



What impact does the Silver CREST Award have on science scores and STEM subject selection?

QUANTIFYING CREST

CREST SILVER AWARD EVALUATION
A PRO BONO ECONOMICS RESEARCH REPORT
FOR THE BRITISH SCIENCE ASSOCIATION

BY

ROSIE STOCK JONES
TOM ANNABLE
ZOE BILLINGHAM AND
CEE MACDONALD

CONTENTS

Forewords	4
The authors	7
Executive summary	8
1. INTRODUCTION	10
2. LITERATURE REVIEW	12
2.1 Existing CREST research	13
2.2 Wider research on school-based science interventions	13
2.2.1 Inquiry-based learning	13
2.2.2 Extra-curricular activities	14
2.3 How this research will fill existing research gaps	14
3. DATA	16
3.1 CREST Awards data	17
3.2 National Pupil Database data	17
3.3 Linking data	17
3.4 Cleaning the data	17
3.5 Final dataset	17
3.6 Limitations	18
4. METHODOLOGY	20
4.1 Propensity score matching (PSM)	21
4.2 Outcome variables	22
4.2.1 Limitations	22
4.3 Control variables	22
4.3.1 Limitations	23
5. RESULTS	24
5.1 Research question one: what are the characteristics of students taking Silver CREST Awards?	25
5.1.1 CREST projects	25
5.1.2 Student demographics	26
5.1.3 Student attainment	27
5.2 Research question two: Does participation in the Silver CREST Award programme have an impact on attainment in science subjects at GCSE?	29
5.2.1 By gender	30
5.2.2 By project delivery type	30
5.2.3 By free school meals status	30
5.3 Research question three: does participation in the Silver CREST Award affect the likelihood of taking a STEM AS level?	31
5.3.1 By gender	31
5.3.2 By project delivery type	32
5.3.3 By free school meals status	32
6. DISCUSSION	34
6.1 Interpreting the results	35
6.2 Demonstrating causality	35
6.3 Wider applicability of results	38
7. RECOMMENDATIONS	40
Bibliography	45
Annex A - Nesta Standards for Evidence	46
Annex B - Further tables	47
Annex C - A detailed look at our methodology choices	52



It is widely recognised that Britain faces a skills shortage amongst businesses requiring high-level or technical science, technology, engineering and maths (STEM) skills that will enable us to compete in a global knowledge economy. Increasing the number of young people who continue in STEM education, therefore, is one of the key education challenges of our times. So too is the need to improve the level of science literacy amongst those who do not go on to work in the STEM industries; to equip a future generation - who will be grappling with issues from cyber security and AI to climate change and the challenge of feeding the global population - with the confidence to question or challenge experts, to evaluate evidence and understand risk and uncertainty.

The British Science Association (BSA) has been running the STEM enrichment programme for young people, CREST Awards, for almost 30 years. CREST is a curriculum-enhancement award scheme that recognises success, and enables students to build their skills and demonstrate personal achievement in project work. It acts as a framework for educators, accredits partner activities, and gives students the freedom to design their own projects and experiments.

At the BSA, we have long-suspected that CREST is a great thing, of course - but creating a firm evidence base was less straightforward. There has been some research - as our team of volunteer economists explain in this report - but nothing that has independently evaluated the impact of the programme using achievement and progression data.

This ground-breaking report, which is the first quantitative evidence of the impact of extra-curricular STEM interventions, reveals that undertaking CREST Silver Award appears to have a positive impact on students' GCSE grades and the likelihood of them continuing on to AS levels in STEM subjects. That in itself is hugely significant. However, I am particularly interested in the finding that there is an even greater impact on students who are or have been eligible for free school meals.

One of the BSA's priorities, set out in our vision and mission, is to grow and diversify the community of people participating, engaged and interested in science. At a schools level this means both broadening the reach of CREST beyond the traditional boundaries of science and design and technology teachers, into geography, history, English and drama, for instance (making a film about climate change could as easily sit as a project outside of the science classroom or club), as well as ensuring that it is not only the most able or privileged students who participate in CREST. It means that we must work harder to remove barriers to access, to support teachers and STEM club leaders, particularly those in rural or disadvantaged areas, and to show all students that CREST is fun, rewarding and applicable to their interests, skills and talents.

Outside of the traditional school setting, we want to ensure home educators have access to information and advice, and we are keen to continue to grow participation by groups such as Brownies, Scouts and youth clubs. To realise this ambition, the BSA is embarking on building a digital platform for the delivery of CREST that will support the expansion of the programme and make it more accessible.

We believe that CREST, with its student-led, hands-on, project-based approach can play an important role in inspiring a new generation of future STEM leaders for the UK. This report is an excellent first step in building a reliable evidence-base to demonstrate the impact of CREST.

I would like to thank Cee, Rosie, Tom and Zoë for this excellent piece of work, as well as the Pro Bono Economics team for enabling this project to happen and for overseeing it. This collaboration means that we have achieved a first in producing evidence of this kind into the impact of CREST on attainment and subject choice, now occupying a unique position as the only extra-curricular STEM activity with this kind of evidence.

I would like to thank all the Local Coordinators, teachers, mentors - and the BSA's education team and its stakeholder group - for delivering CREST so brilliantly, and helping to generate the positive impact seen here. And finally, well done and congratulations to all CREST Silver students who achieved their award during this period.

I commend this report to you, and hope that you will support the BSA to continue the growth of CREST as we work towards its milestone 30th anniversary in September 2016.

Rt. Hon the Lord Willetts
Chair, British Science Association

FOREWORD FROM PRO BONO ECONOMICS



On behalf of Pro Bono Economics, I am delighted to introduce this report for the British Science Association. I would like to both thank and congratulate Tom Annable, Zoë Billingham, Cee MacDonald and Rosie Stock Jones for their hard work in producing this report. We are also grateful to our peer-reviewers for constructive and timely comments.

To date, robust research into the impact of the Silver CREST Award has been limited, with findings often coming from self-reporting or from relatively small datasets. The analysis reported here has been able to make use of the fantastic national resource which is the National Pupil Database (NPD) compiled by the Department for Education (DfE). DfE statisticians were extremely helpful in facilitating appropriately data-protected access to records of all pupils in English state schools who took their GCSEs between 2010 and 2014. The DfE statisticians were able to match the British Science Association's data on pupils taking the Silver CREST Awards with the records of their subsequent GCSE results. The volunteer economists were then able to construct a statistically similar control group of pupils who had not taken the CREST Awards from the NPD using Propensity Score Matching techniques. This enabled the volunteer economists to make a more reliable estimate of the effect of the Award programme on the take-up of science subjects and attainment.

There are always caveats with this type of analysis, but the results suggest that students participating in Silver CREST achieve about

half a grade higher on their best science GCSE result on average compared with a statistically-similar control group. Silver CREST students were also 21% more likely to take a STEM AS level subject than control group students. Further research is recommended in the report to improve the evidence further - e.g. through Randomised Control Trials, conducting a Cost-Benefit Analysis or making more use of data from Bronze and Gold CREST Awards. Nevertheless, the work reported here represents a step up in robustness for evidence of CREST's impact.

As ex-Chief Economist of the Department for Education, I am delighted to see volunteer government economists using and developing their skills to strengthen the evidence base on educational matters. We also hope that the results of this report will support the British Science Association in making a case for encouraging greater CREST uptake in schools.

Karen Hancock
Economist,
Pro Bono Economics
Ex-Chief Economist,
Department for Education

THE AUTHORS

Rosie Stock Jones works as an economist at the Cabinet Office for the central Analysis and Insight team. She currently provides analytical support to the Office for Civil Society.

Tom Annable is also an economist for the Cabinet Office Analysis and Insight team and has worked there since 2014. Before that he obtained a Masters in Economics from the University of Warwick, where he specialised in Labour Market Econometrics.

Zoë Billingham is an economist in the Treasury working on EU reform, having previously worked in the Deputy Prime Minister's Office on social mobility.

Cee MacDonald loves charts and chickpeas. She has worked as an economist for the Department of Work and Pensions and Cabinet Office and is currently based at Defra.

Rosie, Tom and Zoë would like to give particular thanks to Cee, who skilfully led and managed this project over the course of a year, guiding us over analytical obstacles (and keeping us going through our post-day job evening sessions!). Thank you, Cee.

