



St. George's, University of London: Science Stars Programme Evaluation

Impact Report 2022/2023

October 2023



Contents

Executive Summary.....	3
Headline findings.....	3
Recommendations for delivery.....	4
1. Introduction.....	5
About the Organisations.....	5
Programme Overview.....	5
2. Methodology.....	6
Outcome Measures.....	6
Evaluation Design.....	6
3. Programme Delivery.....	11
Tutors and schoolteachers' perception of programme.....	11
Pupil perception of programme.....	13
4. Outcomes.....	18
4a. Academic Outcomes.....	18
4b. Non-cognitive Outcomes.....	26
5. Conclusion.....	40

Executive Summary

Science Stars is a tutoring intervention delivered by St. George's, University of London. It aims to improve the science GCSE attainment of Y11 pupils. 2022/23 was **the fourth year of the Science Stars tutoring programme and this evaluation has found positive results.**

This was the first year that the programme was delivered in two different ways; one school delivered the programme remotely, whilst the other school delivered it in-person. It is important to note that **St George's chose to move their delivery online because of the external restrictions imposed by Covid-19, and have already moved back to delivering the programme in-person in both of their partner schools.** However, the difference in impact between in-person and online is still interesting, and one explored in this evaluation.

This evaluation adopted a mixed-methods approach and used quantitative attainment and non-cognitive surveys as well as qualitative interviews and focus groups. Whilst the first two years of evaluation (2019/20 and 2020/21) showed positive impacts of the programme, last year's evaluation (2021/22) showed the beginning of some negative trends. However, the evaluation of this year's programme (2022/23) shows positive results, both in terms of participating pupils' outcomes and teachers', tutors', and pupils' perception of the programme and its impact. **Many of the negative trends from last year have been reversed this year.** Overall, analysis of academic outcomes showed that **participating pupils improved their science GCSE grades more than their peers** in the comparison group. Interestingly, findings from the analysis of the non-cognitive data appear to suggest that the **in-person Science Stars programme had a greater positive impact on improving pupils' non-cognitive outcomes than those receiving the intervention online.**

Headline findings

-  Science Stars pupils' science GCSE attainment improved more than pupils who did not participate in the Science Stars.
-  A greater percentage of Science Stars pupils achieved their target science GCSE grade (61%) than those who did not participate in Science Stars (42%).
-  2022/23 is the first academic year since 2019/2020 where Science Stars participants have improved their metacognition across the course of the programme.
-  Science Stars participants' self-efficacy increased more than those in the comparison group.
-  Participating pupils' test anxiety level decreased by 9%, whilst comparison pupils' test anxiety levels increased by 1%.

6

Communication and collaboration across stakeholders was successful, even if staff changes in schools made this challenging at times.

7

73% of participants agreed or strongly agreed that they highly valued the Science Stars programme, and 72% of participants agreed or strongly agreed that the Science Stars programme will make a big difference to their GCSE grade.

Recommendations for delivery

Based on the findings of this report, we recommend the following:

- ◆ Run an in-person session for tutors before they start delivering the programme.
- ◆ Establish a comprehensive plan with schools and tutors for logistical considerations and successful setup for sessions for in-school delivery to ensure the smooth running of session. This could include, but is not limited to:
 - Ensuring a reliable and consistent WiFi in schools
 - Providing USB sticks to tutors
 - Ensuring school systems can integrate with Science Stars tutor's USBs
 - Whitelisting certain website and applications so tutors can access Science Stars resources in school.
- ◆ Allocate more time for continuous assessment within sessions.
- ◆ Introduce a new way to deliver continuous assessment that will enable tutors to gain greater clarity of pupils' understanding.
- ◆ When teaching pupils exam-style questions, allow tutors to spend more time coaching pupils to apply their knowledge in a way that will maximise their grades.

It is worth noting that due to the current report cycle, which is restricted by when GCSE results are released, the following recommendations can only be implemented in 2024/25.

1. Introduction

About the Organisations

St George's, University of London, is an independent university dedicated to medical and health science education, training and research, affiliated with the University of London. With a strong historical commitment to widening participation activities, St George's is now increasingly working across the whole student lifecycle to support students from under-represented backgrounds. This year, St George's has run the Science Stars programme for a fourth year, focusing specifically on school-based activities to raise attainment.

ImpactEd is a not-for-profit organisation that exists to improve pupil outcomes by addressing the evaluation deficit in education. ImpactEd works in partnership across the education sector to support high-quality monitoring and evaluation that informs decisions about what will work most effectively to support students. Their work in access and widening participation has included evaluation projects with University College London, Goldsmiths University and London South Bank University among others.

Programme Overview

Science Stars is a sustained tutoring intervention designed to support Year 11 students to prepare for GCSEs and ultimately increase their attainment in science. The programme is delivered both remotely and in-person by current students at St George's, University of London – following a pre-designed curriculum developed by a former science teacher.

The programme aims to improve educational outcomes in GCSE Science for target students in Year 11. The key aims and objectives of the programme for participating students are as follows:

- ◆ More able to answer exam questions.
- ◆ Better understanding of science GCSE content.
- ◆ Increased academic attainment.
- ◆ Improved revision skills.
- ◆ Better understanding of their own learning, strengths, and weaknesses.
- ◆ Increased confidence in science.
- ◆ Less anxious about tests and exams.

2. Methodology

This section will cover the outcome measures, the evaluation design for data collection and analysis, as well as the limitations of the approach.

Outcome Measures

The table below shows the key outcomes in this evaluation and how they will be measured using both quantitative and qualitative measures.

Outcome	Quantitative Measure	Qualitative Measure
Improved revision skills	MSLQ Metacognition	
Increased confidence in science	MSLQ Self-efficacy	
Less anxious about tests and exams	MSLQ Test Anxiety	
More able to answer exam questions	School attainment data	
Better understanding of their own learning, strengths, and weaknesses	MSLQ Metacognition	Focus groups with graduate tutors and interviews with teachers
Better understanding of science GCSE content	GCSE grades and school attainment data	
Increased academic attainment	GCSE grades and school attainment data	

Evaluation Design

This evaluation is the fourth annual evaluation of this programme and was conducted in 2022/23. All the data was collected between Autumn Term 2022 and Autumn Term 2023.

As pupil selection was conducted by the school and through a voluntary sign-up process, a randomised control group design was not possible. As such, we used a **matched comparison group** consisting of students from the same school, same year group and similar target grades (where possible) as the Science Stars participants. This group will be referred to as either the comparison or comparator group throughout the report.

Some important caveats for this evaluation design are worth noting:

- ▶ As the comparison group was not randomised, there may be unobservable characteristics affecting performance beyond prior attainment.

- ▶ Particularly when looking at the schools separately, the overall sample size for both participants and the comparator group is small. As such, results may not be immediately generalisable to other school contexts.
- ▶ The programme is operating in very different contexts and two different delivery formats across the two schools so this report will examine the difference in results.

Notwithstanding these points, the design approach allows us to make relatively robust inferences within these constraints by collecting a range of datapoints to triangulate findings and assess if there was a common pattern across indicators.

In this evaluation we analysed three different types of data:

- ▶ **Attainment data** was used to evaluate the impact of the programme on pupil's academic progress,
- ▶ **Pupil survey data** was used to evaluate the impact of the programme on pupils' non-cognitive outcomes,
- ▶ **Qualitative research and delivery data** was used to evaluate the success of the implementation of the programme.

Attainment data: Design and Sample

The table below shows what attainment data was collected, when it was collected, whose attainment data was collected, as well as the sample size.

Data	When?	Which pupils?	Matched Sample Size	
			Ernest Bevin	Hartsdown Academy
Autumn Mock exam	Autumn Term 2022	Participating	17	16
		Comparator	17	16
Spring Mock exam	Spring Term 2023	Participating	17	16
		Comparator	17	16
Final GCSE results	September 2023	Participating	17	16
		Comparator	17	16

Survey: Design and Sample

The non-cognitive outcomes (self-efficacy, test anxiety and metacognition) were measured because they have predictive validity i.e., they have been shown to be with associated improvements in long-term outcomes such as well-being, academic achievement, and employment destinations. Alongside academic achievement, there is evidence that these skills can be particularly important in closing disadvantage gaps.

These non-cognitive outcomes were measured using psychometrically validated questionnaires, administered to pupils pre and post Science Stars. The evaluation followed a pre-post-test design. Pupils were assessed at the beginning (baseline collection) and end (final collection) of the programme. Collecting data at these two time points allows us to analyse the level of change over the course of the programme for each specific outcome.

Our core outcome measures for this evaluation were:

Outcome	Measurement Details
Metacognition	Metacognition means 'thinking about thinking': pupils' ability to think explicitly about their own learning. It is strongly associated with academic progress and improves other skills required for learning, such as critical thinking (Flavell, 1979; Higgins et al., 2016). We measured metacognition using the Cognitive Strategies Use and Self-Regulation subscales of the Motivated Strategies for Learning Questionnaire.
Self-efficacy	Self-efficacy is a measure of pupils' belief in their ability to achieve a specific task in the future. Self-efficacy is correlated with higher academic achievement and persistence, and also contributes to pupil wellbeing (Gutman & Schoon 2013, DeWitz et. al. 2009). We measured self-efficacy using the Self-efficacy subscale of the Motivated Strategies for Learning Questionnaire.
Test anxiety	Test anxiety is concerned with pupils' emotional responses to tests (Pintrich and De Groot, 1990). Greater levels of test anxiety can result in worse performance in exams but in some situations may be linked to increased motivation.

The results of the psychometrically validated survey will be supplemented by qualitative data that has been drawn out by the two focus groups with Science Stars tutors and two one-to-one interviews with the group assistants, one from Hartsdown Academy and the other from Ernest Bevin College.

The table below summarises what surveys that were completed, at which timepoints, who responded, and the sample size of respondents.

Data	When?	Which pupils?	Matched Sample Size	
			Ernest Bevin	Hartsdown Academy
Meta-cognition baseline	Autumn Term 2022	Participating	9	12
		Comparator	11	10
Meta-cognition endline	Spring Term 2023	Participating	9	12
		Comparator	11	10
Self-efficacy baseline	Autumn Term 2022	Participating	12	12
		Comparator	14	12
Self-efficacy endline	Spring Term 2023	Participating	12	12
		Comparator	14	12
Test anxiety pre-survey	Autumn Term 2022	Participating	12	12
		Comparator	14	12
Test anxiety post-survey	Spring Term 2023	Participating	12	12
		Comparator	14	12
Custom questions	Spring Term 2023	Participating	17	16

Custom Questions

This academic year (2022/23), pupils participating in the Science Stars programme were asked four questions around their thoughts and feelings on the programme. They were provided with four statements, and they had to rate them from a scale of 1 to 5, with 1 = strongly disagree, 2 = disagree, 3 = neither disagree nor agree, 4 = agree, and 5 = strongly agree.

When looking at the results in the programme evaluation section, it is worth noting that for some questions, there were a few pupils who did not provide any response at all.

Qualitative Research: Design, Sample and Analysis

Focus groups were conducted with tutors leading sessions at both schools. 3 tutors participated in the focus group for Ernest Bevin, whilst 2 tutors participated in the focus group for Hartsdown Academy. 1:1 interviews were conducted with the relevant schoolteacher in each of the schools.

The qualitative data was analysed using a deductive thematic approach, meaning that we systematically 'code' the data to find common themes and present these, drawing on examples where appropriate.

National Benchmarks for non-cognitive outcomes

Benchmarks used for non-cognitive comparisons use data from the ImpactEd platform.

Baseline benchmarks come from all surveys taken by Year 11 pupils across Autumn Term 2022, as this was when Science Stars participants completed their baseline surveys. Each survey baseline benchmark has the following sample sizes:

- ▶ Metacognition (n = 2663)
- ▶ Self-efficacy (n = 925)
- ▶ Test Anxiety (n = 967)

Endline benchmarks come from all surveys taken by Year 11 pupils across Summer Term 2023, as this was the closest period to when Science Stars participants completed their endline surveys. Each survey endline benchmark has the following sample sizes:

- ▶ Metacognition (n = 288)
- ▶ Self-efficacy (n = 289)
- ▶ Test Anxiety (n = 832)

Analysis Terminology: Percentage vs Percentage Points

Throughout the report, the terms percentage vs percentage points will be used. Please note the difference between the meanings in the definitions below:

- ▶ **Percentage change** quantifies the change we observed as a proportion of the value we started from.
- ▶ **Percentage point change**, on the other hand, quantifies the change we observed in absolute terms (i.e. not relative to the starting point). For example, if 50% of pupils answer 'yes' to a certain question in our baseline survey, but then, later on, 55% of pupils answer 'yes' to that same question in our endline survey, this is a change of **5 percentage points** but a change of 10% (since the difference, 5, is 10% of the starting value, 50).

3. Programme Delivery

This section will be split into two sections. The first will focus on tutors and schoolteachers' perception of the delivery of the programme, whilst the second will focus on participating pupils' perception of the programme through the answers to custom survey questions.

Tutors and schoolteachers' perception of programme

Key finding: Approach to pupil recruitment varied across the two schools which may have impacted outcomes.

The fact that the two participating schools had different approaches to recruiting pupils for the programme may have been a factor in varying pupil outcomes across the two schools. Hartsdown Academy allowed pupils to choose to participate in the intervention, but the **two top sets were encouraged to participate**. Ernest Bevin selected **pupils who were disadvantaged**, recipients of the Pupil Premium or on the edge of grade boundaries. This may be a reason that participating pupils from Hartsdown Academy have a higher average grade in their final GCSE results in comparison to the participating pupils at Ernest Bevin, even though Ernest Bevin participating pupils showed greater percentage improvement than Hartsdown participating pupils.

Key finding: Tutors were motivated by the desire to give back to their local community and found training sessions beneficial.

Tutors chose to get be part of the programme for a variety of reasons. Some had been inspired by friends tutoring in similar programmes, other wants to develop their teaching skills or to understand whether they enjoyed teaching. Many tutors stated that participated because they wanted **to give back to the community**.

Tutors were generally positive about the pre-programme support received through Science Stars. They were happy with both the pre-programme training sessions and the coaching call and reported that they had applied the skills they had acquired from the first session. Within the training, tutors highlighted that the following elements had been particularly beneficial: going through different techniques, how to respond to different situations in the classroom, and going through exam question approaches. Tutors were also satisfied with the resources provided; they stated that they contained good content, were well structured and were easily accessible.

Tutors explained a few ways that they could have been better supported. From a logistical perspective, they noted that it was challenging to set-up the coaching calls and it may have helped to include the person delivering the coaching calls in the WhatsApp group. Some tutors stated they would have preferred to receive the coaching in-person; this was not every tutor's opinion. Tutors would have liked to **have a 'taster session' before they went into schools just by themselves**. In a similar vein, tutors expressed that they would have

benefitted from some feedback from someone who had seen them teach in-person to know if they were delivering sessions well.

