those of the general school population. If this is the case, it is possible that there is less scope for the intervention to lead to greater improvement. This may be especially true for the final sample of schools used in the analytical sample, if those that failed to integrate the baseline and endline surveys into their lessons and thus dropped out were less interested in innovative teaching approaches for computing.

- **Changes in intended outcomes are not observable immediately after the intervention:** Given the measurement challenges associated with the target outcomes, it is possible that this intervention set off a chain of mechanisms which will eventually contribute to improved attitudes towards computing or actual subject choice, but that changes in these outcomes were not yet apparent when measured immediately after the intervention has finished.

There is not sufficient evidence to confidently determine the extent to which each of these factors may have influenced the evaluation results. The data collected as part of the evaluation also confirmed the relevance of interventions targeting gender balance in computing at Key Stage 2 level: in our sample, at baseline the difference between boys and girls was largest in their reported intention to study computing in the future (71% vs 56%).

7.2 Recommendations

Recommendations for the *My Skills My Life* intervention

The following steps could help to build on the positive teacher and pupil response to the *My Skills My Life* intervention:

1. **Facilitate teacher delivery of the intervention through reduced content or number of lessons**
   The feasibility of delivering the intervention could be improved by reducing the amount of content per lesson and time required to prepare for the lessons, or reducing the number of lessons, to reduce the disruption with the usual curricular computing lessons..

2. **Provide additional opportunities for pupil engagement with role models**
   Given the high pupil engagement with the session involving a role model and the very positive feedback from the participating role models that volunteered, providing more opportunities for direct engagement between pupils and female role models working in computing-related fields could reinforce any positive effects of the intervention while lessening the burden for teachers, though this is conditional on the availability of role models. Ensuring these interactions take place in person rather than online whenever possible would help with pupil engagement.

Recommendations for the *Code Stars* intervention

The following actions could improve the delivery and potential impact of the *Code Stars* intervention:

1. **Increase number of lessons to reinforce targeted mechanisms of impact**
   The teacher feedback on the action guide and the rest of the resources provided for the AI-themed lessons, and pupil engagement in the lessons were both positive in the

case study schools, but more lessons may be necessary to more substantially increase girls' interest in computing and expose them to prosocial aspects of computing. The positive response to the current two lessons is an encouraging sign that additional lessons may be well received.

2. **Review the mechanisms for teachers to communicate messages to parents**
   The communication mechanism initially envisioned for this intervention, SMS messages, had to be revised based on feedback from schools and to avoid data privacy concerns. The IPE findings suggest that the alternative mechanisms made available to teachers, such as emails and hard copies given to children as homework tasks, should be refined, and that providing clear guidance and additional reminders to teachers may be more effective than providing flexibility to choose the communication format and timing.

Broader recommendations for future interventions and evaluations

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

1. **Explore potential positive interactions between GBIC initiatives**
   For example, the IPE identified potential synergies between the Belonging interventions and the Informal Learning - Code Club intervention. This link was made in one of the *Code Stars* intervention messages to parents. Providing opportunities for pupils to access both types of interventions could enable girls to further explore their new interests sparked by initiatives like the *My Skills My Life* and *Code Stars* interventions through extracurricular computing activities such as code clubs. This could make it easier for the girls to continue research into computing in their free time and would also expose them to prosocial aspects of computing, which were two of the barriers targeted by the interventions.

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

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

3. **Identify strategies to measure outcomes targeted by the intervention further into the future**
   Tracking relevant behavioural outcomes (in this case, subject choice from Year 10

onwards) multiple years after the intervention requires planning, collaboration with schools, longer and more flexible evaluation timelines. However, it would also greatly enhance the ability to evaluate the impact of early interventions over a time horizon in line with the mechanisms and barriers hypothesised, and thus identify the most impactful ones. In this case, attempting to collect and analyse data on whether pupils in the evaluation sample select computer science as a GCSE subject once the choice arises would enable the estimation of the impact of the intervention on the long-term outcomes targeted, in addition to the short-term proxy indicators used in this evaluation.