The authors would also like to thank Richard Garbitz, Aidan Mews, Georgina Eaton, Julie Moote, Karen Hancock, Fiona Ramsay, Michael Reiss, Iain MacDonald, Leena Philips, Radana Crhova, Hannah Kirk, Adrian Fenton, Katherine Mathieson, Alex Taylor, Amy MacLaren and Louise Ogden.

We would also like to thank our peer reviewers - Tom Battrick and Ravi Kanabar, from FTI Consulting - for their helpful comments and suggestions.

This analysis was independently undertaken as a Pro Bono Economics project, and we received no compensation for this work.

If using this report for future research please reference it as:

Stock Jones, Annable, Billingham and MacDonald (2016), "Quantifying CREST: what impact does the Silver CREST Award have on science scores and STEM subject selection?"
A Pro Bono Economics research report for the British Science Association.

EXECUTIVE SUMMARY

Introduction

The British Science Association's (BSA) CREST Awards are a practical science intervention for school-age children, which seeks to broaden their interest in science and encourage them to continue with STEM subjects. This analysis focuses on students in English state schools the majority of whom were aged 14-16 and who took part in CREST Silver Awards between 2010 and 2013. The report addresses three research questions:

1. What are the characteristics of students taking Silver CREST Awards?
2. Does participation in the Silver CREST Award programme have an impact on attainment in science subjects at GCSE level?
3. Does participation in the Silver CREST Award impact on the likelihood of taking a STEM AS level?

Research to date

To date, research regarding the effectiveness of CREST has been limited to studies with self-reported outcomes. For example, Moote (2014) found participation in CREST led to students reporting a positive impact on their self-regulated learning. Related studies looking at the impact of inquiry-based learning and extra-curricular studies have also shown positive impacts. For example, Metsapelto and Pulkkinen (2012) demonstrate the impact of extra-curricular activities on academic attainment. However, previous research has also set out the difficulties of attributing causality, given broader factors such as teacher and pupil enthusiasm.

Methodology

In order to conduct this research, data collected by the BSA on students starting Silver CREST Awards between 2010 and

2013, was linked to data in the National Pupil Database. This enabled us to bring together information on CREST participants with pupil characteristics, attainment and subject selection data. The total sample sizes for this analysis were 2.4 million students at Key Stage 4 (KS4), of whom 3,800 undertook a Silver CREST Award and 1.0 million at Key Stage 5 (KS5), of whom 2,300 participated in Silver CREST. The number of Silver CREST students in the sample who were eligible for Free School Meals (FSM) was 380 at KS4 and 190 at KS5.

Propensity Score Matching was used to create a control group of students who did not take part in CREST, but had similar characteristics to those who did. Key Stage 2 (KS2) SATs results were used to control for prior attainment. There is a strong relationship between KS2 attainment and GCSE attainment. However, these exams were taken around five years prior to GCSEs and A level choice, and do not give a detailed picture of attainment. Other control variables included gender, ethnicity, region of school, year of GCSEs, participation in triple award science, type of school, free school meal status in the six years prior to taking GCSEs and the Income Deprivation Affecting Children Index (IDACI).

The highest science point score at GCSE was chosen as the outcome variable for research question two (impact on attainment). This is a pre-existing variable in the NPD and helped to remove difficulties comparing performance across differing numbers of science GCSEs taken. For research question three (STEM take-up at KS5), we created a binary indicator for whether AS-level students had selected a STEM AS level subject. The main science, maths, technology and engineering subjects were considered qualifying choices. Analysis for this research question was restricted to non-vocational AS levels.

EXECUTIVE SUMMARY

Results

In regards to the first research question, we found that students taking Silver CREST were broadly representative of the wider pupil population in terms of gender and ethnicity. However, the CREST students were substantially less likely to have been eligible for free school meals (10% versus 22%) or have special educational needs (7% versus 16%) than the wider pupil population. Further, CREST students were more likely to have achieved stronger results at KS2 across all subjects. Their higher attainment was also reflected in their results at GCSE; students who undertook Silver CREST had higher average GCSE grades than those who did not do Silver CREST.

The second research question looked at the impact of Silver CREST Award participation on GCSE science results. We found that students who took CREST achieved half a grade higher on their best science GCSE result, compared to a statistically matched control group. We looked at a subset of students who had been eligible for free school meals (FSM) in the six years prior to their GCSEs. The CREST students eligible for FSM saw a larger increase in their best GCSE science score (two thirds of a grade) compared to a matched control group of other students who were also eligible for FSM. These results were statistically significant.

Finally, we looked at the subject choices students made in their AS levels. 82% of CREST students took a STEM AS level compared to 68% of a statistically matched control group. CREST students were therefore 21% (or 14 percentage points) more likely to take a STEM AS level than students in the control group. For students who had been eligible for free school meals this difference was larger (38% or 21 percentage points). These results were statistically significant.

Discussion and recommendations

To demonstrate causality i) the treatment must have occurred before the outcome, ii) there must be a demonstrable statistical link between the presence of the treatment and the outcome, and iii) there must be no alternative plausible explanation. This research satisfies the first condition for the majority of students (a small proportion are likely to have taken CREST after their GCSEs and AS level choices). The second condition is satisfied, but not the third. We cannot rule out the possibility that other unobserved variables are affecting GCSE results and AS level subject choice. For instance, the motivation and enthusiasm of students and teachers for science are likely to have an effect, yet we have not been able to control for these. There may also be selection bias into Silver CREST if only high achieving students are permitted to participate, or if the Award only appeals to very bright students. Further research would be required to determine whether CREST has a causal effect on GCSE science attainment and AS level subject choice.

We make several recommendations for further work, including replicating this analysis through a Randomised Control Trial, broadening it to cover Discovery, Bronze and Gold Award types and conducting a cost-benefit analysis for schools. Additionally we make three broader recommendations: that charities ensure accurate and usable data collection, that young people consider taking part in project/inquiry-based learning such as CREST and finally that the BSA consider targeting CREST at students from low income families.

To download a full copy of this report, please visit: bsa.sc/CRESTImpact

INTRODUCTION



1 INTRODUCTION

Whilst many of society's biggest risks and challenges, such as pandemics and climate change, are scientific in nature¹, the UK currently has a science skills shortage. To help address this, it is important to understand what can support students to achieve in science subjects and encourage them to continue with science, technology, engineering and mathematics (STEM) study. However, given limited resources and teacher time, schools have to make trade-offs about which activities will best encourage their students to participate and achieve in STEM subjects.

The CREST Awards programme seeks to address this skills shortage by engaging pupils in STEM projects. The British Science Association (BSA) describes this as "hands-on science", which "builds transferable skills for further education and future employment"². Operating UK-wide, CREST "offers educators an easy-to-run framework for curriculum enhancement and is student-led, which means that young people take ownership of their projects and choose to undertake them in areas they enjoy or see as relevant"³.

There are four levels of Awards in the CREST programme; Discovery, Bronze, Silver and Gold, which each require increasing amounts of teacher and student time and mentor involvement.

The scheme has been running since 1986 and around 30,000 young people achieve an Award each year. Of the 33,000 participants in 2014: 10,700 did Discovery, 15,400 did Bronze, 4,300 did Silver and 2,600 Gold.

Previous research to evaluate the impact of CREST focused on the process of the programme and the self-reported experience of participants. Our report is a quantitative analysis of the impact of the Silver Award on attainment of students in English state schools in GCSE science subjects and on the probability of them taking a STEM AS level. Specifically, the research questions we set out to answer were:

1. What are the characteristics of students taking Silver CREST Awards?
2. Does participation in the Silver CREST Award programme have an impact on attainment in science subjects at GCSE level?
3. Does participation in the Silver CREST Award impact on the likelihood of taking a STEM AS level?

This report is structured as follows: firstly we review the existing literature on CREST and science interventions more broadly; then we explain our data and methodology, before reporting and discussing our results in detail. Finally, we make recommendations as a result of our research findings.

1. www.britishtscienceassociation.org/vision

2. www.britishtscienceassociation.org/crest-awards

3. Ibid.

Table 1.1 CREST Award levels and requirements

Type of CREST Award	What it involves
Discovery	Around 5 hours' project work; typically undertaken by 11-to-14-year-olds
Bronze	Around 10 hours' project work; typically undertaken by 11-to-14-year-olds
Silver	Around 30 hours' project work; typically undertaken by 14-to-16-year-olds
Gold	Around 70 hours' project work; typically undertaken by 16-to-19-year olds



LITERATURE REVIEW

2 LITERATURE REVIEW

2.1 Existing CREST research

Existing qualitative studies have linked CREST to positive outcomes for students, including increased self-regulated learning and the motivation to study science further.