Key finding: Communication and collaboration across stakeholder was successful, even if staff changes in schools made this challenging at times.

All stakeholders were positive about **communicating** with each other. Tutors noted that this was well-facilitated by the WhatsApp group chat as well as Science Stars staff being so responsive.

A challenge that was raised by tutors was that there were **some staff changes** at schools which meant that communicating with school staff was difficult because tutors did not always have the most up to date contacts. Another concern tutors raised **around group assistants is that did not seem to always know the pupils** which might be linked to the staff changes.

Tutors were generally positive about the group assistants, and they have enjoyed working with them. **Tutors delivering sessions online were particularly positive** – with one tutor saying that **Group Assistants played a “crucial role”** as they helped handling technical issues.

Key finding: Inconsistent attendance was in issue for both remote and in-person sessions.

The group assistant where the school was delivering the programme in-person reported that the Science Stars’ start time session starting at 3:45 (45 minutes after the end of the school day) caused **logistical and attendance issues**. They reported that this would change for next academic year (23/24).

Both tutors and group assistants named **non-attendance** as a recurring issue. This was attributed to various factors ranging from logistical issues and pupils’ homelife. This issue occurred both across the online and in-person sessions.

Pupils’ **abilities** within a classroom were named as an impeding factor by tutors and teachers but for different reasons. Some tutors reported that having **varied abilities within a group made teaching sessions challenging**. Group assistants, on the other hand, reported that they were **planning to mix pupils with different abilities** within a class as they thought this would **improve overall achievement**.

One teacher mentioned that next year they would implement **seating plans** to prevent disruption caused by friends sitting next to each other.

Key finding: Tutors flagged that a range of technical and pedagogical issues hindered delivery of sessions.

Tutors delivering sessions in-person repeatedly mentioned that a barrier to successfully delivering the session was the school's **Wi-Fi** not being good enough to deliver the sessions.

Tutors also mentioned that another barrier to successful delivery **was pupils not being able to access the relevant resources**. They also flagged session timings being tight was another obstacle for successful session delivery. Tutors flagged that a consequence of tight timings was **no time for continuous** assessment which they thought was important for the programme's success. Tutors delivering the programme in-person reported that they had not found the Socratic quizzes to be useful; they thought they were time consuming and that its results presented pupils' strength and weakness as incredibly spread, meaning that tutors were not able to steer their teaching in one specific direction as is the intention of using a Socratic quiz. Tutors delivering the session online, conversely, thought the Socratic quizzes had helped them direct their teaching in the most relevant direction.

Group assistants expressed that they found sharing the data with Science Stars easy and not time-consuming. One teacher reported that one element they had found challenging was collecting data from the comparator group as there was not a natural time to do it as is the case with participating pupils.

Pupil perception of programme

Key finding: 73% of participants agreed or strongly agreed that they highly valued the Science Stars programme.

Most Science Stars participants (72.73%) agreed or strongly agreed that they highly valued the Science Stars programme. A small proportion of participants (9.09%) neither agreed nor disagreed with the statement that they highly valued the programme. An even smaller proportion (3.03%) disagreed or strongly disagreed that they highly valued the programme. This trend was the same for individual schools.

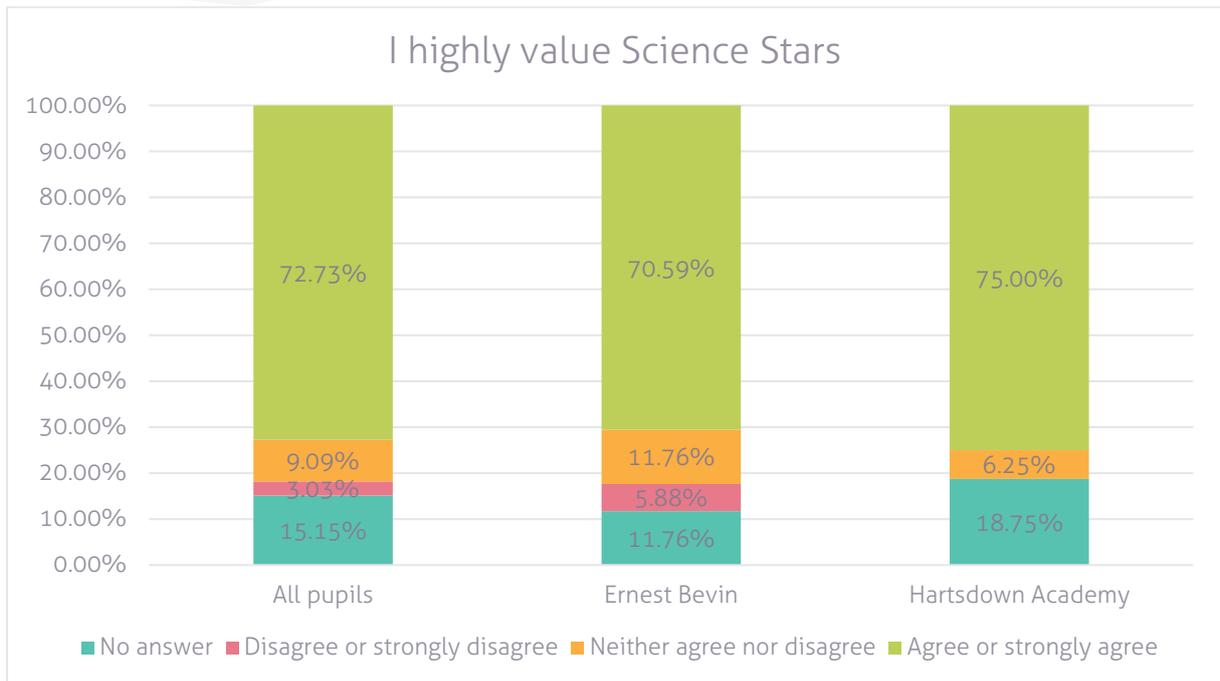


Figure 1 – All pupils: n = 28; Ernest Bevin: n = 15; Hartsdown Academy: n = 13

Key finding: 75.75% of participants agreed or strongly agreed that they were very engaged in the Science Stars tutoring session.

Most Science Stars participants (75.75%) agreed or strongly agreed that they were very engaged in the Science Stars sessions. A small proportion of participants (6.06%) neither agreed nor disagreed with the statement that they were very engaged in the Science Stars programme. An even smaller proportion (3.03%) disagreed or strongly disagreed that they were very engaged in the Science Stars sessions. Looking at schools at an individual level, the trend is broadly the same.

It is worth highlighting that at Ernest Bevin 82.35% agreed or strongly agreed that they were very engaged in the programme, whilst only 68.75% of participating pupils at Hartsdown Academy agreed or strongly agreed with the statement.

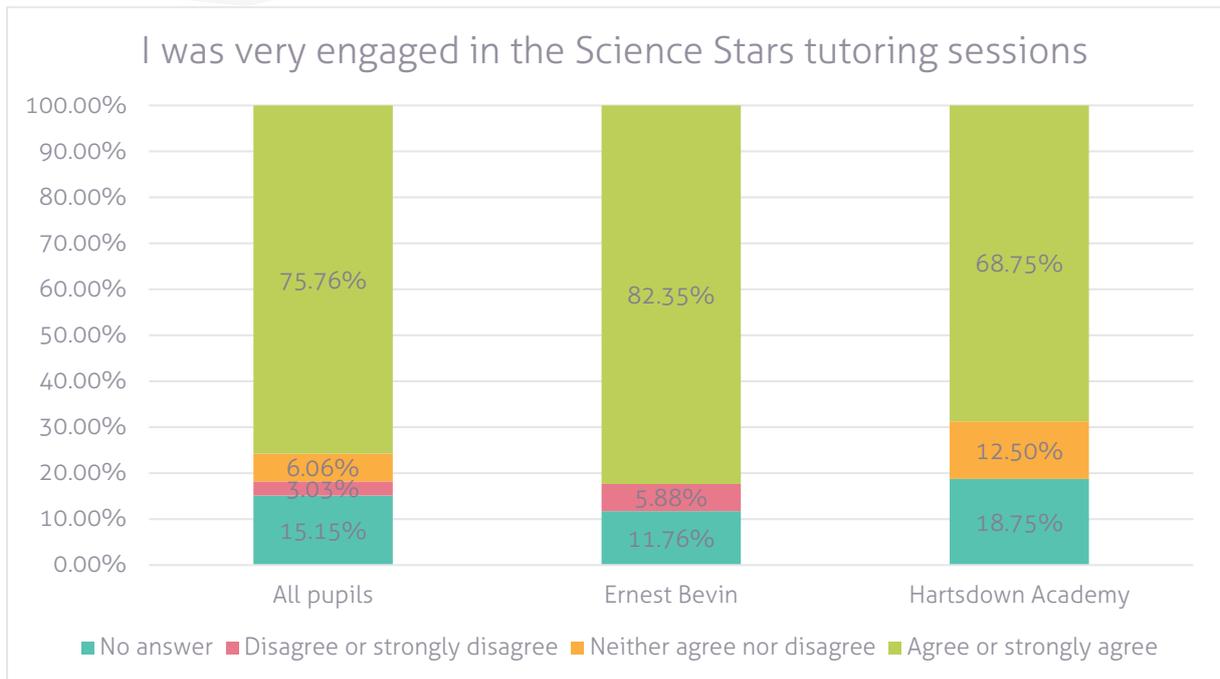


Figure 2 - All pupils: n = 28; Ernest Bevin: n = 15; Hartsdown Academy: n = 13

Key finding: 72.73% of participating pupils agreed or strongly agreed that the Science Stars programme had greatly motivated them to study outside of lessons.

Most Science Stars participants (72.73%) agreed or strongly agreed that the Science Stars programme greatly motivated them to study outside of lessons. A small proportion of participants (9.09%) neither agreed nor disagreed with the statement that the programme greatly motivated them to study outside of lessons. An even smaller proportion (3.03%) disagreed or strongly disagreed that the Science Stars programme greatly motivated them to study outside of lessons. Looking at schools at an individual level, the trend is broadly the same.

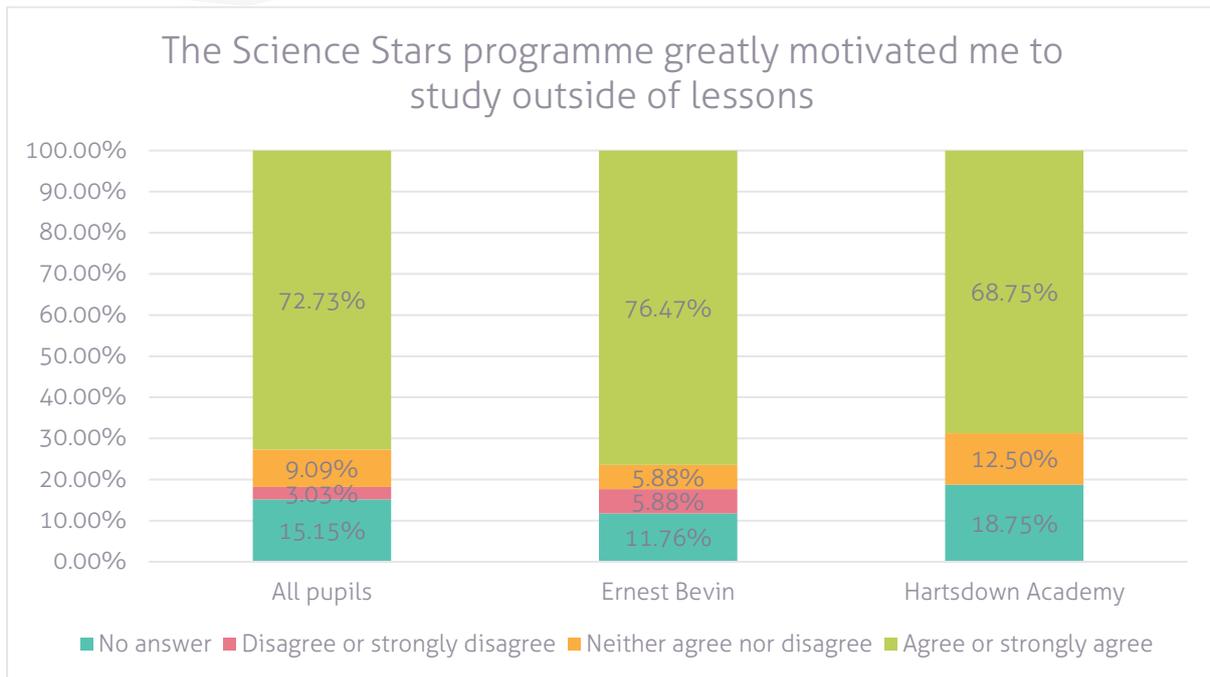


Figure 3 - All pupils: n = 28; Ernest Bevin: n = 15; Hartsdown Academy: n = 13

Key finding: 72.73% of participants agreed or strongly agreed that the Science Stars programme will make a big difference to their GCSE grade.

Most Science Stars participants (72.73%) agreed or strongly agreed that they thought that the Science Stars programme will make a big difference to their GCSE grades. A small proportion of participants (12.02%) neither agreed nor disagreed with the statement that the programme would make a big difference to their grades. No pupils (0%) disagreed or strongly disagreed that the Science Stars programme will make a big difference to their GCSE grade.

Looking at schools at an individual level, the trend is broadly the same but worth highlighting that a greater proportion of Ernest Bevin’s participating pupils (17.65%) neither agreed nor disagreed than those participating at Hartsdown Academy (6.25%).