4. **For any future adaptations or new interventions, consider additional small-scale piloting to refine delivery prior to a full-scale impact evaluation**
   Piloting interventions in school is complicated given the school staff involvement and coordination with schools it requires, particularly as it does not provide the promise of findings on the causal impact of the programme. In the recent COVID-19 context in schools, it would have been even more challenging. However, the possible improvements to the delivery of both interventions identified through the IPE illustrate the value of small-scale piloting to inform improvements to the impact potential of any intervention before moving to a full-scale impact evaluation. Where possible, strategies to evaluate interventions at incremental scale and cost should be explored to maximise learning and resource efficiency.

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

Reference list

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117,* 497–529.

Cassidy, R., Cattan, S., Crawford, C., & Dytham, S. (2018) How can we increase girls' uptake of maths and physics A-level? *The Institute for Fiscal Studies, R149.*

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

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of personality and social psychology, 102*(4), 700.

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

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

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

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

Rubin, D.B. (2004) Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons Inc., New York

Appendices

Appendix 1: Model specification

Primary outcome: SCSAS scores - Belonging Subscale - 4-point mean score

The primary outcome was calculated from averaging pupil scores on the 5 belonging questions of the survey. For the purpose of analysis, we treated it as continuous and used a linear regression to assess the Intention-To-Treat (ITT) effect of our treatment on this outcome. Owing to the clustered nature of the data, and because we randomised at the cluster level, we will use cluster-robust standard errors in analysis and clustering at the school level.

$$Y_{is} = \alpha + \beta_1 T_i + \beta_2 BL_i + \beta_3 propFSM_s + \beta_4 Ofsted_s + \epsilon_{is}$$

Where:

- is the belonging subscale SCSAS mean survey score for pupil *i* in school *s*
- α is the constant
- T_i is a tertiary indicator of treatment assignment for pupil *i in school s*
- BL_i is the baseline belonging SCSAS score for pupil *i* in school *s* collected before the intervention
- $propFSM_s$ is the proportion of pupils eligible for FSM in school *s*
- $Ofsted_s$ is an indicator of Ofsted rating in school *s*, comprising (i) “Outstanding”; (ii) “Good”; (iii) “Below good” (the combination of “Requires improvement” and “Inadequate”) and (iv) “missing ofsted rating”[31]
- ϵ_{is} is the error term for pupil *i* in school *s*

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

The secondary outcome will also be calculated from averaging different scores. For the purpose of analysis, we treat it as continuous and therefore we will use a linear regression to assess the Intention-To-Treat (ITT) effect of our treatment on this outcome. Owing to the clustered nature of the data, we will use cluster-robust standard errors in analysis, clustering at the school level.

---

[31] Not all schools had ofsted ratings given to them.

$$Y_{is} = \alpha + \beta_1 T_i + \beta_2 BL_i + \beta_3 propFSM_s + \beta_4 Ofsted_s + \epsilon_{is}$$

Where:

- Y_{is} is the Total SCSAS survey mean score for pupil *i* in school *s*
- α is the constant
- T_i is a tertiary indicator of treatment assignment for pupil *i in school s*
- BL_i is the baseline SCSAS score for pupil *i* in school *s* collected before the intervention
- $propFSM_s$ is the proportion of pupils eligible for FSM in school *s*
- $Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) “Outstanding”; (ii) “Good”; and (iii) “Below good” (the combination of “Requires improvement” and “Inadequate”) and (iv) “missing ofsted rating”
- ϵ_{is} is the error term for pupil *i* in school *s*

Secondary outcome: stated intention to study computing

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

$$Y_{is} \overset{Indep.}{\sim} bernoulli(p_{is}); logit(p_{is}) = \alpha + \beta_1 T_i + \beta_2 BL_i + \beta_3 propFSM_s + \beta_4 Ofsted_s$$

Where:

- Y_{is} is a binary indicator for pupil *i* reflecting intention to study computing in school *s*
- p_{is} is the probability of a positive intention for pupil *i* in school *s*
- α is the constant
- T_i is a tertiary indicator of treatment assignment for pupil *i in school s*
- BL_i is the baseline SCSAS score for pupil *i* in school *s* collected before the intervention
- $propFSM_s$ is the proportion of pupils eligible for FSM in school *s*
- $Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) “Outstanding”; (ii) “Good”; and (iii) “Below good” (the combination of “Requires improvement” and “Inadequate”) and (iv) “missing ofsted rating”

Exploratory outcome: stated intention to study science, maths or computing

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

$$Y_{is} \overset{Indep.}{\sim} bernoulli(p_{is}); logit(p_{is}) = \alpha + \beta_1 T_i + \beta_2 BL_i + \beta_3 propFSM_s + \beta_4 Ofsted_s$$

Where:

- Y_{is} is a binary indicator for pupil *i* reflecting intention to study at least one of science, maths, or technology in school *s*
- p_{is} is the probability of a positive intention for pupil *i* in school *s*
- α is the constant
- T_i is a tertiary indicator of treatment assignment for pupil *i in school s*
- BL_i is the baseline SCSAS score for pupil *i* in school *s* collected before the intervention
- $propFSM_s$ is the proportion of pupils eligible for FSM in school *s*
- $Ofsted_s$ is a tertiary indicator of Ofsted rating in school *s*, comprising (i) “Outstanding”; (ii) “Good”; and (iii) “Below good” (the combination of “Requires improvement” and “Inadequate”) and (iv) “missing ofsted rating”

Appendix 2: SCSAS subscale scores

Table 10: Baseline and endline SCSAS subscale and overall scores by treatment group for girls who completed both surveys

Subscale	Survey	Group	N (non-missing)	Mean (SD)
Confidence	Baseline	Control	1,046	2.82 (0.58)
	Baseline	WISE Treatment	438	2.83 (0.55)
	Baseline	Code Stars Treatment	604	2.83 (0.63)
	Endline	Control	1,046	2.80 (0.63)
	Endline	WISE Treatment	438	2.78 (0.64)
	Endline	Code Stars Treatment	604	2.82 (0.64)
Interest	Baseline	Control	1,046	2.93 (0.65)
	Baseline	WISE Treatment	444	2.97 (0.60)
	Baseline	Code Stars Treatment	604	2.97 (0.65)
	Endline	Control	1,046	2.93 (0.70)
	Endline	WISE Treatment	444	2.91 (0.66)
	Endline	Code Stars Treatment	604	2.93 (0.70)
Belonging	Baseline	Control	1,046	3.03 (0.51)
	Baseline	WISE Treatment	444	2.93 (0.46)
	Baseline	Code Stars Treatment	605	3.00 (0.51)
	Endline	Control	1,046	3.04 (0.56)
	Endline	WISE Treatment	444	3.03 (0.49)
	Endline	Code Stars Treatment	605	3.03 (0.55)
Usefulness	Baseline	Control	1,039	2.97 (0.54)
	Baseline	WISE Treatment	440	2.98 (0.51)
	Baseline	Code Stars Treatment	591	2.94 (0.56)
	Endline	Control	1,039	3.00 (0.61)
	Endline	WISE Treatment	440	3.10 (0.57)
	Endline	Code Stars Treatment	591	3.00 (0.61)

Encouragement	Baseline	Control	1,039	2.65 (0.64)
	Baseline	WISE Treatment	440	2.59 (0.63)
	Baseline	Code Stars Treatment	591	2.60 (0.65)
	Endline	Control	1,039	2.67 (0.70)
	Endline	WISE Treatment	440	2.65 (0.66)
	Endline	Code Stars Treatment	591	2.67 (0.70)
Overall SCSAS score	Baseline	Control	1,046	2.87 (0.46)
	Baseline	WISE Treatment	444	2.86 (0.42)
	Baseline	Code Stars Treatment	605	2.87 (0.47)
	Endline	Control	1,046	2.89 (0.52)
	Endline	WISE Treatment	444	2.89 (0.48)
	Endline	Code Stars Treatment	605	2.89 (0.52)