Julie Moote (2014)⁴ found that participating in Bronze CREST Awards increased students' self-reported levels of self-regulated learning and career motivation in comparison to a control group. The increases were retained six months following programme completion, however, after nine months no statistically significant effect was found.

The University of Liverpool's survey based evaluation (2007)⁵ found that among Silver Award students participating in CREST, 78% rated it as good or very good and 79% regarded CREST as worthwhile. The University of Liverpool also found that of the Silver CREST students sampled, 16% said that doing CREST increased their take up of further study in science, engineering or technology subjects. This is directly relevant to our third research question. The CREST Expansion report (2011)⁶ also discussed future science careers among CREST participants but found that many of the Bronze Award students were not yet interested in thinking about career choices at that age (11-14).

Finally, there is some evidence that the framework of CREST supported the development of more general employability skills. Elaine Hendry's project for the BSA (2013)⁷ found that the emphasis on employability skills varied across the different levels of Award but suggested the young people would see some skills improve.

2.2 Wider research on school-based science interventions

2.2.1 Inquiry-based learning

Inquiry-based learning is a process where students answer a question or solve a problem to obtain knowledge, rather than being presented the information directly, which is similar to how CREST is run. There is some strong evidence on the positive impact that inquiry-based learning has on students' learning and attainment, predominantly from studies in the USA.

Robinson, Dailey, Hughes and Cotabish (2013)⁸ looked at the impact of training teachers in the USA in how to integrate inquiry-based learning into their STEM lessons. They found that students in the treatment group had better science concept and content knowledge, and a better understanding of science process skills. These results were statistically significant and the teachers included in the study were randomly assigned, allowing the authors to demonstrate causal effect. However, because the study matched students based on attainment, the number of observations was low with 154 in the treatment group and 130 in the control.

A similar study by Lui, Lee and Linn (2010)⁹ measured the impact of inquiry-based instruction on students' science scores and had large enough sample sizes to be deemed reliable (2,685 treated, 2,060 in control). They found that the treatment cohort performed better than the control in both tests that assess knowledge directly linked to the intervention and tests that assess wider application. However, Lui, Lee and Linn found that teachers with more experience delivered better results across both cohorts, and that students did better with teachers who did not need mentoring. These findings demonstrate the importance of teacher quality, a variable not available to us for our analysis, and the difficulty isolating the impact of inquiry-based learning. These studies suggest that inquiry-based learning projects like CREST can have a positive effect on attainment and understanding across a variety of subject areas.

4. Moote, J. (2014), 'When Students Lead: Investigating the Impact of the CREST Inquiry-Based Learning Programme on Changes in Self-Regulated Processes and Related Motivations Among Young Science Students', PhD University of Edinburgh.

5. Grant, L (2007), 'CREST Awards evaluation: Impact study', University of Liverpool Science Communication Unit.

6. Fisher S, Tifford R & Gammon B (2011) CREST Expansion Scheme 2009-2011, London: British Science Association.

7. Hendry, E. (2013) 'CREST Awards and Employability Skills: A project for the British Science Association'.

8. Cotabish, A., Dailey, D., Robinson, A. and Hughes, G. (2013), 'The Effects of a STEM Intervention on Elementary Students' Science Knowledge and Skills. School Science and Mathematics', 113: 215-226. doi: 10.1111/ssm.12023.

9. Lui, Lee & Lin (2010), 'Multifaceted Assessment of Inquiry-Based Science Learning', Educational Assessment, 15, 69-86.

2 LITERATURE REVIEW

2.2.2 Extra-curricular activities

Other studies have looked at the impact of extra-curricular activities on attainment. In our sample CREST was delivered to around three in ten students in lesson time, one fifth in a club format, and one tenth as an off timetable event¹⁰.

Herbert Marsh (1992)¹¹ found that the take-up of extracurricular activities, including school based subject activities comparable to CREST, was “significantly and favorably [sic] related” to a range of outcomes including academic achievement. However, Marsh concedes that the self-selecting nature of extra-curricular participation is a source of bias that cannot be removed entirely by control variables. Further, and of interest to this CREST analysis, his analysis of variable interactions “indicates that students from lower-SES families benefit more than students from higher-SES families”. More recently, Metsäpelto and Pulkkinen (2012)¹² have used longitudinal data on the Integrated School Day Programme in Finland from 2002 - 2005 to show that participation in performing arts and academic clubs was related to higher academic attainment. Their study has several strengths as it uses panel data and controls for the grade level and the initial level of the outcome variable. However, their sample size is relatively small at 302 students, and the non-experimental design means it is difficult to draw causal conclusions due to the interaction of complex factors such as parental encouragement and socioeconomic background.

These findings demonstrate the potential for a positive connection between extra-curricular activities and attainment. However it also shows that evaluating school-based learning interventions is difficult. Problems surrounding access to data, the control of important variables, finding a large enough sample and finding a suitable control group

are common, and have often meant the number and findings of these studies are limited. Furthermore, there is a specific gap in the CREST literature as there have not yet been quantitative studies on the impact of CREST using externally validated attainment measures.

2.3 How this research will fill existing research gaps

Nesta has developed a Standards of Evidence framework to help different stakeholders understand the degree of confidence there can be that an intervention is having an impact (for more details see Annex A)¹³.

The framework consists of a scale from one (the evidence can give an account of likely impact that is based on a logical theory) to five (the evidence is robust enough that the intervention can achieve the same impact when replicated faithfully by others, and/or at scale). The existing evidence on the impact of CREST could be classified as level one or two. This is because studies to date have widely been based on surveys, which are prone to response biases and have not included control groups. Moote's research included a control group, but had a relatively small sample of 200 in the treatment group and a self-reported, subjective, outcome measure.

10. This is likely to be a lower end estimate as data on delivery type is missing for just under half of our sample.

11. Marsh, H. (1992). Extracurricular activities: Beneficial extension of the traditional curriculum or subversion of academic goals? *Journal of Educational Psychology*, 84(4), 553-562.

12. Metsäpelto, R., & Pulkkinen, L. (2012). Socioemotional Behavior and School Achievement in Relation to Extracurricular Activity Participation in Middle Childhood. *Scandinavian Journal Of Educational Research*, 56(2), 167-182.

13. Puttick, R & Ludlow, (2013) 'Standards of evidence: an approach that balances the need for evidence with innovation'.

2 LITERATURE REVIEW

This piece of analysis fills three key gaps in the CREST literature so far:

1. **Robust methodology.** It uses Propensity Score Matching to create a matched control group. This is a more robust methodology than a 'before and after comparison' and can provide stronger evidence that any impact found was due to CREST rather than other factors.
2. **Sample sizes.** There are large numbers in both the treatment and control group, which make it less likely that our findings occurred by chance and more likely that they will be replicable.
3. **Objective outcomes.** This analysis measures quantitative, objective outcomes rather than self-reported metrics. Data on both GCSE attainment and AS level study choice come from the National Pupil Database and therefore are externally validated rather than self-assessed.

We consider that our analysis meets the Level 3 criteria under the Nesta Standards of Evidence.



A great teacher at my school ran a science club... so I got involved from there. I believe it really helped me to get into a good university and experience in industry. It definitely makes studying easier when you're doing something that you enjoy. The sense of achievement when the project was completed and you were presenting to others is a great feeling.

Rachel, Hertfordshire

(Age 15 when she undertook her CREST Silver project which explored different spot creams)



3 DATA



3 DATA

3.1 CREST Awards data

The BSA is responsible for running CREST Awards and collects data on participants. This data includes information about: the Award level, date, student name, school name, school address as well as information about the student project such as type and subject. Forms used to collect this data are filled out by students and teachers and uploaded to the database by CREST Local Coordinators. However, not all of the fields are filled in, for example only 23% of CREST records had a Unique Pupil Number (UPN). For this research, BSA sent records of 11,700 students to the Department for Education (DfE) for linking with the NPD. The dataset included all the students in England who started a CREST Silver Award between January 2010 and December 2013, apart from a small number of students and teachers who opted out of our study. Only two schools with relevant data, comprising a total of 36 records, opted out of the study and as a result we do not expect non-response bias.