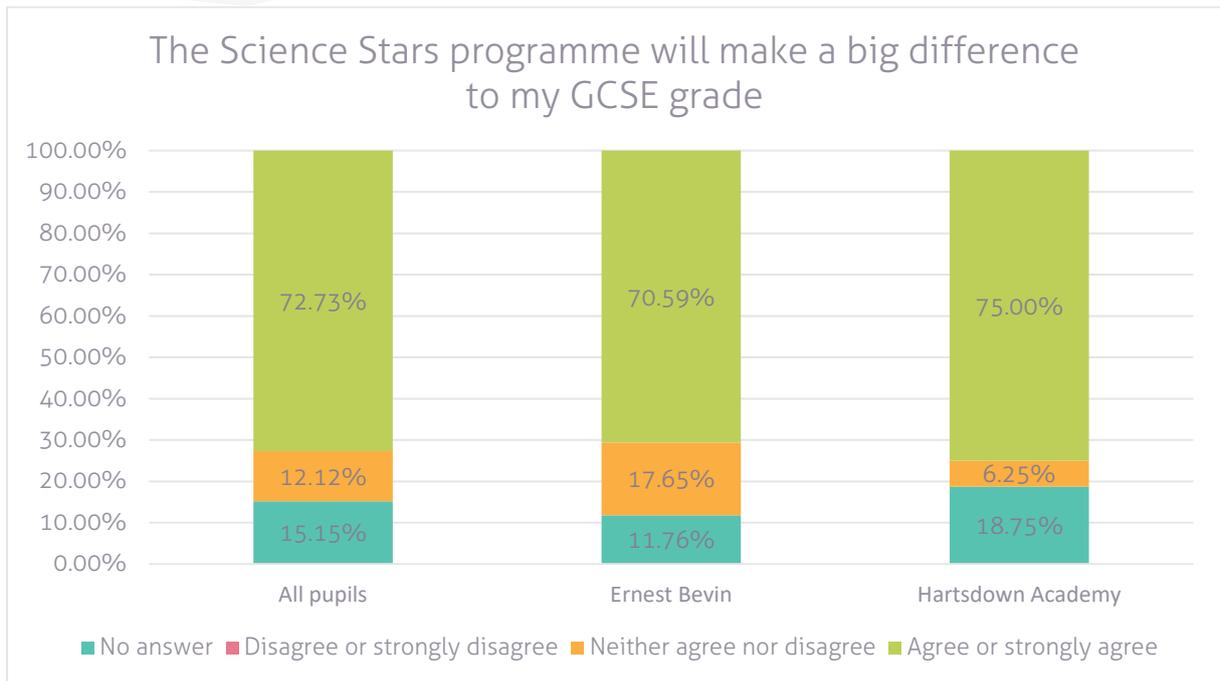


Figure 4 - All pupils: n = 28; Ernest Bevin: n = 15; Hartsdown Academy: n = 13

4. Outcomes

4a. Academic Outcomes

Key finding: Tutors and teachers thought that pupils are more able to answer exam questions.

Many tutors and teachers reported that the Science Stars programme's focus on making pupils cover and practise more exam questions had been useful and had made pupils more confident in answering exam questions. Tutors said that they had been able to teach pupils how to approach different exam questions and that this had worked well, with pupils starting to use keywords in answers.

Teachers noted that the programme's focus on helping pupils understand how to structure their answers had also been beneficial and pupils writing up model answers on the board had been a great way to help pupils improve their ability to answer exam questions. Some tutors, however, wondered whether pupils would be capable of applying their knowledge in a way that examiners wanted.

Key finding: Tutors and teachers thought that pupils had a better understanding of Science GCSE content.

Both tutors and teachers were positive that Science Stars had increased pupils' understanding of Science GCSE. They noted that pupils displayed more familiarity and comfort talking about concepts and using key words, and that their answers to questions also demonstrated better understanding.

Tutors reported that pupils explicitly stated they understood concepts at the end of session and were good at completing recall activities. Teachers also reported pupils having better understanding of Science GCSE following doing some analysis on the positive difference in pupils' score between last year and this year's paper. Teachers mentioned that pupils were better at evaluating and analysing as well as identifying key commands and knowing what calculations to use in questions.

Both tutors and teachers theorised as to which activities helped pupils to gain a better understanding of the Science GCSE. Tutors thought that pupils completing past papers contributed to their improved understanding, whilst Teachers thought that going through model answers helped pupils understand how to provide the key details of an answer as well as having the right structure to an answer. Both tutors and teachers noted that one particularly effective method for improving pupils' understanding was group activities where pupils collaboratively wrote a model answer on the board, where for every step, a new pupil wrote it up on the board.

Some tutors did express some uncertainty as to whether pupils truly understood the concepts they were being taught.

Key finding: Science Stars pupils’ attainment improved more than pupils who did not participate in the Science Stars

Across every measured time interval, Science Stars participants showed a more positive change in results than those who did not participate in Science Stars.

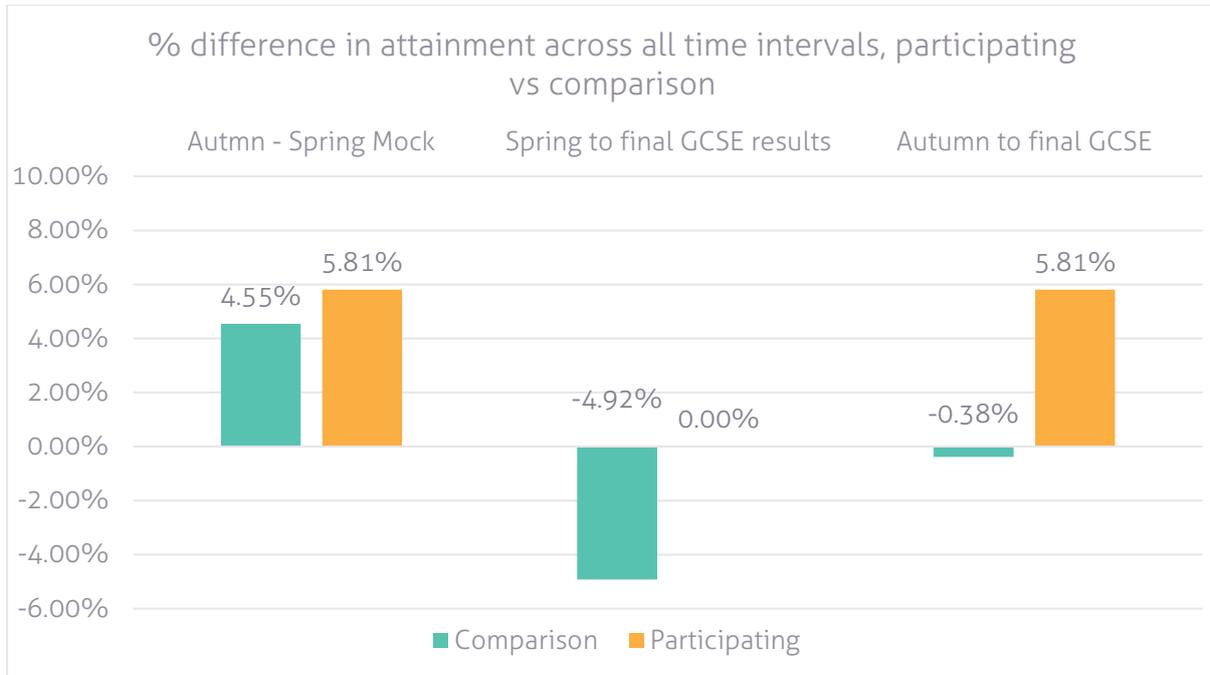


Figure 5 - Comparison: n = 33; Participating: n = 33

Key finding: In 2022/23 Science Stars participants, on average, improved at least one third of a grade, across their Science examinations compared to students in the programme comparison group.

Firstly, it is important to state that the equivalent to one third of a grade is 0.33. The graph below shows that, on average, comparison pupils’ grades **decreased by 0.03** across the year, whilst, on average, Science Stars’ pupils’ grades **increased by 0.46**. Therefore, on average, Science Stars pupils’ grades improved by 0.49 more than their comparison peers, which is more than 0.33, meaning that Science Stars pupils did improve by at least one third of a grade than the comparison group.

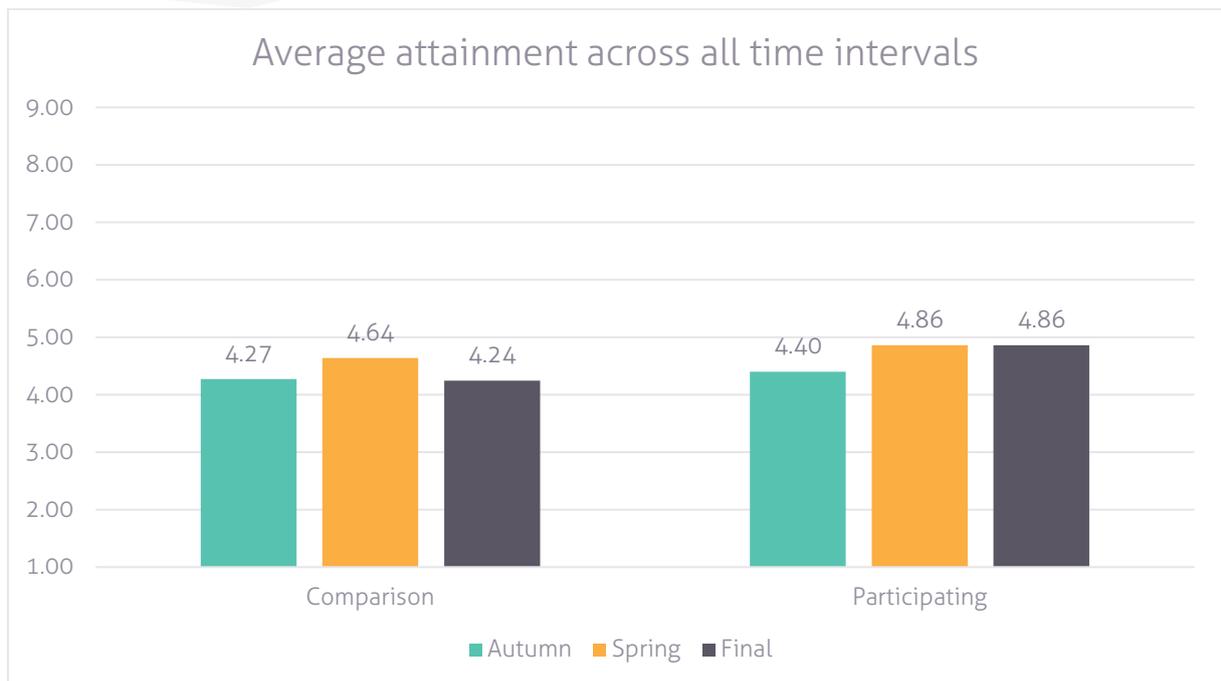


Figure 6 - Comparison: $n = 33$; Participating: $n = 33$

Between the Autumn and Spring mock, the mean percentage increase in grades of comparator pupils was 4.55% whilst the mean percentage increase of participating pupils was slightly higher at 5.81%. Both increases were statistically significant; the increase within comparator pupils had a p value 0.0436 ($n=33$) and the increase within the participating pupils had a p value of 0.0024 ($n=33$).

A difference-in-difference analysis between comparator and participating pupils reveals, however, that the difference in increase between the two groups of pupils is not statistically significant ($p = 0.84$, $n=66$). This general trend was reflected in both schools.

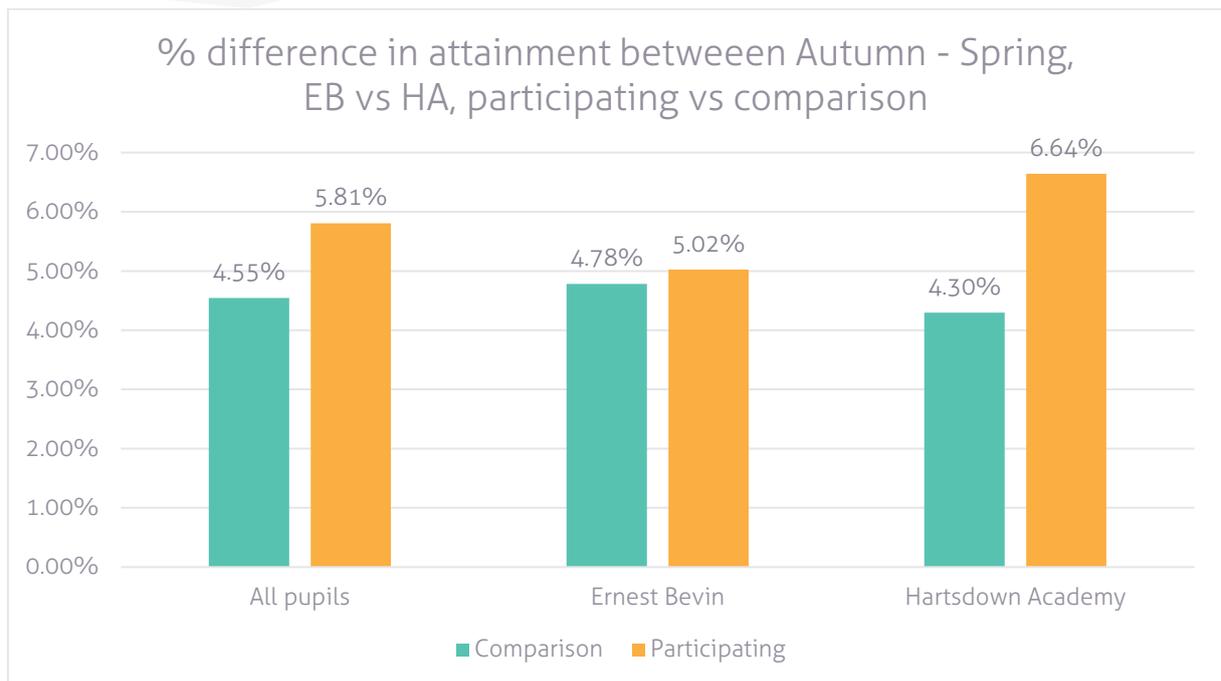


Figure 7 - Comparison - All pupils: n = 33, Ernest Bevin: n = 17, Hartsdown Academy: n = 16; Participating - All pupils: n = 33, Ernest Bevin: n = 17, Hartsdown Academy: n = 16.

Between the Spring mock and final GCSE grades, the mean grades across all comparator pupils decreased by 4.92% whilst for participating pupils' mean grade stayed the same. Neither of these changes were statistically significant and the difference-in-difference analysis also revealed no statistical significance between the two.

These results, however, obscure the difference in trends between the schools. At Ernest Bevin, the mean grades of comparator pupils decreased by 10.66% whilst the mean grades of participating pupils only decreased by 0.74%. At Hartsdown Academy, however, the mean grades of comparator pupils increase by 1.17% whilst participating pupils' grades increased by 0.78%.