Appendix 3: Regression tables with all model specifications[32]

Appendix 3.1: My Skills My Life regression tables

Outcome: Sense of Belonging in Computing (SCSAS belonging Subscale)	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	3.03		3.04
Treatment-group mean (WISE)	3.00		3.02
Estimated treatment effect (standard error)	0.024 (0.031)	-0.028 (0.035)	0.034 (0.031)
N	2,599	2,606	2,095

Outcome: Attitudes towards computing (Total SCSAS score)	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	2.89		2.88
Treatment-group mean (WISE)	2.85		2.89
Estimated treatment effect (standard error)	-0.006 (0.031)	-0.035 (0.037)	0.025 (0.028)
N	2,599	2,606	2,095

[32] Models (1) & (2) for all outcomes have slightly different sample sizes, with the multiple imputation model having slightly a slightly smaller sample. This is likely because two of the variables used to impute (baseline SCSAS scores and baseline SCSAS belonging scores) were missing some values in the data, due to the fact that we did not drop incomplete observations that could provide information for at least one of our outcome measures (e.g. a pupil answered the intention to study stem question, but did not complete the whole survey). For this reason, we believe the multiple imputation model failed to impute on some values, causing the discrepancy between sample sizes.

Outcome: Intention to study computing	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	50%		49%
Treatment-group mean	50%		52%
Estimated treatment effect	0.16pp	-1.1pp	2.2pp
N	2,600	2,616	2,096

Outcome: Intention to study maths, science or technology	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	76%		76%
Treatment-group mean	76%		77%
Estimated treatment effect	-0.14pp	-0.5pp	0.6pp
N	2,614	2,633	2,110

Appendix 3.2: Code Stars regression tables

Outcome: Sense of Belonging in Computing (SCSAS belonging Subscale)	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	3.03		3.04
Treatment-group mean	3.01		3.03
Estimated treatment effect (standard error)	0.007 (0.031)	-0.003 (0.037)	0.003 (0.030)
N	2,599	2,606	2,095

Outcome: Attitudes towards computing (Total SCSAS score)	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	2.89		2.88
Treatment-group mean	2.88		2.89
Estimated treatment effect	0.020 (0.028)	0.003 (0.035)	0.021 (0.027)
N	2,599	2,606	2,095

Outcome: Intention to study computing	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	50%		49%
Treatment-group mean	55%		56%
Estimated treatment effect	5.6pp	4.6pp	6.4pp
N	2,600	2,616	2,096

Outcome: Intention to study maths, science or technology	(1) Multiple Imputation Model	(2) Baseline Missingness Indicator	(3) Complete Case Analysis
Control-group mean	76%		76%
Treatment-group mean	77%		76%
Estimated treatment effect	-0.43pp	-0.08pp	1.4pp
N	2,614	2,633	2,110

Appendix 4: Survey measures

Appendix 4.1: Survey explanation and changes made

This version of the SCSAS survey was updated based on comments by the DfE on 02/09/2021. A detailed overview of the changes made to the previous iteration of the survey is detailed here. While it is possible these changes may have affected the validity of the survey, we believe this is unlikely to be the case. The changes made weren't substantial, and mainly focused on re-wording the questions to be more 'English', rather than American.

The original SCSAS was designed for pupils aged 13+ in the US[33]. For the current trial, the language of the SCSAS was adapted to a UK sample and one is that younger (in Years 4, 5, 6). The aim was to achieve a Flesch-Kincaid (FK) grade level score of 5 or below, which was possible, or all sub-scales except Usefulness and Encouragement (Kincaid et al., 1975)[34].

Note: All references to "computer science" have been changed to "computer" in the version of the SCSAS we are using.