3.2 National Pupil Database data

In order to measure the impact of CREST on attainment and uptake of STEM subjects, DfE linked students in the CREST database to their record in the National Pupil Database (NPD). The NPD holds information about students in schools in England. It includes information on attainment as well as pupil characteristics such as gender and ethnicity.

3.3 Linking data

The Department for Education used a combination of UPNs, forenames, surnames, dates of birth and school names to link the students in the CREST database with the NPD. The 2,600 (23%) student records with UPNs in the CREST data were matched exactly. The remaining students were linked on other characteristics using 'fuzzy' matching techniques. Fuzzy matching does not require matches to be perfect, but instead looks at several fields such as name and date of birth to link records together. This can take a long time and requires manual checking. The way that records

were linked may have led to some selection bias, however it was not possible to test for this as we did not have access to the complete set of CREST records sent to the DfE by the BSA.

The DfE were able to link around 4,800 of the CREST Silver Award students to records in the NPD at Key Stage 4¹⁴ (KS4), with 4,500 linked to Key Stage 5¹⁵ (KS5) data (a subset of both of these groups had data available at both stages). DfE also supplied a complete extract of students who took their KS4 exams between 2009/10 and 2013/14. In total the datasets comprised of 3.2 million KS4 records and 3.8 million KS5 records.

3.4 Cleaning the data

The data received included duplicated observations (i.e. the same pupil appearing multiple times in the data), which were then removed. The data was then cleaned, with any records removed when information on control and/or outcome variables was not available. Students from independent schools were also removed at this point, as the NPD did not have all the necessary control variables for these pupils. Around 230,000 students (roughly 10% of the final KS4 sample size) were taken out as a result.

3.5 Final dataset

The final dataset used contained 2.4 million KS4 students, of whom 3,800 took Silver CREST, and 1.0 million KS5 records, of whom 2,300 had taken a Silver CREST Award.

¹⁴. Key Stage 4 covers the two years of schooling students usually undergo in school years 10 to 11 (ages 14 to 16). GCSE (or equivalent) exams are taken by students at the end of year 11 (age 15 or 16).

¹⁵. Key Stage 5 covers the two years of schooling students usually undergo in school years 12 and 13 (ages 17 and 18). A-level (or equivalent) exams are taken throughout this period; students typically pick 3 or 4 A-level subject choices.

3.6 Limitations

During the process of constructing the final dataset we (and DfE) discarded a number of records¹⁶. This was principally due to difficulties in linking the CREST data to the NPD, missing variables and duplicate records. In addition, KS5 data did not have variables for pupil characteristics and so needed to be linked to KS4 data. This was not possible for a surprisingly large proportion of students. If the records lost at each stage are biased in a certain way then there is a risk that the results will also be biased. It has not been possible to check for this bias as the linked data we received from DfE had already dropped a number of CREST records.

Our working assumption throughout much of the project was that CREST Silver Awards take place during KS4 (as stated in table 1.1) and therefore would be expected to take place prior to final exams and AS level subject selection. However, BSA analysis later showed 13% of the Silver CREST Award

records they had sent to DfE had birthdates suggesting students were 17 or over when they started the Award. We were not able to exclude these records from our analysis¹⁷, which will have a small impact on the implications of our results for our second and third research questions (see discussion in section 6.2).

It was also not possible to distinguish where students had taken part in multiple CREST Awards, such as Bronze and Silver or two Silver Awards. As a result, the estimates should be interpreted as the effect of taking 'at least one Silver CREST Award'. The BSA however believes it is very rare for a student to take multiple Silver Awards.

16. The 3,800 Silver CREST pupils with KS4 data in the final dataset represents around a third of the 11,700 records provided by DfE. The 2,300 Silver CREST pupils with both KS4 and KS5 data available represents around a fifth of that total.

17. Because the data we received had been anonymised, with dates of birth removed.

CREST definitely inspired me to work harder in STEM subjects and to aim high in further education

Jonathan, Cheshire

(Age 16 when he undertook his CREST Silver project which investigated the sensitivity of people's taste buds)

METHODOLOGY



4 METHODOLOGY

Linking the CREST data to the National Pupil Database provided a rich source of information to understand the characteristics of students starting Silver CREST Awards, and make inferences about the effect this had on their science attainment and STEM subject selection. For the latter two research questions we used Propensity Score Matching (PSM) to construct a control group. We have included a detailed discussion of why we chose this methodology, how we implemented it, and its limitations and benefits in comparison with other methods, in Annex C.

4.1 Propensity Score Matching (PSM)

PSM involves selecting a statistically-similar comparison group of students who did not receive 'treatment' - as it is termed in the literature. In this case, the treatment is CREST participation. The outcomes for the treated group (the CREST participants) are compared with the outcomes for the non-treated, but otherwise similar, group. Any differences between the outcomes of the treated and non-treated comparison group are interpreted as the "Average Treatment effect on the Treated" (ATT). It should be noted that any differences found between the two groups cannot automatically be assumed to apply to CREST participation by other cohorts of students or in other places. One key reason for this is that the delivery of the CREST programme may have changed. But there may be other changes that could affect the results.

The PSM methodology is deployed across a variety of academic fields including the education literature. Pingault et al (2015)¹⁸, used PSM to study the effect of maternal expectations on the probability of high school graduation in Canada and

Nagengast et al (2013)¹⁹ looked at the effects of single-sex schooling on broad educational outcomes. Anderson (2013)²⁰ also assessed the impact of HIV education on the behaviour of youths, finding some evidence that exposure to HIV education decreases risky sexual activity. Whilst the use of PSM within an education context seems to be widespread, we have not found any studies which use PSM to assess the efficacy of an inquiry-based learning programme.

PSM can only ensure that treated and control groups are similar on characteristics for which data are available. Unobserved characteristics, such as student enthusiasm, cannot be controlled for. This leaves scope for results to be biased due to unobserved differences between the treatment and control groups. As set out by Bryson, Dorsett & Purdon (2002)²¹, this is the main disadvantage of PSM relative to randomised control trial (RCT) methods. In principle, RCTs control for unobservable variables by assigning treatment at random. An RCT was not possible at this stage of research as we were working with historical data, but is recommended as a next step of analysis in this report (see section 7). OLS regressions are used as an alternative methodology to PSM to enable comparison of results between methods.

18. Pingault, J., Côté, S., Petitclerc, A., Vitaro, F. & Remblay, R. (2015) 'Assessing the Independent Contribution of Maternal Educational Expectations to Children's Educational Attainment in Early Adulthood: A Propensity Score Matching Analysis', *PLoS One*, 2015; 10(3).

19. Nagengast, B., Marsh, H. & Hau, K. (2013). 'Effects of single-sex schooling in the final years of high school : a comparison of Analysis of Covariance and Propensity Score Matching.'

20. Anderson, DM, (2013) 'The Impact of HIV Education on Behavior among Youths: A Propensity Score Matching Approach'

21. Bryson, A., Dorsett, R. & Purdon, S. (2002). 'The use of Propensity Score Matching in the Evaluation of Active Labour Market Policies: A study carried out on behalf of the Department for Work and Pensions'.

4.2 Outcome variables

The outcome variable for the second research question (science attainment at GCSE) was the highest scoring science GCSE. Keeping the indicator of science performance restricted to one science subject removed the difficulties of comparing performance across variable numbers of science GCSEs. Point scores relate directly to GCSE grades. Each grade is equivalent to 6 points, meaning that 2 points is equal to a third of a grade. An “A” grade is worth 52 points.

For the third research question (AS subject selection), the scope of the study was limited to non-vocational subject choices. Less common qualification types (e.g. IBs, BTEC) were not included. By limiting the study in this way, comparisons of students’ subject selections are less complex and more likely to be reliable. We created a binary indicator for whether the student had taken a STEM AS level which is a subset of students taking any non-vocational AS level.

For the purpose of this study, a broad definition of STEM was used, as agreed with the BSA. The DfE also suggested a narrower definition and the specifications for these are detailed below. Results according to this narrow definition are available in Annex B.

4.2.1 Limitations

There are some limitations associated with the choice of outcome variables. The decision to use the ‘highest’ science GCSE attained means that a student who got an ACC in triple science is treated the same as one who got AAA in triple science. Also by restricting our analysis to non-vocational subjects when studying impacts on STEM uptake at AS level, the interpretation of our results is narrower, but the validity is not affected.