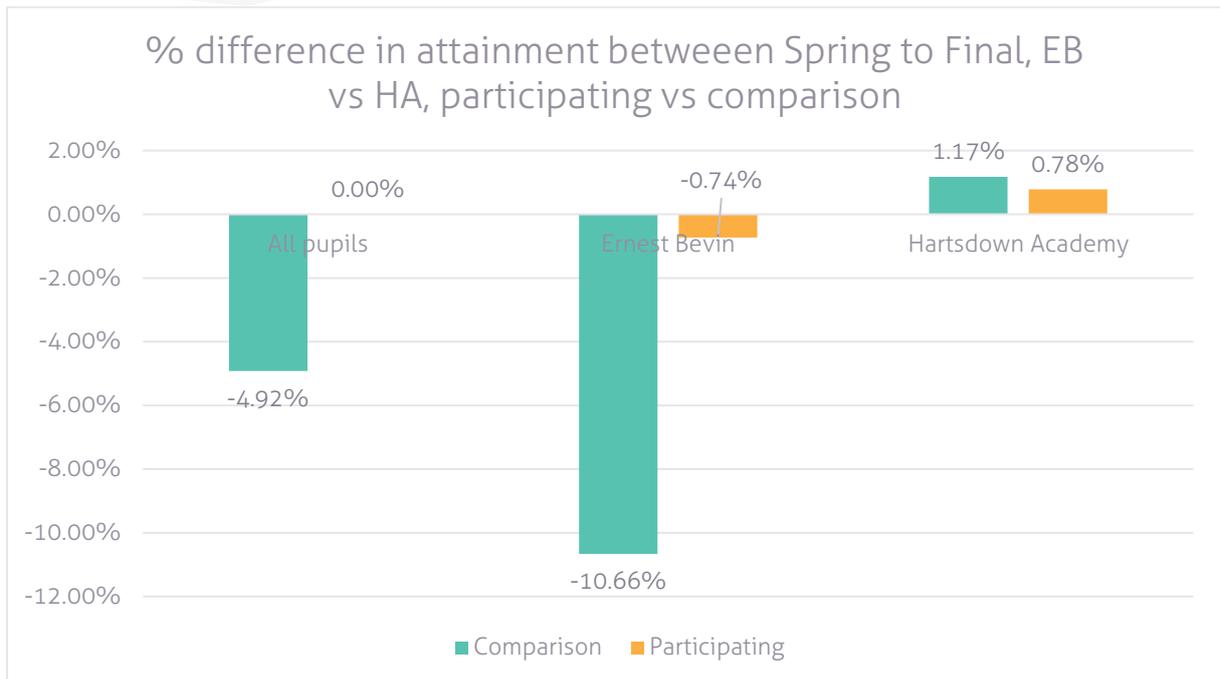


Figure 8 - Comparison - All pupils: $n = 33$, Ernest Bevin: $n = 17$, Hartsdown Academy: $n = 16$; Participating - All pupils: $n = 33$, Ernest Bevin: $n = 17$, Hartsdown Academy: $n = 16$.

Between the Autumn mock and final GCSE grades, the mean grades across comparator pupils decreased by 0.38% whilst the mean grades across participating pupils increased by 5.38%. The comparator pupils' percentage decrease in grades was not statistically significant ($p = 0.911$, $n=33$) whilst the participating pupils' percentage increase in grades was statistically significant ($p = 0.0005$, $n=33$). However, a difference-in-difference (DiD) analysis demonstrates that the difference between the two groups was not statistically significant ($p > 0.05$, $n=66$).

These overall trends, however, also obscures the difference in the two schools. At Ernest Bevin, the mean grades of comparator pupils decreased by 5.88%, whilst the mean grades of participating pupils increased by 4.29%. At Hartsdown Academy, the difference between the percentage difference in grades between comparator and participating pupils was less marked; comparator pupils' mean grades increased by 5.47% and participating pupils increased by 7.42%. When conducting DiD analysis on individual schools, a different story around statistical significance appears. The DiD analysis demonstrates that the difference between the two groups at Hartsdown Academy was not statistically significant ($p > 0.05$, $n = 32$). Whilst the DiD analysis demonstrates that the difference between the two groups at Ernest Bevin was statistically significant ($p = 0.007$, $n = 34$).

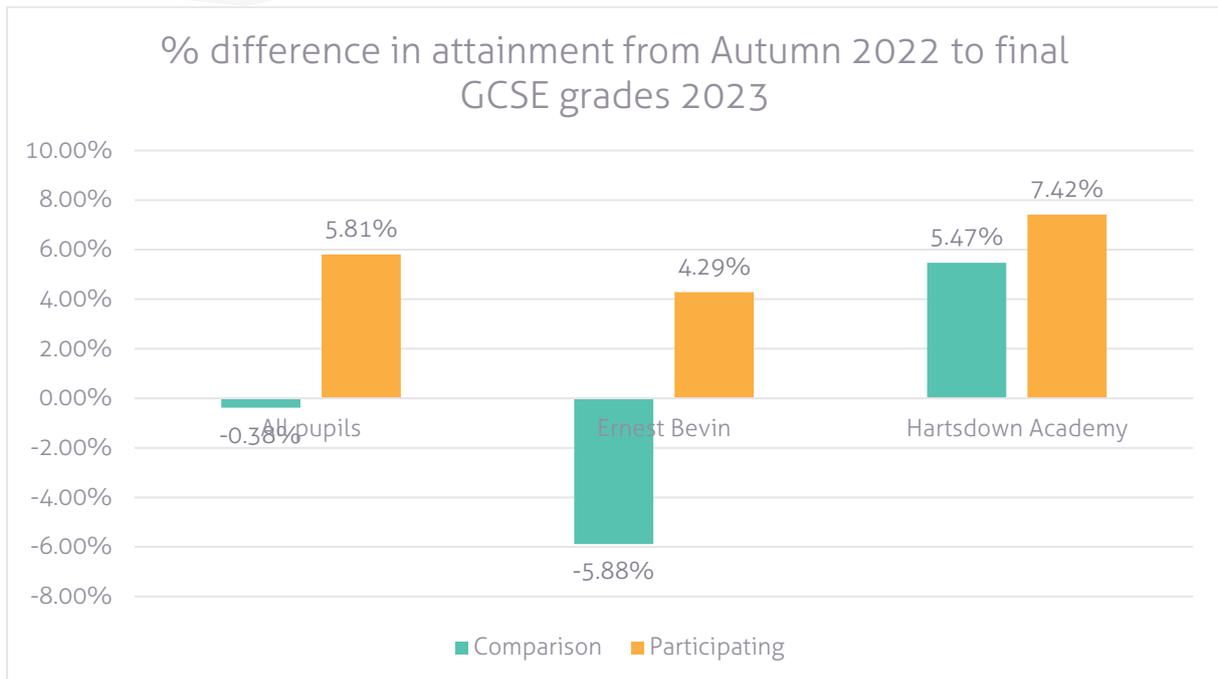


Figure 9 - Comparison - All pupils: n = 33, Ernest Bevin: n = 17, Hartsdown Academy: n = 16; Participating - All pupils: n = 33, Ernest Bevin: n = 17, Hartsdown Academy: n = 16.

Key finding: Since last year, there is a positive trend in Science Stars pupils improving their grades more than comparison pupils

The graph below shows the percentage change in grades of the participant and comparator group pupils for the past four years. From 2019 to 2022, for both participating and comparison pupils, the percentage change increase in their overall grade from Autumn mock to GCSE result stayed the same or increased from year to year. From 2021/22 to 2022/23, however, the percentage change in overall grade decreased for both participating and comparison pupils.

This is seemingly a downward turn for this academic year (2022/23), however, is not a fair representation of the impact that the Science Stars programme seemed to have had on participating pupils this year.

The graph below, illustrating the *difference* in improvement between participating and comparison pupils, shows that in the programme’s first two years (2019/2020 and 2020/21), participating pupils improved their grades more than comparison pupils. However, in the programme’s third year (2021/22), comparison pupils improved their science grades more than their Science Stars counterparts. This year, however, the positive trend of participating pupils progressing more than comparator pupils has returned.

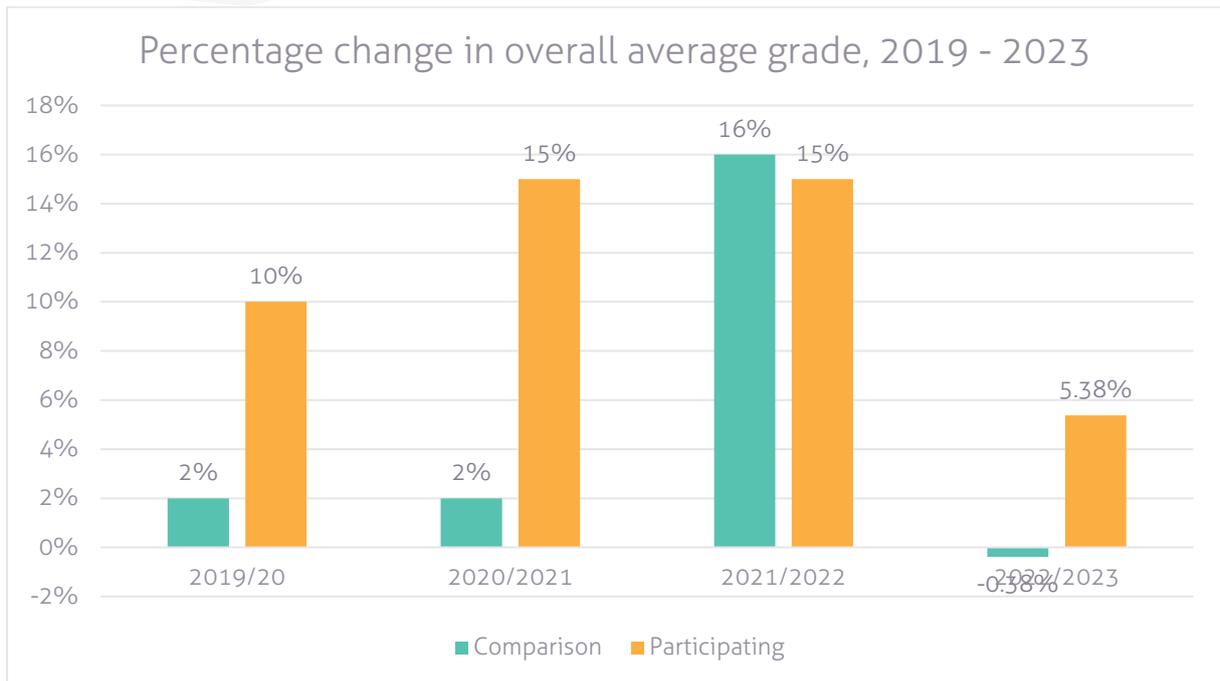


Figure 10 – 2022/23 Control: n = 33; 2022/23 Participating: n = 33

The chart below shows that this year (2022/23) the narrative of comparison pupils improving more than their Science Stars counterparts was reversed. **Science Stars** participants improved their grades more than comparison pupils, which is a positive narrative to draw out of the attainment data.

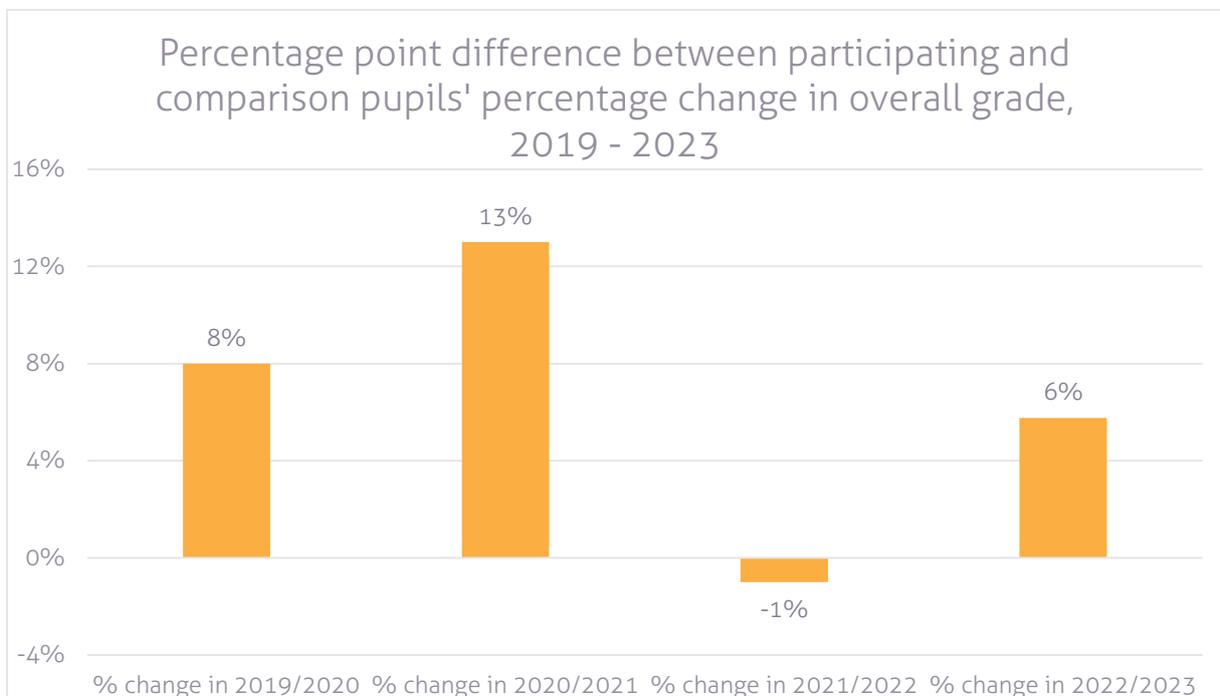


Figure 11 – 2022/23 Control: n = 33; 2022/23 Participating: n = 33

Key finding: A greater percentage of Science Stars pupils achieved their target grade (61%) than those who did not participate in Science Stars (42%).

60.61% of Science Stars participants achieved their target grade whilst on 42.42% of comparator pupils did the same. This trend was broadly the same for both schools; a greater percentage of participating pupils achieved their target grade than their comparator pupils.

At Hartsdown Academy, 81.25% of participants achieved their target grade, whilst only 56.25% of comparator pupils achieved this; this is a difference of 25 percentage points. At Ernest Bevin, 41.18% of participating pupils achieved their target grade whilst only 29.41% of comparator pupils achieved this; this is a difference of 11.77 percentage points.

A larger percentage of participating pupils achieving their target grade in comparison to the participating pupils at Ernest Bevin may have been influenced by Hartsdown Academy’s decision to recruit participants from their top sets.