Survey

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

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

[33] Haynie and Packman (2017). Available at: https://csedresearch.org/tool/?id=156.
[34] Calculated here: https://goodcalculators.com/flesch-kincaid-calculator/

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

2.1	I feel happy in computing class	Strongly disagree	Disagree	Agree	Strongly Agree
2.2	I feel like I belong in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.3	I have lots of friends in my computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.4	I know someone who uses computing in their job	Strongly disagree	Disagree	Agree	Strongly Agree
2.5	I have friends who think computing is interesting.	Strongly disagree	Disagree	Agree	Strongly Agree
2.6	I am confident that I can do computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.7	I am confident that I can solve problems by using computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.8	I can learn computing skills without much help	Strongly disagree	Disagree	Agree	Strongly Agree
2.9	I am good at solving hard questions in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.10	I think I will do well in computing	Strongly disagree	Disagree	Agree	Strongly Agree

2.11	I would choose more computing lessons if I could	Strongly disagree	Disagree	Agree	Strongly Agree
2.12	In the future I'd like to do more computing	Strongly disagree	Disagree	Agree	Strongly Agree
2.13	I like to use computing to solve problems	Strongly disagree	Disagree	Agree	Strongly Agree
2.14	Solving questions in computing lessons makes me feel happy	Strongly disagree	Disagree	Agree	Strongly Agree
2.15	I like computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree

Page 3
Well done! Keep going - you are already half way through.

2.16	I feel happy in computing class	Strongly disagree	Disagree	Agree	Strongly Agree
2.17	I feel like I belong in computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.18	I have lots of friends in my computing lessons	Strongly disagree	Disagree	Agree	Strongly Agree
2.19	I know someone who uses computing in their job	Strongly disagree	Disagree	Agree	Strongly Agree
2.20	I have friends who think computing is interesting.	Strongly disagree	Disagree	Agree	Strongly Agree
2.21	Knowing about computing will help me get a job.	Strongly disagree	Disagree	Agree	Strongly Agree

2.22	To get the job I want I will need computing skills.	Strongly disagree	Disagree	Agree	Strongly Agree
2.23	I can use things I learn in computing lessons in other lessons too.	Strongly disagree	Disagree	Agree	Strongly Agree
2.24	I'll need to be good at computing for my lessons as I get older.	Strongly disagree	Disagree	Agree	Strongly Agree
2.25	Computing is an important subject.	Strongly disagree	Disagree	Agree	Strongly Agree

Page 4
Almost done!

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

Page 5
You have completed this survey! Thank you for taking the time to answer this survey.

Appendix 5: *Code Stars* teacher survey

Explanation and motivation

Following completion of the intervention, the BIT evaluation team designed and sent out a short feedback survey to teachers, to try and further understand the following:

- Whether teachers administer the lessons as intended, and whether there were any specific challenges teachers faced with the lessons.
- Whether, how, and how often the conversation prompts were shared with parents, and whether there were any specific challenges associated with this.
- We also asked teachers to provide any feedback they had about the programme more generally.

The rationale behind designing and sending out this survey was that, in contrast with the *My Skills My Life* intervention, we had less information on how the intervention was implemented. The teacher survey was seen as a low cost way to try and obtain any information we could, while placing a low burden on teachers.

The survey

Thank you for participating in *Code Stars*!

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

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

Thanks!

1. Overall, how did you find the project in general? Was there anything in particular you found useful, or that went particularly well?

- [Free text feedback]

2. How many lessons did you give as part of the *Code Stars* project (the main lesson and the follow up lesson count as one lesson each)?
- 0
- 1
- 2

3. What challenges did you face when giving the lessons, if any?

- [Free text feedback]

4. Did you try out any of the tips or activity ideas in the action guide for inclusive teaching?
- Yes, all of them
- Yes, most of them
- Yes, some of them
- No

5. How often did you share the conversation prompts with parents?
- Once a week
- Every two or three weeks
- Once a month
- Less than once a month
- I shared them all at once
- I didn't share the conversation prompts

6. How did you share the conversation prompts?
- I didn't share the conversation prompts
- By email
- I gave hard copies of the prompts directly to parents
- I sent hard copies of the prompts home with students
- By text
- Other (Please specify)

7. What challenges did you face in sending the prompts, if any
- [Free text feedback]

Thank you for your feedback!
This will be very helpful in informing our research