4.3 Control variables

Information from descriptive statistics and the literature review was used to generate a list of relevant control variables. Both of these sources indicated that prior attainment was key. In this paper, we used KS2 SAT results. We found that together KS2 point scores in English, maths and science can explain 40% of the variation in GCSE point scores (see Annex B). KS2 point scores are commonly used as a measure of prior attainment in the education literature, for example in Blanden et al’s (2015) paper on the ‘London effect’²³. We also found that CREST students were more likely to have higher attainment at KS2.

23. Blanden, J., Greaves, E., Gregg, P., Macmillan, L., & Sibbels, L., 2015) ‘Understanding the improved performance of disadvantaged pupils in London’. Social policy in a Cold Climate: Working Paper 21, September 2015

Other variables controlled for included: gender, ethnicity, region of school, year of GCSEs, whether students have taken double or triple award science, type of school, whether the student had free school meal status in the six years prior to taking GCSEs, income deprivation affecting children index (IDACI)²⁴, whether or not students spoke English as a first language, and the special educational needs statuses of students. OLS regressions use the same control variables for research question 2²⁵, but additionally control for GCSE performance for research question 3²⁶.

4.3.1 Limitations

Controlling for prior attainment is important for this evaluation. Whilst KS2 results are commonly used they are also an imprecise measure. Assessments were taken around 5 years before GCSEs and contain limited detail of students’ performance. For instance in the data, students could achieve levels 2 to 5 and 83% of CREST students achieved level 5 in their KS2 science, demonstrating the lack of variation within the measure. As a result, prior attainment can be controlled for only to a limited extent. Results at KS3²⁷ would have given

a fuller picture of student’s prior attainment, but this data was not available for all students as exams at this level were abolished mid-way through the sample period.

There are also a number of factors that could not be controlled for at all in this study. These could have a significant effect on CREST participation and the outcomes being studied. They include: teacher quality, and enthusiasm for science; school quality, and enthusiasm for science; parental enthusiasm for science; student enthusiasm for science and whether or not the student had also participated in other CREST Awards. These have the potential to cause an upward bias in the estimates. As noted in section 4.1, to fully control for these a Randomised Control Trial (RCT) would be required.

24. The percentage of children aged 0-15 living in income-deprived households.
25. See [bsa.sc/CRESTimpact](https://www.bsa.sc/CRESTimpact) for tables and full list of control variables in regressions and corresponding regression coefficients.
26. See further discussion in section 6.2.
27. Key Stage 3 covers the three years of schooling students usually undergo from school years 7 to 9 (ages 11 to 14). Externally-marked exams used to be taken in English, maths and science by students at the end of year 9 (age 13 or 14), but these were stopped in 2008 in favour of internal assessments.

Table 4.1 Explanation of STEM and science definitions

Definition	Subjects included
GCSE science	Highest point score achieved (GCSE equivalencies) in science (Full GCSE, Full Intermediate or Foundation GNVQ and Vocational GCSE). Does not include maths or design and technology.
AS level – STEM Broad - BSA	Majority of GCE AS levels in science, maths and technology subjects including: biology, human biology, chemistry, physics, maths, further maths, statistics, additional maths, pure maths, electronics, environmental science, geology, public science, statistics, computer studies, graphics, textiles, 3D studies, food technology, design systems, and production design.
AS level – STEM Narrow - DfE	Biology, human biology, chemistry, physics, maths, further maths.

“CREST students are 21% (or 14 percentage points) more likely to take a STEM AS level than students in a matched control group”

5 RESULTS

5 RESULTS

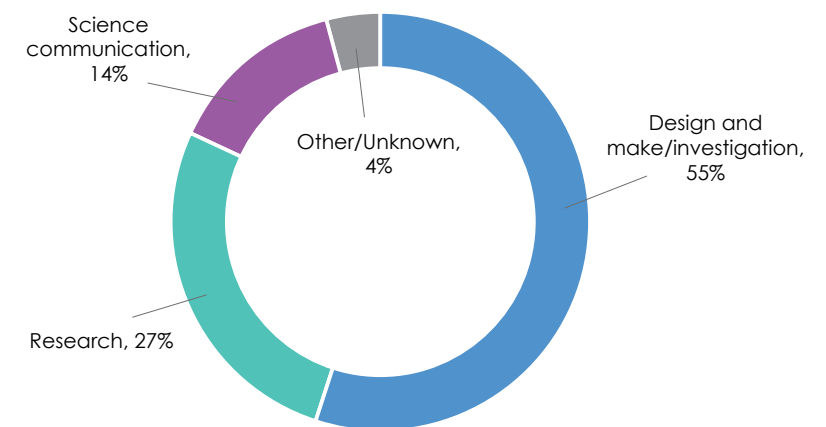
This section presents the results of the analysis, taking each of the three research questions in turn. The implications of these results are discussed in more detail in section 6.

5.1 RESEARCH QUESTION ONE: What are the characteristics of students taking Silver CREST Awards?

5.1.1 CREST projects

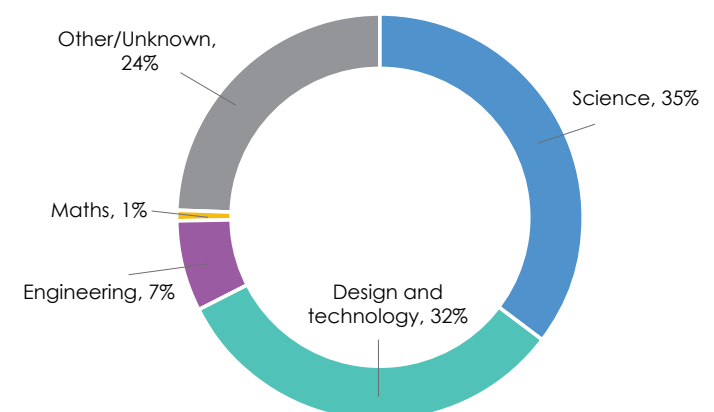
There are many different types of projects students can do when taking part in the Silver CREST scheme. Just over half of the students in our dataset focus on a 'design and make' investigation, with a quarter focusing on research. The rest undertook a project focused on science communication, with a small proportion of project focuses remaining unknown.

Chart 5.1 Silver CREST Awards 2010-2013 by project type



Where the project area was recorded most students (in approximately equal numbers) based their CREST project either on design and technology, or on science, with a smaller proportion looking at engineering, and a very small number of students undertaking a maths project.

Chart 5.2 Silver CREST Awards 2010-2013 by project area

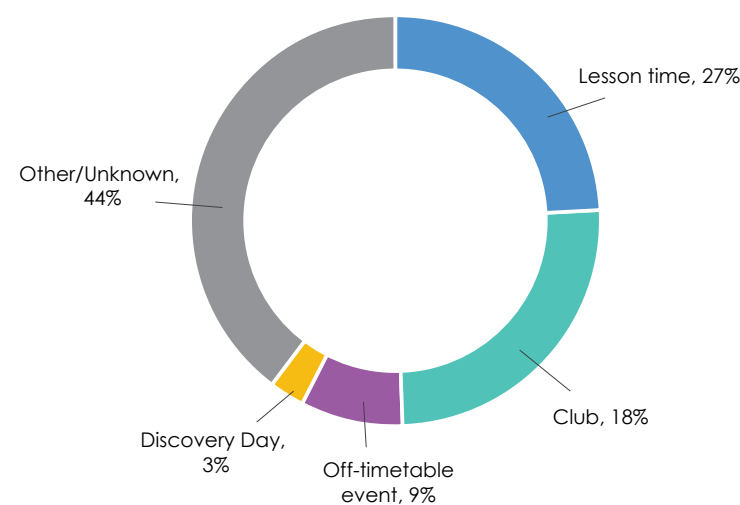


5 RESULTS

CREST projects can also be delivered in different ways, for example as part of a lesson time, or as an after school club. Unfortunately the delivery type is unknown for 44% of students in the data, meaning that this presents an incomplete picture of delivery type. We do know that 27% of the total did their Award as part of lesson time, 18% did it as part of a club, and 9% did it as part of an off-timetable event. A small number of students were recorded as doing the Award through a 'Discovery Day'; however, this option was not available to CREST students during this period. We believe these records have been incorrectly entered and as a result all results relating to project delivery type should be treated with caution.