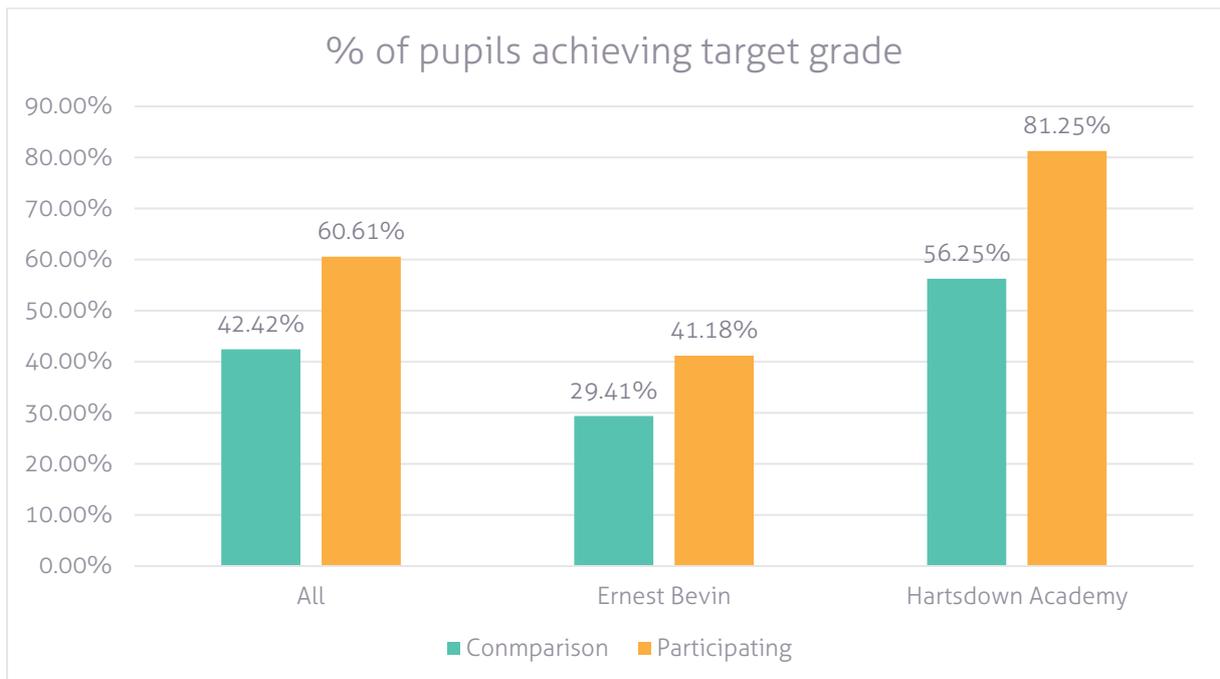


Figure 12 - Comparison - All pupils: n = 33, Ernest Bevin: n = 17, Hartsdown Academy: n = 16; Participating - All pupils: n = 33, Ernest Bevin: n = 17, Hartsdown Academy: n = 16.

Key finding: The gap between Science Stars participants and the comparison group achieving their target grades is increasing in participants’ favour.

The graph below illustrates how the impact of the programme on pupils achieving their target grades from 2020/21 until 2022/23. The general trend since 2020/21 is that a greater percentage of Science Stars participants achieved their target grade than those who were in the comparison group.

Last year (2021/22) the percentage of comparison pupils achieving their target grade was behind by a small gap of 3 percentage points. This year, however, the percentage of Science Stars participants achieving their target grade was a much larger 18.19% percentage points more than those in the comparison group. This is a positive improvement in the difference between comparator and participating pupils from last year.

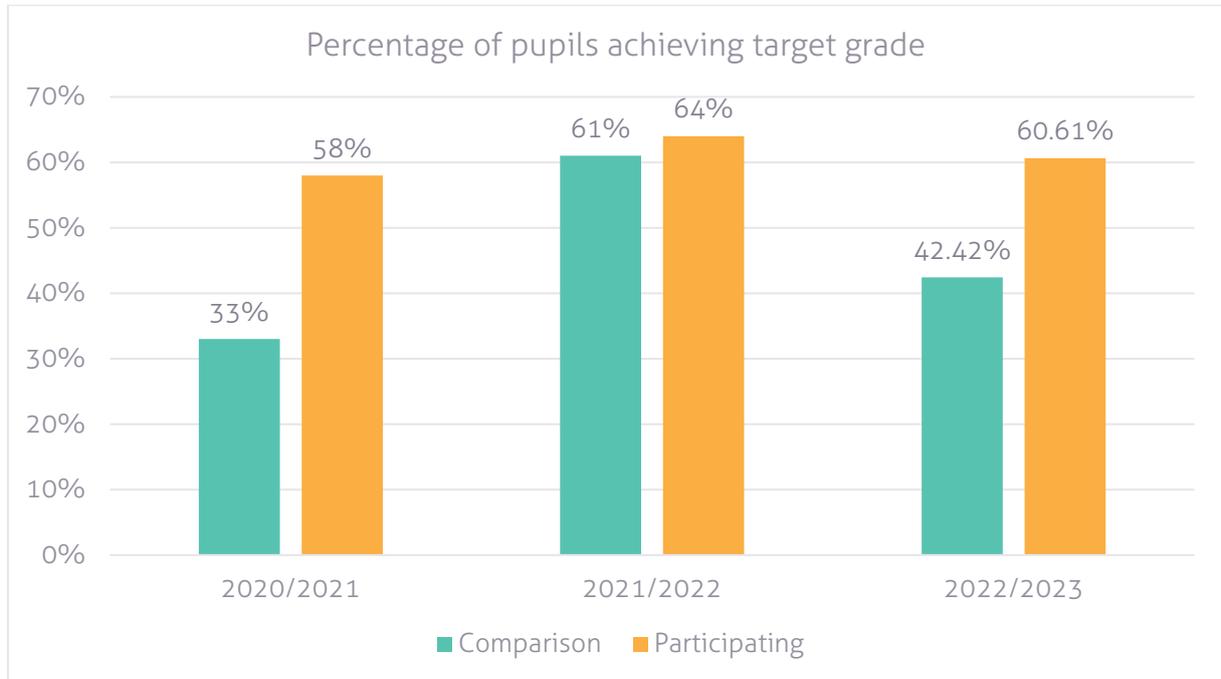


Figure 13 – Comparison: n = 33; Participating: n = 33

4b. Non-cognitive Outcomes

Key finding: Metacognition was the only non-cognitive outcome where comparison pupils made greater improvement than Science Stars participants.

The graph below shows that Science Stars participants saw their metacognition levels increase by 2.04% which was a smaller increase than comparator pupils whose metacognition levels increased by 6.58% between pre and post programme survey. This difference, however, was not statistically significant ($p = 0.61$, $n = 42$.)

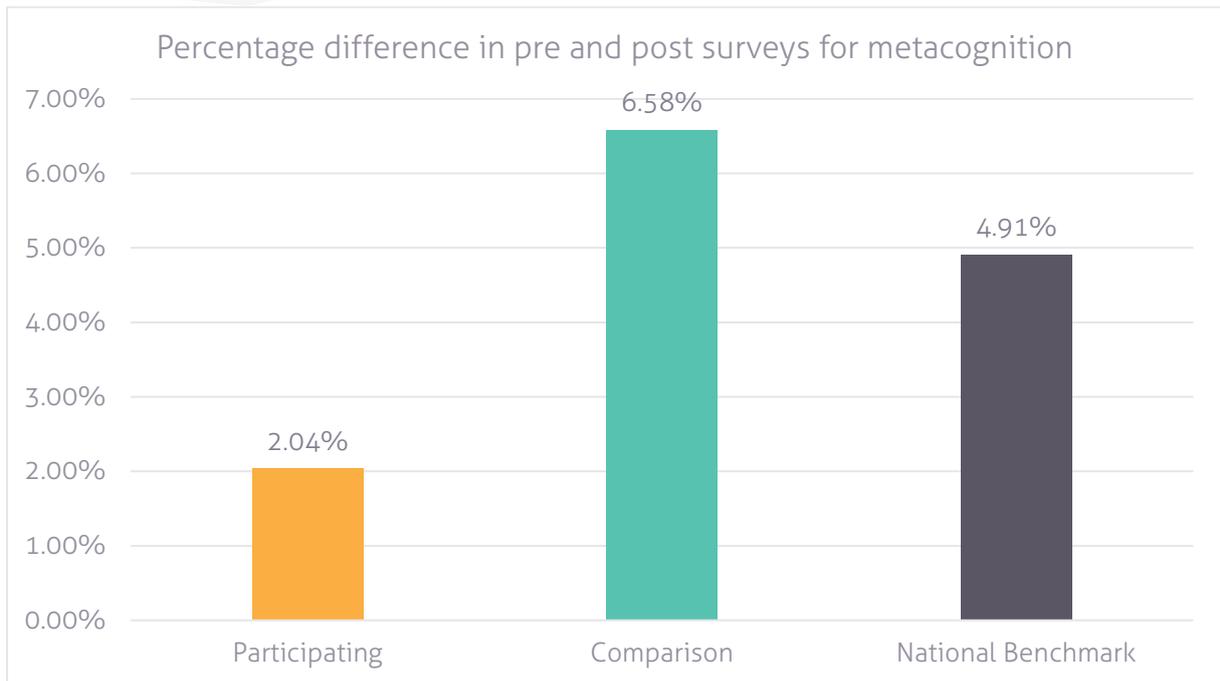


Figure 14 – Participants: n = 21; Comparison n = 21; National Benchmark – Baseline: n = 2663, Final: n = 288.

When looking at the score averages for metacognition between baseline and endline, however, although participating pupils made less progress, they still ended up with a higher metacognition score than their comparison counterparts. Both sets of pupils at both baseline and endline fared better than the national benchmark.

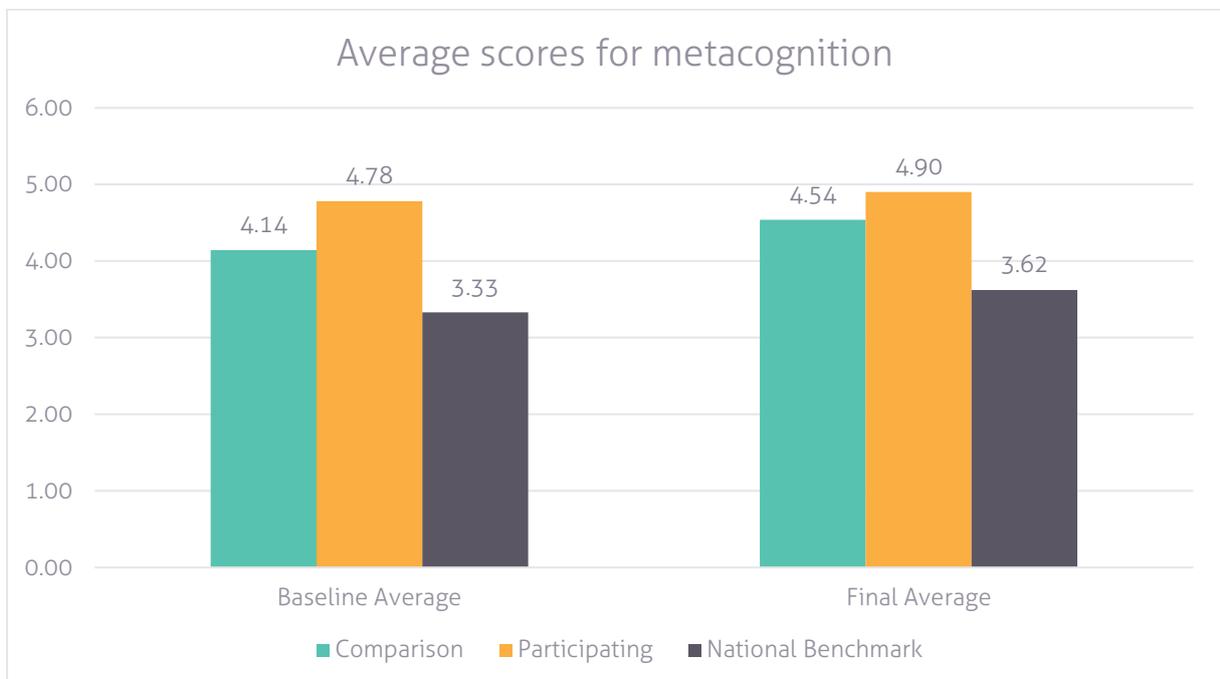


Figure 15 - Participants: n = 21; Comparison n = 21, National Benchmark - Baseline: n = 2663, Final: n = 288.

When we look at the schools separately, we see different trends. Ernest Bevin participating pupils' metacognition scores increased by 1.09% whilst their comparator pupils' score

increased by 0.76%. In contrast, Hartsdown Academy saw comparison pupils' metacognition scores increase by 12.99% whilst their participating pupils' metacognition scores increased by 2.75%.

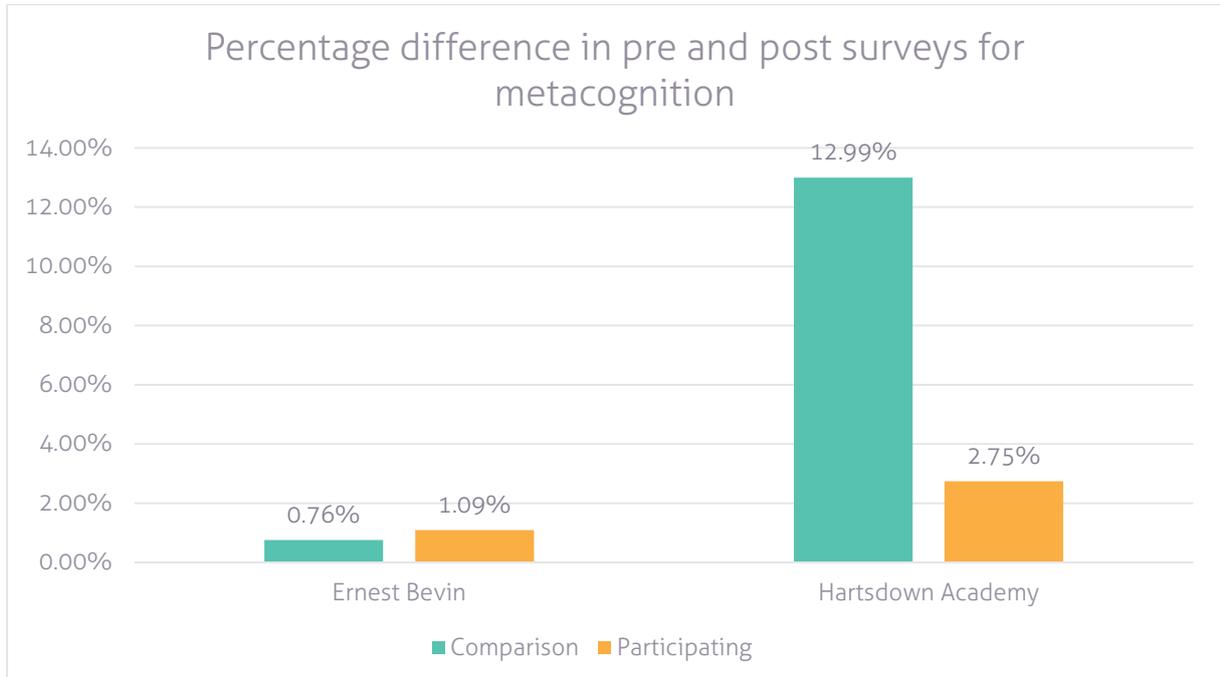


Figure 16 - Comparison - Ernest Bevin: n = 11, Hartsdown Academy: n = 10; Participating - Ernest Bevin: n = 9, Hartsdown Academy: n = 12.

Although Hartsdown Academy comparison pupils progressed more than their participating counterparts, the graph below shows that Hartsdown Academy participants still had a higher endline score than the comparison pupils. It also shows that all pupils from both schools were higher than the national benchmark in their average metacognition schools.

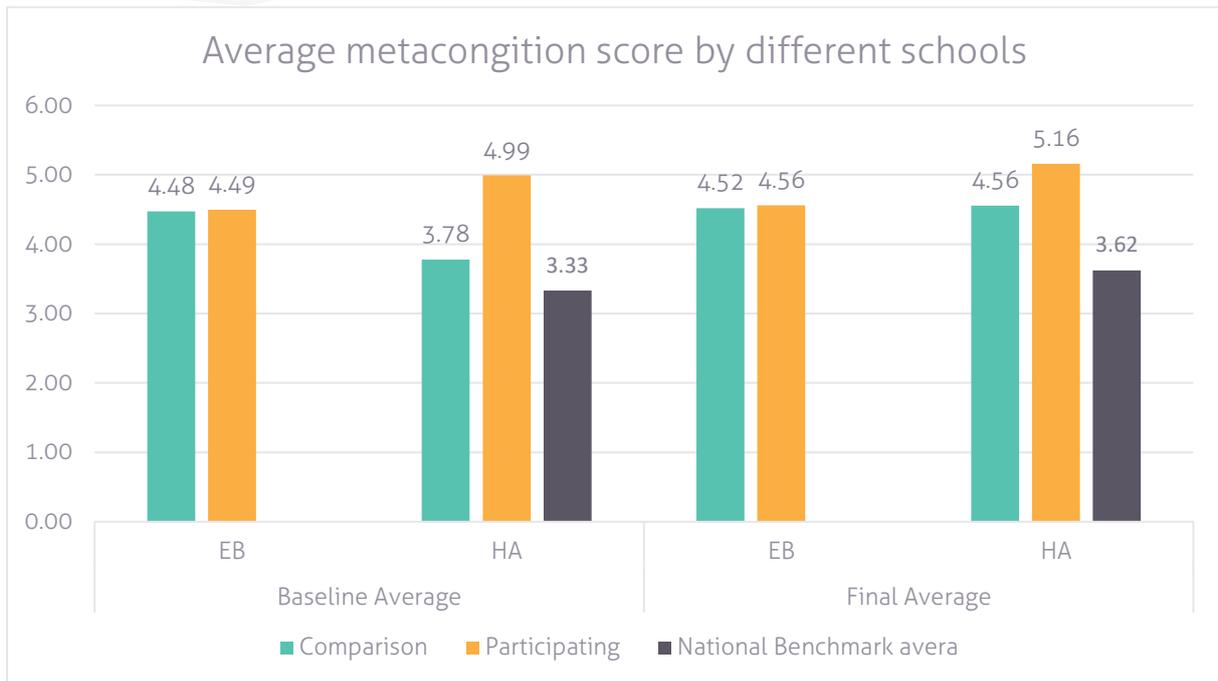


Figure 17 - Comparison - Ernest Bevin: n = 11, Hartsdown Academy: n = 10; Participating - Ernest Bevin: n = 9, Hartsdown Academy: n = 12; National Benchmark - Baseline: n = 2663, Final: n = 288.

Key finding: Tutors and teachers recognised that pupils had a better grasp of implementing more successful revisions techniques.

Zooming in on pupils’ revision skills, across the focus groups and interviews with tutors and teachers, there was some indication that pupils’ revisions skills had improved. Tutors reported giving revision tips to their pupils. They also noted that pupils were engaging and experimenting with an increasing range of revision techniques: practising more exam questions, using acronyms and visual tools. Some tutors noted that although pupils were trying different ways of revising, they were unwilling to implement and embed new approaches into their current way of revising. Other tutors said that revision techniques were not discussed at all and that they had found it difficult to share proper revision strategies with their pupils.

Some teachers at the school noted that pupils had started actively asking about revision tips. They also noted that the Science Stars sessions got pupils into the routine of revising and that they had seen improvements in their pupils’ study skills. There was uncertainty from the teachers to assign these changes to the Science Stars programme because of two key reasons: firstly, they were aware that pupils who had signed up to Science Stars were already motivated to study and secondly, they had not been in the sessions enough to see this change.

Key finding: Tutors and teachers thought that pupils were better at understanding their strengths and weaknesses.

In terms of pupils having a better understanding of their own learning, tutors and teachers indicated that the Science Stars programme had contributed to pupils gaining a better understanding of their own learning. Tutors said that some pupils were good at identifying strengths and weaknesses and then feeding these into exam strategies such as focusing on questions they were good at and then doing the ones they found trickier later.

Teachers also thought that pupils had become capable of connecting exam questions to areas of study, and then knowing with what they are struggling. Pupils also became much better at bringing work and questions they couldn't complete to gain extra support on this.

Tutors reported that they had been able to initially help pupils trying to understand their strengths and weaknesses and then encouraged practice in those areas. At points tutors still had to sign-post pupils' own knowledge to them, perhaps indicating that pupils did not have a complete understanding of their own learning.

Key finding: 2022/23 is the first academic year since 2019/2020 where Science Stars participants have improved their metacognition across the course of the programme.

The graph below shows this year (2022/23), Science Stars participants have once again, after two years of decreasing metacognition scores, shown a positive change in their metacognition levels across the duration of the programme.

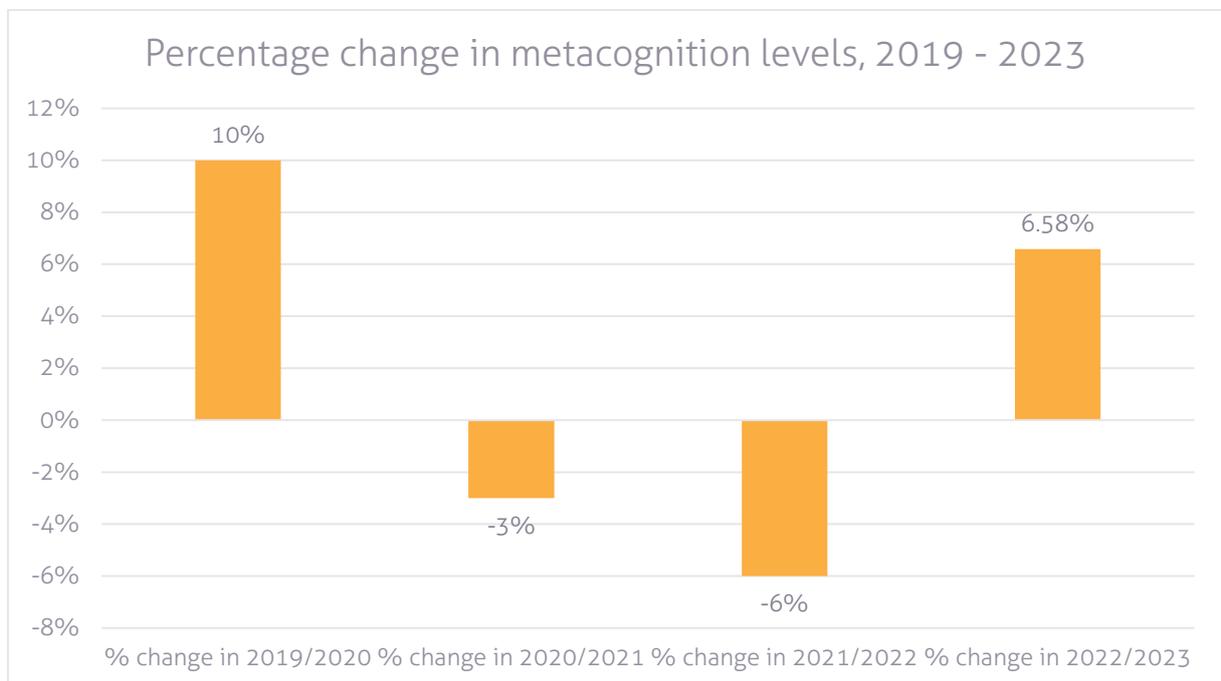


Figure 18 - 2022/23 - Participating: n = 21

Key finding: Science Stars participants' self-efficacy increased more than those in the comparison group.

Pupils' confidence in science was measured by the MSLQ sub-questionnaire on self-efficacy. This survey, administered at the start and end of the programme, found that on average participating pupils' self-efficacy levels increased by 4.13% whilst comparator pupils' self-efficacy levels decreased by 5.63%. This change was not statistically significant with a p -value of 0.08 ($n=50$).

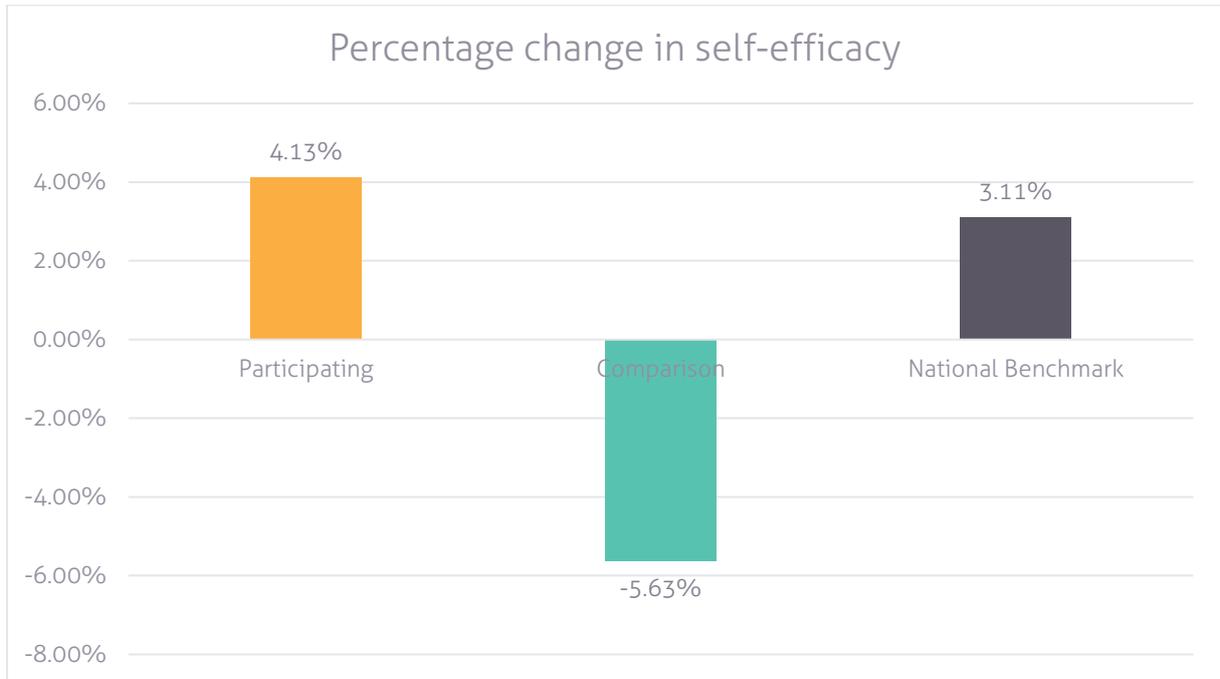


Figure 19 - Participants: $n = 24$; Comparison $n = 26$; National Benchmark – Baseline: $n = 925$, Final: $n = 289$

The graph below shows that participating pupils also had higher endline self-efficacy scores than their comparison peers as well as the national benchmark.

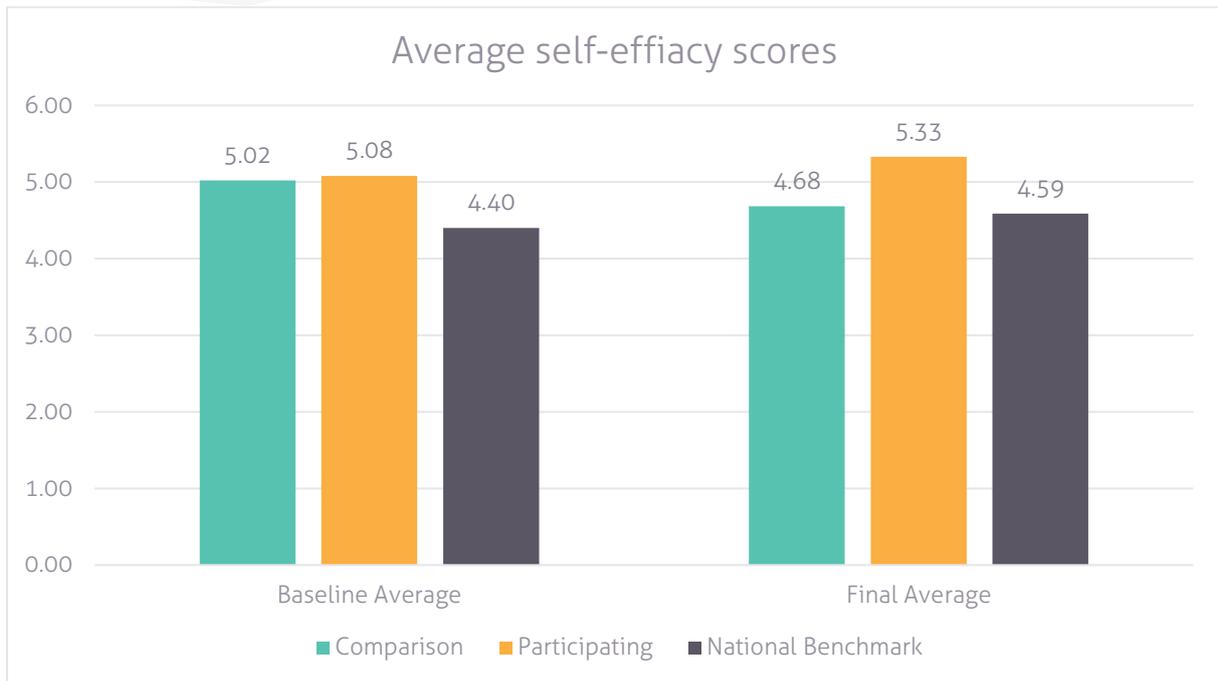


Figure 20 - Participants: n = 24; Comparison n = 26; National Benchmark – Baseline: n = 925, Final: n = 289

Key finding: The Science Stars programme had a much bigger positive impact on Ernest Bevin’s participating pupils’ self-efficacy than Hartsdown Academy’s participating pupils.

When we look at the schools separately, we can see that the overall percentage changes mask two quite different trends occurring at each school. Hartsdown Academy Science Stars participants saw a slightly larger decrease in their self-efficacy scores (-9.49%) than the comparator pupils (-8.02%).

Ernest Bevin Science Stars participants saw a substantial increase in their self-efficacy scores (+17.75%) whilst the comparator pupils saw a small decrease (-3.57%). Significance testing between participating and comparator groups for individual pupils was not possible due to sample size.