Later in this analysis the effect of different delivery types on attainment is explored. However, these relationships should also be treated with caution due to the incomplete and unreliable nature of this data.

Chart 5.3 Silver CREST Awards 2010-2013 by project delivery type



5.1.2 Student demographics

The gender and ethnic breakdown of CREST students was similar to the larger non-CREST student population. Half (50%) of students taking Silver CREST Award were young women, reflecting the gender breakdown of the student population as a whole. This is notable, as the government and many other science bodies are interested in getting more women to take up and continue science careers and qualifications. However, as the extent of self-selection onto CREST is unclear, it is also not clear whether CREST is attractive to young women interested in science, or if it is the by-product of teachers deciding to do CREST with a whole class.

Chart 5.4 shows that the proportion of CREST students who are white is very similar to that of students who did not take Silver CREST. There are proportionately slightly more Asian and Chinese students who took CREST than we see in the wider student population, and slightly fewer students of black and mixed ethnicities.

However, in other ways CREST students were notably different from their peers. CREST students were less likely to have been eligible for Free School Meals in the last six years than other students (10% versus 22%) and less likely to have Special Educational Needs (7% versus 16%).

5 RESULTS

Chart 5.4 Ethnicity of CREST and non-CREST students (2009/10 to 2013/14)



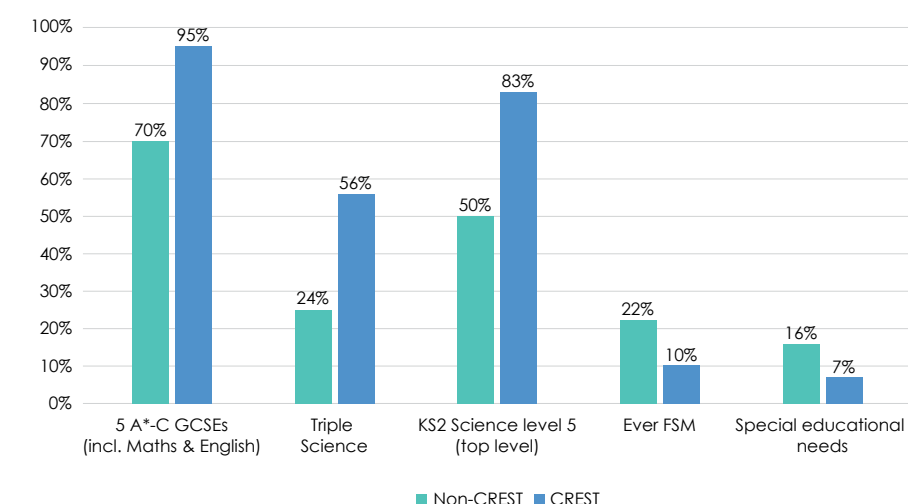
5.1.3 Student attainment

This section compares the attainment of CREST and non-CREST students. These differences cannot be taken as a measure of impact, as they will reflect differences in composition of the two groups of students. Research questions two and three look at some of the same differences whilst controlling for some of these factors.

The results for KS2 exams (taken at age 10-11) did not vary much across subjects. Students who had done well in KS2 science were likely to have done well in KS2 English and maths. In the data, students achieved levels 2 to 5 in their KS2 exams, with higher levels indicating higher achievement. CREST students performed higher than non-CREST students across all subjects. 83% of students who did CREST achieved level 5 in their KS2 science exam, compared to half of non-CREST students. A full breakdown of KS2 results can be seen in Annex B.

It is therefore not surprising that students who do CREST do better in their GCSEs, with 95% achieving the benchmark of 5 A*-C GCSEs including English and maths, compared to 70% of non-CREST students. Students who take Silver CREST Award are also much more likely to take, or be taking, triple science at GCSE (56% versus 24%).

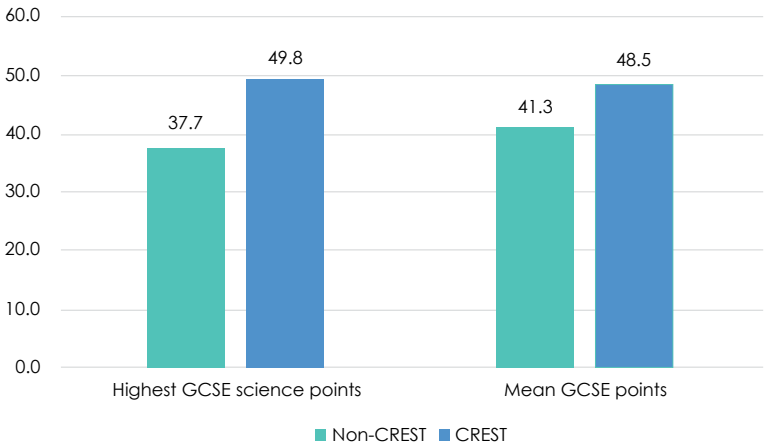
Chart 5.5 Comparing Silver CREST students with non-CREST students



5 RESULTS

The higher proportion of students getting 5 A*-C grades at GCSE among CREST students is also reflected in their point scores (which translates to grades, see section 4.2 for more details). Students who undertook Silver CREST Awards had higher average GCSE points than those who did not do CREST. Similarly, the number of points CREST students received for their highest science GCSE result was also higher.

Chart 5.6 GCSE points for CREST and non-CREST students



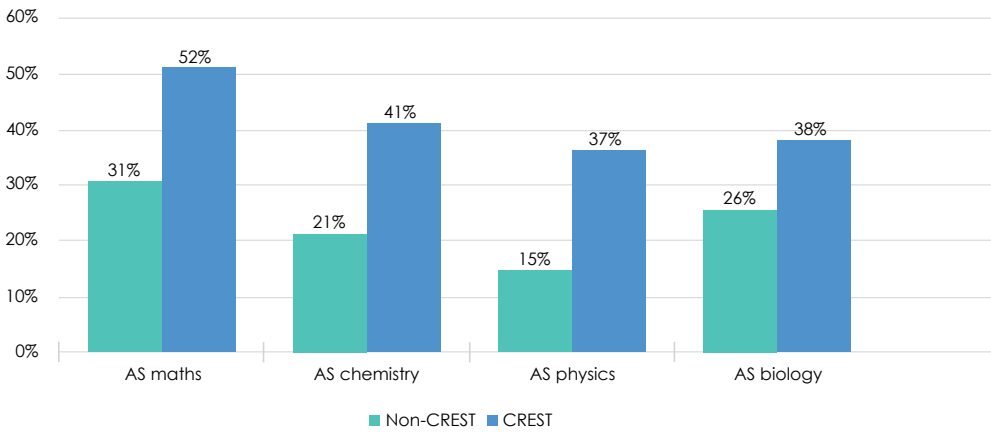
Given that CREST students: did comparatively well in their GCSEs, were more likely to take triple rather than double science, were less likely to be deprived, and exhibited higher achievement in science at KS2, it is not surprising we find a correlation between students who participated in Silver CREST Award and the likelihood of continuing with academic qualifications at Key Stage 5 and take-up of STEM AS levels.

The NPD has KS5 information about students who do AS levels, but does not contain information about some other post-16 qualifications such as apprenticeships. 94% of students who did Silver CREST Awards went on to take GCE AS level qualifications compared to 77% of students who did not take CREST Awards (where their KS5 outcome was known). In this sample there were complete records for analysis of 2,300 students who took CREST and around 980,000 students who did not take CREST.

The CREST students who took academic A level qualifications were far more likely to have undertaken science or maths at AS level, as demonstrated in the next chart. The difference is particularly stark in the take up of AS level physics, which 37% of CREST students took (of those taking AS levels), compared to 15% of students taking AS levels overall. However, this comparison does not take into account the different characteristics of the two groups of students; this is covered in research question three.

5 RESULTS

Chart 5.7 Take up of certain STEM subjects at AS level



5.2 RESEARCH QUESTION TWO: Does participation in the Silver CREST Award programme have an impact on attainment in science subjects at GCSE?

It is clear from the above analysis that CREST students have different characteristics from the student population as a whole, and do better academically. This section looks at the relationship between participation in CREST and students' highest points score for their science GCSEs, using PSM to construct a statistically matched control group.

As discussed above, the average CREST Silver student achieved 49.8 points on their highest GCSE science grade, 12.1 points higher than the average non-CREST student in the data - this is equivalent to around two GCSE grades. However, this difference does not take into account the systematic difference between the characteristics of students who do and do not take CREST.

Once we had constructed a control group using PSM (as detailed in the methodology section and Annex C) the difference between the average GCSE point score of CREST and non-CREST students reduced from 12.1 points, to 3.3 points (see table below). Therefore, compared to a matched control group, students who took a Silver CREST Award achieved half a grade higher on their best science GCSE result. This finding was statistically significant. The control group was matched for prior attainment (using KS2 SATs results), gender, ethnicity, region of school, year of GCSEs, whether students have taken double or triple award science, type of school, whether the pupil ever had free school meal status, income deprivation affecting children index (IDACI), whether or not students spoke English as a first language, and the special educational needs statuses of students.

Table 5.8 PSM results for effect of CREST participation on highest science GCSE point score

	Mean highest science GCSE point score		Difference	Standard Error
	CREST students (N=3,774)	Control (N=2,363,679)		
Unmatched	49.8	37.8	12.1*	0.3
PSM matched control	49.8	46.5	3.3*	0.2

Notes: *=statistically significant at the 5% significance level

The results were quality assured against a standard OLS linear regression, which also found a difference of around half a grade. The results of the OLS regression can be found in Annex B.

5 RESULTS

5.2.1 By gender

We re-ran the PSM process, this time breaking down the sample of CREST students by gender. There was a similar statistically significant effect for both men (3.5 points) and women (3.2 points). However, regression results suggested that any gender differences were not statistically significant (for more detail see Annex B).

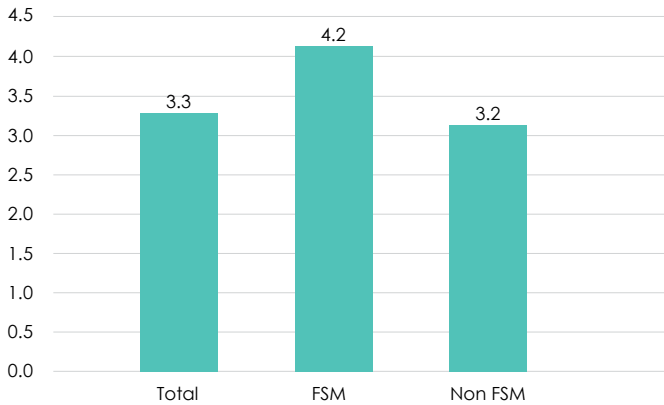
5.2.2 By project delivery type

We also wanted to understand if the CREST effect varied by project delivery type. Students who did CREST in a club may be more likely to have elected to do so than students who did CREST as part of a class, where the decision to participate may have been taken by a teacher. Differences in attainment by delivery type could indicate that students who did CREST through a club were more likely to volunteer and therefore be more motivated. PSM results showed the effect of clubs turned out to be stronger, though the difference was relatively small. The average CREST effect size on a student's highest science GCSE was 3.9 points when they did it in a club compared to 3.2 points when they did it as part of lesson time. The impact of CREST was statistically significant for both delivery types individually, but regression results suggested that any difference in effect size was not statistically significant. As noted in section 5.1.1, results relating project delivery type should be treated with caution due to data limitations.

5.2.3 By free school meals status

Students are eligible for free school meals (FSM) at secondary school when their family receives a qualifying income related benefit. The CREST students in this study who were eligible for free school meals at any point in the six years before their GCSEs had a larger difference in their highest science GCSE points score (4.2 points) than those CREST students who had not been eligible for free school meals (3.2). This means in this study, the impact of CREST on non-FSM students was around half a GCSE grade, but around two thirds of a grade for students who had been eligible for free school meals (both findings were statistically significant individually). Regression results suggested a similar difference in effect sizes, and, importantly, that this difference was statistically significant (for more detail see Annex B). Having said this, due to the small proportion of CREST pupils who were eligible for FSM (n=382), these results should be treated with caution.

Chart 5.9 Differences in highest science GCSE points score between CREST students and statistically matched control groups, split by free school meals eligibility



5 RESULTS

5.3 RESEARCH QUESTION THREE:
Does participation in the Silver CREST Award affect the likelihood of taking a STEM AS level?

The third and final research question in this study looks at the proportion of Silver CREST students who went on to take a STEM AS level (as a proportion of all CREST students who chose to do academic AS levels) relative to a similar group of students who did not take CREST. This uses a wider definition of STEM for this section; details of the narrower definition can be found in Annex B .

Overall, 82% of people who did Silver CREST and went on to take academic AS levels chose a STEM subject, compared to 55% of students who did not take CREST. Again, one would expect CREST students to be more likely to take a STEM subject because of their different characteristics, primarily their better KS2 attainment in science. However, when we construct a statistically similar control group of non-CREST students using PSM we find 68% take a STEM subject, still 14 percentage points less than the CREST students. This means that CREST students are 21% (or 14 percentage points) more likely to take a STEM AS level than students in a matched control group.

Table 5.10 PSM results for effect of CREST participation on STEM uptake at AS level

	Percentage of students taking STEM AS levels		Absolute difference (% points)	Standard Error	Percentage difference
	CREST students (N=2,332)	Control (N=975,922)			
Unmatched	82%	55%	26*	0.3	47%
PSM matched control	82%	68%	14*	0.2	21%

Notes: *statistically significant at the 5% significance level

These results are similar to those we recorded from OLS regressions (using both linear probability and logit models) on take up rates. A summary of those results can be found in Annex B.

5.3.1 By gender

When we examined attainment, there was no significant difference in effect size between males and females. For subject choice CREST appears to have a marginally stronger impact on take-up of a STEM AS level for female students (15 percentage points) than for male students (13 percentage points). As the take-up of STEM AS levels is lower among women than for men, the proportionate effect of this is larger. It is important to note however, that regression results were inconclusive as to whether or not the gender difference was statistically significant (tables in Annex B show linear probability and logit models to give contrasting results).

Table 5.11 PSM results for effect of CREST participation on STEM uptake by gender (summary table – full table available in Annex B)

	STEM uptake			Difference in uptake between CREST and matched control	
	CREST students	Non-CREST students	Matched control group	Absolute difference (% points)	% difference
Overall	82%	56%	68%	14*	21%
Men	89%	65%	76%	13*	16%
Women	74%	47%	59%	15*	25%

Notes: *statistically significant at the 5% significance level

5.3.2 By project delivery type

Looking at STEM AS level take up, the results by project type show a similar pattern to those seen in attainment. The CREST effect size is almost twice as big for students who did their CREST Awards as part of a club (18 percentage points) as it is for those did them in lesson time (10 percentage points). The impact of CREST was statistically significant for both delivery types individually and regression results suggested that the difference between the impacts of each delivery type was statistically significant. As noted in section 5.1.1, results relating project delivery type should be treated with caution due to data limitations.

5.3.3 By free school meals status

On average, the difference in AS level take up between the Silver CREST pupils and those who did not take CREST was 14 percentage points (or 21%). However, this difference was larger for students who had been eligible for free school meals in the six years before they participated in CREST.

CREST students who were eligible for FSM had a 78% take-up of STEM AS levels compared to 57% in the statistically matched control group of non-CREST FSM eligible students. This is a 21 percentage point (or 38%) difference. The impact of CREST was statistically significant for both FSM and non-FSM students individually and regression results showed a statistically significant difference in effect sizes on each FSM status. However, as noted in section 5.2.3, due to the small proportion of CREST pupils who were eligible for FSM (n=193), these results should be treated with caution.

Table 5.12 PSM results for effect of CREST participation on STEM uptake split by FSM status (summary table – full table available in Annex B)

	AS level STEM uptake			Difference in uptake between CREST and matched control	
	CREST students	Non-CREST students	Matched control group	Absolute difference (% points)	% difference
Overall	82%	56%	68%	14*	21%
FSM	78%	48%	57%	21*	38%
Non FSM	82%	57%	69%	13*	19%
Notes: *statistically significant at the 5% significance level					

“
Half (50%) of students taking Silver CREST Award were young women
”

DISCUSSION



6 DISCUSSION

6.1 Interpreting the results

The robustness of these results is around level three in the Nesta Standards for Evidence framework, as discussed in section 2 (for a full description of the standards and levels, see Annex A). We were able to establish robust control groups and use large sample sizes. This is in addition to cross checking PSM results with OLS regressions. We were also able to break down the results by project type, gender and free school meal status. These breakdowns indicate that delivery of CREST through a club seemed to have a stronger impact than delivery during lesson time (although regression results suggested that the difference in effect sizes on GCSE science attainment was not statistically significant). The club result may be unsurprising, as the BSA suggests that this delivery type is often opt-in, catering for enthusiasts, and will sometimes be selective if demand is high. The impact of Silver CREST on science attainment seems relatively consistent across genders but on STEM participation may be higher for girls than boys³¹.