As Ernest Bevin’s Science Stars programme was delivered in-person, so having in-person tutors may have influenced this substantial increase in self-efficacy in comparison to the decrease in Hartsdown Academy participants who received the intervention remotely.

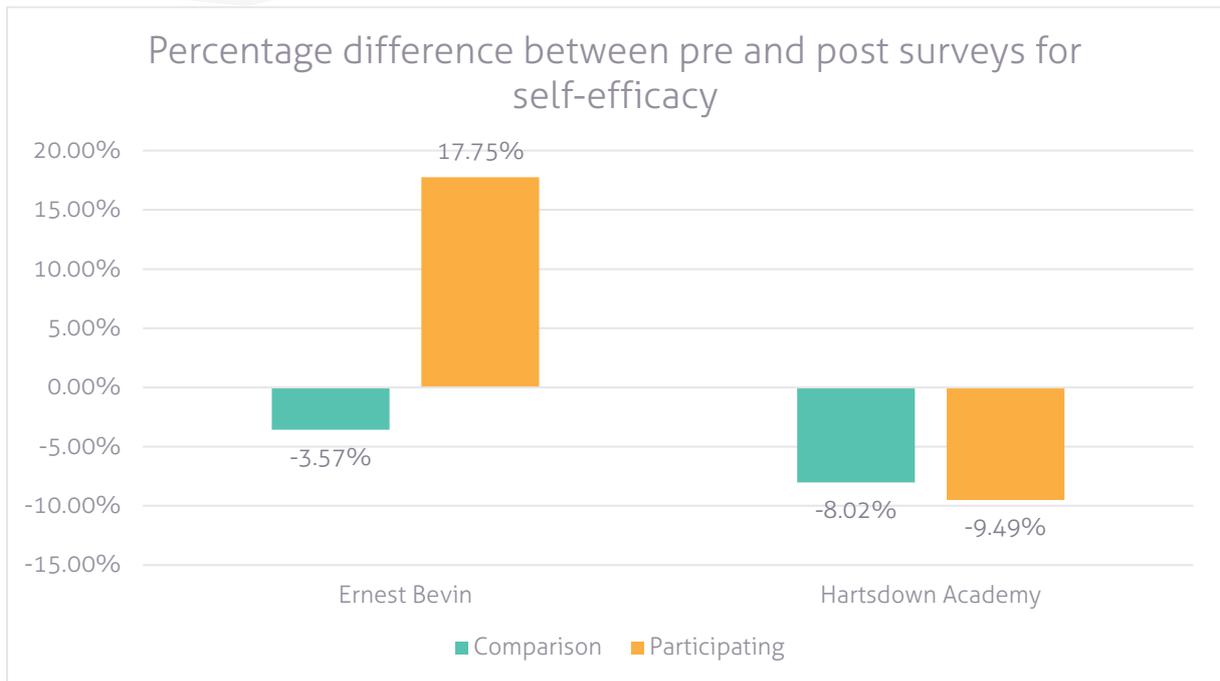


Figure 21 - Comparison - Ernest Bevin: n = 14, Hartsdown Academy: n = 12; Participating - Ernest Bevin: n = 12, Hartsdown Academy: n = 12.

The graph below shows that although participating pupils had a higher average self-efficacy score at baseline because of the huge improvements that Ernest Bevin participants made throughout the programme. In comparison to the Hartsdown Academy participants, the Ernest Bevin pupils end up with a higher average score than the Hartsdown participants at the endline.

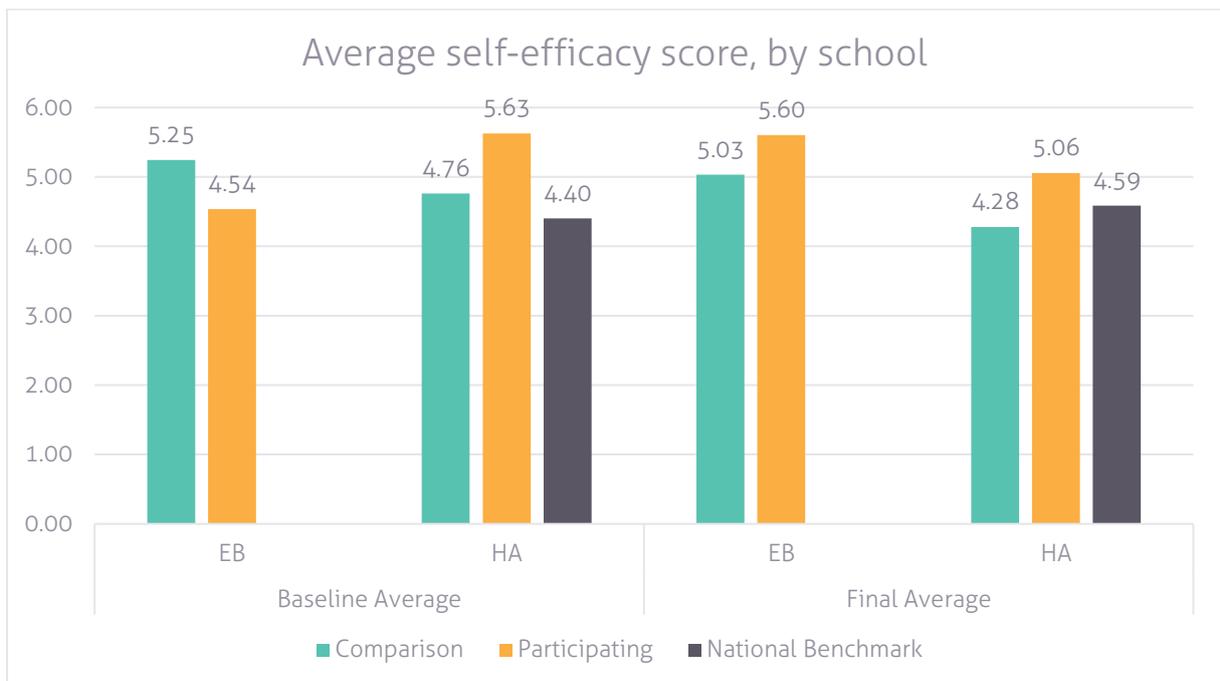


Figure 22- Comparison - Ernest Bevin: n = 14, Hartsdown Academy: n = 12; Participating - Ernest Bevin: n = 12, Hartsdown Academy: n = 12; National Benchmark - Baseline: n = 925, Final: n = 289.

Key finding: Tutors and teachers noted that pupils had increased confidence in certain aspects of science.

Tutors and teachers did report some kind of increase in pupils' confidence in science. Tutors expressed those pupils who had felt frightened by science seemed much more comfortable by the end of the programme. Tutors also noted that along with increased confidence in science, **pupils had increased motivation and enthusiasm**. The increase in pupils' confidence could be seen by tutors in the increase in pupils asking questions. It was expressed by teachers that pupils were more likely to ask their tutors questions rather than their teachers. Pupils were also more willing to try to grapple with more challenging concepts than before.

One key example of a pupil's increase in confidence is a pupil that, at the beginning of Science Stars, was unable to write in full sentences, but across the Science Stars programme, they learnt a helpful approach: spend time thinking about the answer, planning it out, and then writing it out. This made them more confident in writing sentences in science.

Tutors did note, however, that pupils continued to lack confidence in fill-in-the-gap exercises, as they felt they needed multiple choice options to answer these confidently.

Teachers noted that pupils seemed more confident about science because they knew they were receiving additional support. This was demonstrated by **pupils' increased interest in taking A-Level science** in comparison to the beginning of the year. Teachers also noted that the Science Stars programme had a positive impact on pupils' confidence more widely; there were mentions of increasing pupils' self-esteem due to pupils feeling like they were special and being invested in.

Key finding: After two years of participants' self-efficacy level decreasing across the programme, this year has seen participants show an increase in self-efficacy.

The graph below illustrates how the impact of the programme on the self-efficacy levels of participants has changed over time. In 2019/2020, Science Stars had a positive impact on participants' self-efficacy levels, with a +8% increase between the beginning and at the end. In both 2020/21 and 2021/22, participants' self-efficacy did not improve across the duration programme, their levels decreased. In 2022/23, however, Science Stars participants' self-efficacy have seen a positive improvement again, with participants' experiencing an average +4.13 percentage change in their self-efficacy across the duration of the programme.

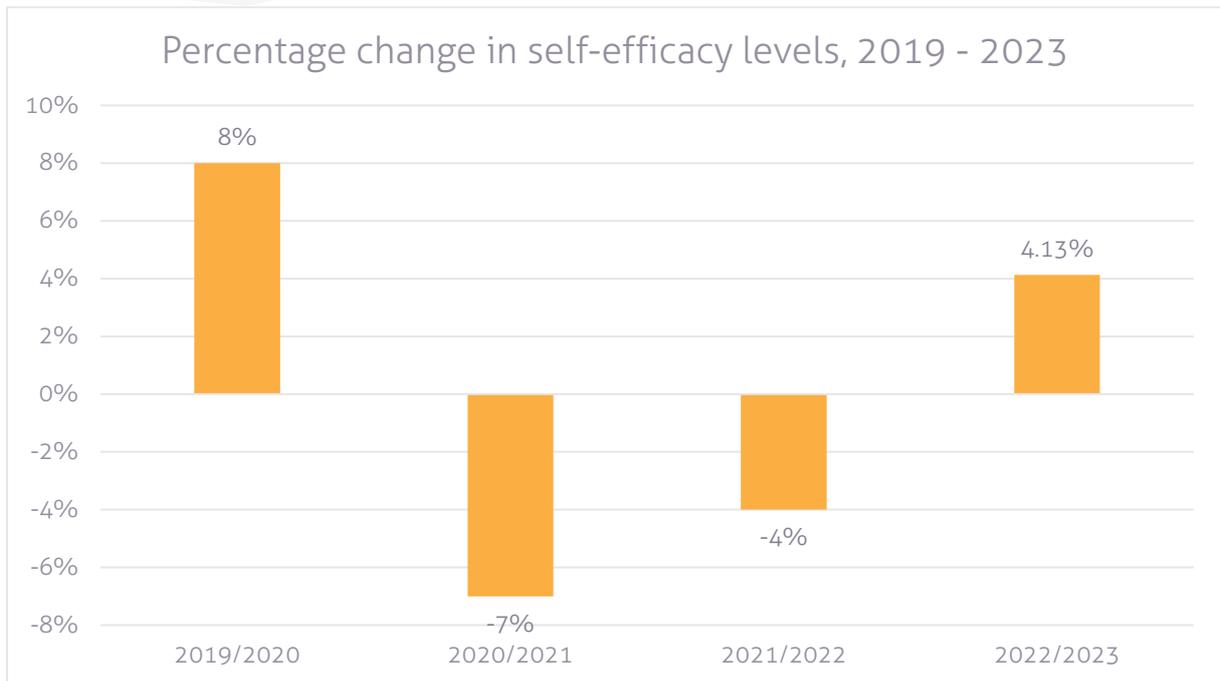


Figure 23 - 2022/23 - Participating: n = 24

Key finding: Participating pupils’ test anxiety level decreased by 9.38%, whilst comparison pupils’ test anxiety levels increased by 1.76%

Pupils’ anxiousness was measured by the MSLQ sub-questionnaire on test anxiety. It is important to explicitly state that a decrease in score for this measure means a decrease in test anxiety which is the more positive outcome. This skill survey administered at the start and end of the programme found that, on average, participating pupils’ test anxiety levels decreased by 9.38% whilst comparator pupils’ test anxiety levels increased by 1.76%. This change was not statistically significant with a *p*-value of 0.28 (n=50).

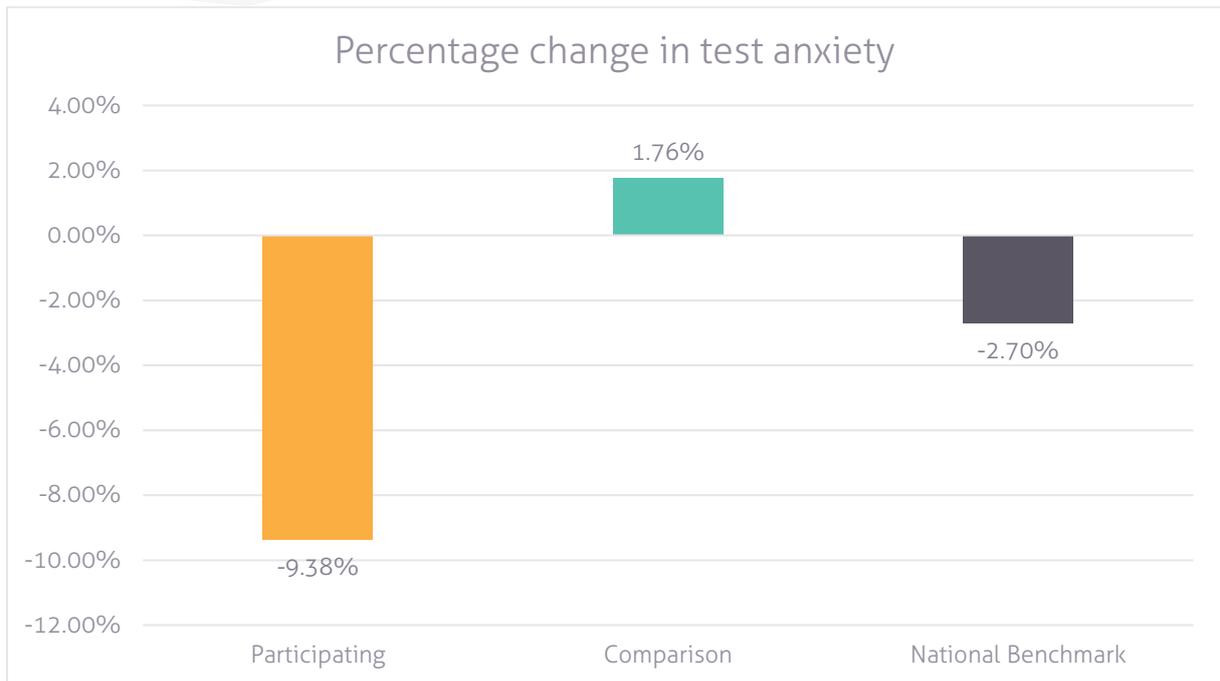


Figure 24 - Participating: n = 24, Comparison: n = 26, National Benchmark - Baseline: n = 967, Final: n = 832

The graph below shows another positive story for participating pupils' test anxiety levels. At the baseline survey, participating pupils' test anxiety was higher than their comparator peers and even higher than the national benchmark. By the endline, however, not only did participating pupils' test anxiety levels reduce, but they were also lower than the comparison groups', and almost become as low as the national benchmark.

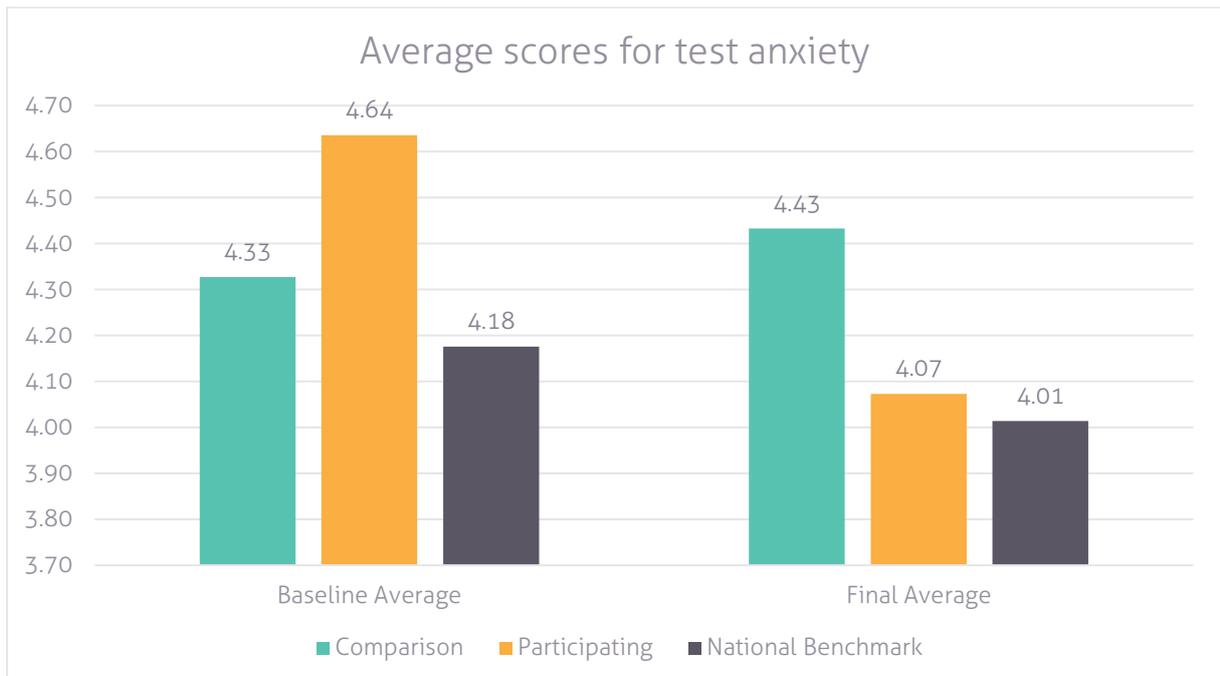


Figure 25 - Participating: n = 24, Comparison: n = 26, National Benchmark - Baseline: n = 967, Final: n = 832

Key finding: The difference between the change in test anxiety levels between participant and comparison pupils at Ernest Bevin was much larger than at Hartsdown Academy.

Looking at the schools individually, however, it becomes apparent that the overall comparator and overall participating pupils test anxiety scores obscures the different trends between the two schools. Pupils participating in the Science Stars programme at Hartsdown Academy saw their test anxiety levels decrease by 6.94% whilst the comparator pupils' test anxiety levels decreased by 7.20%. Ernest Bevin comparator pupils' test anxiety increased by 8.33% but the Science Stars participants' test anxiety levels decreased by 11.81%.

Once again, it is worth highlighting that, Ernest Bevin's Science Stars programme was delivered in-person, so having in-person tutors may have influenced this substantial decrease in test anxiety in comparison to the less substantial decrease in Hartsdown Academy participants who received the intervention remotely.

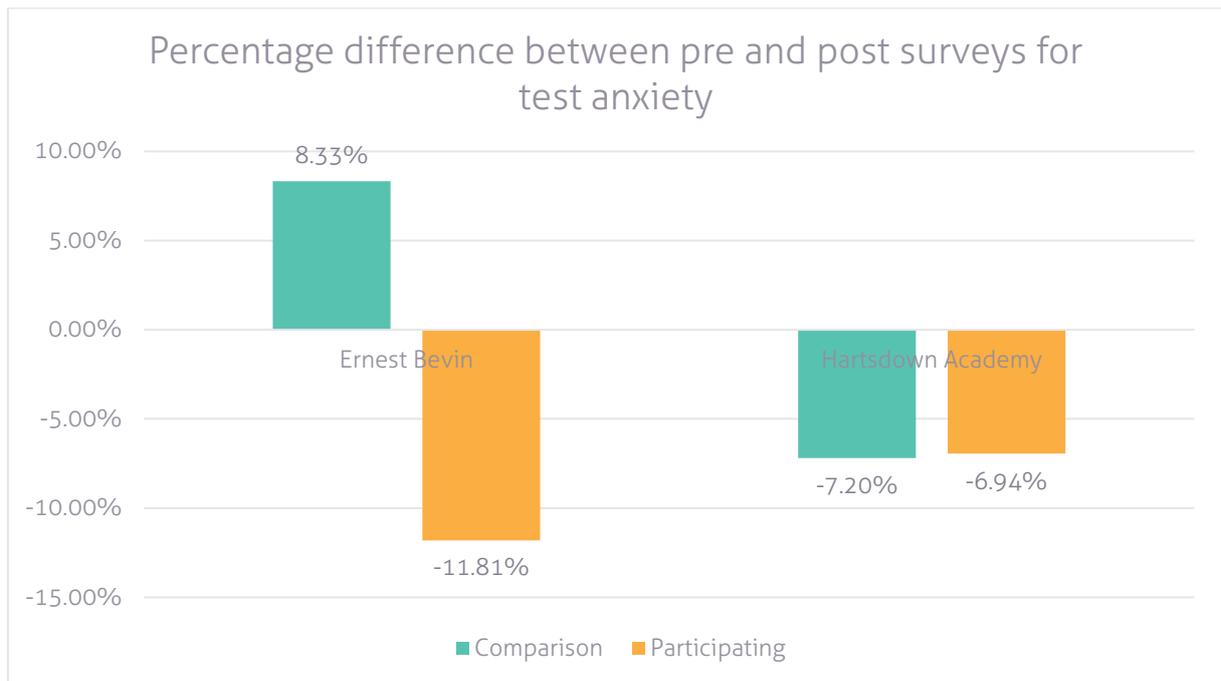


Figure 26 - *Comparison* - Ernest Bevin: n = 15, Hartsdown Academy: n = 11, *Participating* - Ernest Bevin: n = 12, Hartsdown Academy: n = 12. Although the above graph shows that participating pupils at Hartsdown Academy did not reduce their test anxiety levels as much as their comparison peers, the graph below shows that the impact of the Science Stars programme on their test anxiety got them to beneath the national benchmark, which their comparison peers did not achieve.

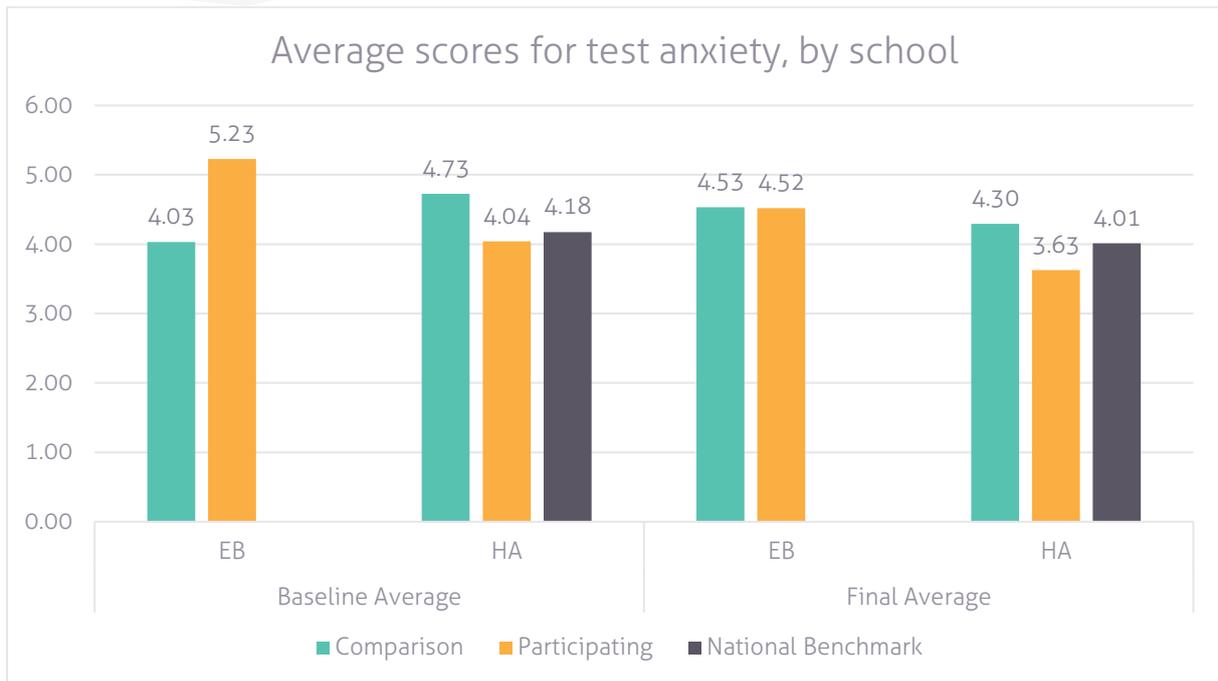


Figure 27 - Comparison - Ernest Bevin: n = 15, Hartsdown Academy: n = 11, Participating - Ernest Bevin: n = 12, Hartsdown Academy: n = 12; National Benchmark - Baseline: n = 967, Final: n = 832.

Key finding: Tutors and teachers both thought that Science Stars had reduced test anxiety levels.

Tutors were uncertain as to whether the Science Stars programme had positively impacted pupils’ anxiety around tests and exams, whilst teachers in schools seem to have a more positive perception of the impact had on pupils’ exam anxiety. Tutors reported that they had not spoken much to pupils about their anxiety levels, so it was hard to gauge whether the levels had decreased although looking at survey results it clear that all participating pupils experienced a decrease in test anxiety. Tutors noted that pupils were quieter after completing their mocks which they had interpreted as pupils being anxious. They did note, however, that **pupils were practising exam questions because of the programme** which may contributed towards decreasing their anxiety levels.

Teachers reported that pupils did seem a little less anxious because of the Science Stars programme. They attributed any reduction in anxiety to pupils knowing that they were receiving additional support and had more knowledge sources to ask questions. Pupils knowing that they had have access to formulas during exams decreased anxiety, although it is unclear whether this knowledge can be directly attributed to Science Stars. Teachers also noted that it was difficult to assess anxiety levels around test and exams as these had only just begun.

Key finding: Test anxiety levels in participants decreased more in participants this year than last year.

The graph below shows what the impact of the programme on the anxiety of participants has changed over time. From the first year of the programme (2019/2020), the general trend has been that Science Stars participants' test anxiety levels have decreased.

In 2020/21, however, participants' test anxiety levels remained the same. Across last year (2021/22) and this year (2022/23), participants' test anxiety from the beginning of the Science Stars programme to end of the programme decreased again. The average percentage decrease in participants' anxiety this year (2022/23) was -9.38% which shows that the impact of Science Stars on participants is returning to the same level of impact it was having in the first year of the programme.

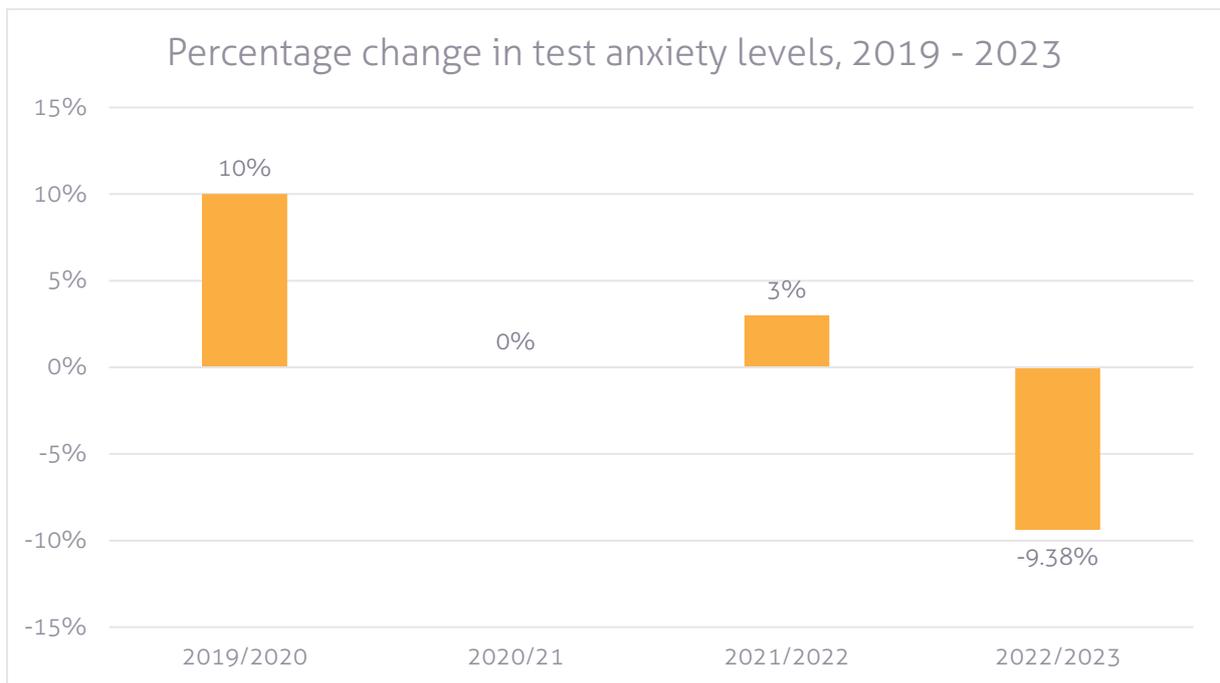


Figure 28 - 2022/23: - Participating: n = 24

5. Conclusion

The Science Stars programme in 2022/23 provided **many positive results**. Data showed that Science Stars participants **improved their science GCSE grades more**, and a **greater percentage of them achieved their target science GCSE grade**, than their comparator peers.

When it came to non-cognitive skills, Science Stars participants **made greater improvement than their comparator peers in self-efficacy and test anxiety**. In-person participants saw greater improvement in non-cognitive skills than their remote participant peers. Although the difference between in-person and remote delivery have been highlighted, SGUL have already decided to return all delivery to in-person having already recognised the value of doing this. It is important to note that **many of the negative trends we saw from last year's evaluation have been reversed** which is promising for future years of the programme.



Partners and supporters



**Improving pupil outcomes
by working with schools to
address the evaluation deficit.**

Get in touch
hello@impacted.org.uk