The finding that the impact on students who have started Silver CREST and have been eligible for free school meals (FSM) was higher than for non-FSM eligible CREST students is likely to be of significant interest to policy makers and schools. The result is in line with the result found by Marsh that the impact of extracurricular activities is more pronounced for those from families with lower socioeconomic status³². However, the FSM result must still be treated with caution, as it is possible that there are missing variables biasing the result. For example, it may be that teacher enthusiasm and quality has particularly pronounced effects on students eligible for free school meals and this is affecting both participation in CREST and AS level choices. The sample size of FSM eligible students is also relatively small.

It is important to note that our PSM analysis for our second and third research questions can only be interpreted as the average treatment effect of CREST on those pupils who took Silver CREST Awards (the treated) during the period we studied. We cannot be certain the impact would be the same if the treatment was given to all students. This means our PSM analysis shows what is known as the "Average Treatment effect on the Treated", not the "Average Treatment Effect".

6.2 Demonstrating causality

Despite these results, proving causality remains complicated. In order to demonstrate causality three conditions must be met:

- Temporal condition: the treatment must have occurred before the outcome. If we tried to claim for example, that participation in Silver CREST (at age 15/16) increased a student's Key Stage 2 results (at age 11) then this condition clearly would not be met.
- Co-variation condition: there must be a demonstrable statistical link between the presence of the treatment (Silver CREST), and the impact on the outcome (GCSE science attainment and AS level subject choice).
- No plausible alternative: this is key; a good research design will allow the researcher to rule out, as far as possible, other explanations for the results (e.g. selection bias) so deductive reasoning can be used to conclude that the treatment has caused the outcome.

³¹. As noted in section 5.3.1, regression results including interaction terms were inconclusive as to whether there was a statistically significant gender difference in effect on STEM participation.

³². Marsh, H. (1992) Extracurricular activities: Beneficial extension of the traditional curriculum or subversion of academic goals? *Journal of Educational Psychology*, 84(4), 553–562.

6 DISCUSSION

This research meets the first condition for the majority of students as most participants should have undertaken their CREST projects prior to taking their GCSE exams and making final decisions on their AS level subjects. However, analysis from BSA has shown that around 13% of Silver CREST participants completed the award during KS5 (as discussed in section 3.6). This leaves scope for reverse causality, as student's GCSE results and AS-level choices could affect their selection onto Silver CREST. This may mean that estimates of the effect of CREST are overstated.

To test the extent of this issue we reran our analysis on Silver CREST students who took part in the award between 2010 and 2013 and took their GCSEs in 2013/14, and therefore could not have taken Silver CREST in KS5. For this subset we found a slightly smaller difference in highest GCSE science compared to the control group (3.0 compared to 3.2 points) - however, this is still equal to a difference of a half a grade. Due to the way the cohorts worked in our data it was not possible to replicate this approach for our third research question (STEM subject selection), but our analysis suggested that the cohorts which were likely to have a smaller proportion of pupils who took CREST after making AS level choices had a similar, but slightly smaller effect size.

Similar issues are present if GCSE module exams or coursework are taken prior to CREST participation and make a significant contribution towards students' highest Science GCSE grade. This is not something we are able to control for in our analysis and the BSA suggests that CREST projects are usually scheduled to avoid clashing with coursework or exam preparation.

The second condition is met. The PSM analysis showed that among the students in the sample, CREST students were more likely to take STEM AS levels than those who didn't take CREST and had higher scores in their best GCSE science subjects.

OLS regression analysis also showed that this link was statistically significant.

However, the third condition is trickier. Even if the first two conditions are met, other factors that could explain the link between the treatment and outcome need to be ruled out. If it is possible to perfectly construct a control group, with no missing or unobserved variables then the researchers are likely to be able to say that the treatment has caused the outcome. This is almost never possible with statistically matched control groups. This is why Randomised Control Trials are considered the gold standard for causality, because if they are carried out well the only difference between the two groups is the treatment.

There are three reasons why our attempts to control for all other 'plausible' factors may have been unsuccessful:

- Unobserved characteristics. We cannot reject the hypothesis that there were other important variables that affected the likelihood of doing CREST and had an impact on GCSE attainment and AS level subject choice. These variables could include the quality and enthusiasm of the school and teachers, in general and specifically for science. This problem is similar to those mentioned in the broader literature on school-based interventions such as Lui, Lee and Linn (2010) and Marsh (1992).
- Weaknesses in our prior attainment variable. As mentioned in section 4, results at KS2 are an imprecise indicator of prior attainment, an important control variable. 83% of CREST students received level 5 in their KS2 Science exam, compared to around half of non-CREST students. This suggests a lack of granularity in the variable, therefore it was not possible to distinguish (and match) the very high achieving students.

6 DISCUSSION

- The way PSM works did not allow us to take into account GCSE results in our third research question. The matching in PSM works by giving every student a score for his or her likelihood to take up Silver CREST. This score is based on a regression that looks at various control variables, however, GCSE results can't be used as a control variable because they occur after a student starts their Silver CREST Award. In our PSM analysis we found the difference in STEM take up rates at AS levels between CREST and non-CREST pupils to be 14 points. When we checked this in a OLS regressions which did control for GCSE results, this found a smaller effect size – 9 percentage points ³³.

Our initial hypothesis was that CREST would have an impact on GCSE science attainment. Interestingly, we also find evidence that it is linked to an increase in overall GCSE attainment – as shown in chart 6.1.

There are two possible explanations for this. The first is we are not fully able to control for variables such as intelligence, pupil and teacher motivation and school quality, which means our results would still be biased from these factors. This interpretation is in line with the limitations mentioned by

authors Lui, Lee and Linn (2010) and Marsh (1992). The second explanation is that CREST is having a broader effect than just on science subjects. This could be because the experience of project based learning gives students skills which they can apply to their other studies. This would be in line with Philip Adey's work on Cognitive Acceleration through Science Education (CASE) in the mid-1990s ³⁴. Adey finds that CASE method interventions have an impact on students' cognitive gains across subjects. It is also consistent with the result found by Metsäpelto and Pulkkinen that non-academic extracurricular activities increased academic attainment ³⁵.

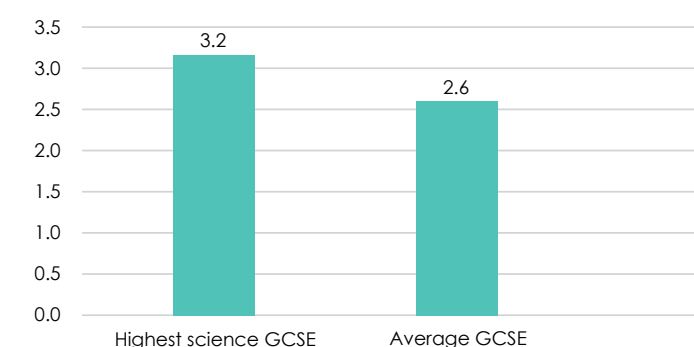
The impact of CREST on overall GCSE attainment might be an indicator that we have not managed to meet the third condition for causality mentioned above. Or it may be that CREST has a wider benefit. The data does not include enough variables to be able to discriminate between either explanation.

33. 9 percentage points in logit models, 8 percentage points in linear probability model.

34. Adey, P. S. (1995) 'The Effects of a Staff Development Program: The Relationship between the Level of Use of Innovative Science Curriculum Activities and Student Achievement.'

35. Metsäpelto, R., & Pulkkinen, L. (2012) Socioemotional Behavior and School Achievement in Relation to Extracurricular Activity Participation in Middle Childhood. *Scandinavian Journal Of Educational Research*, 56(2), 167-182.

Chart: 6.1 OLS regression showing impact of CREST on highest science GCSE points score and average GCSE points score



6.3 Wider applicability of results

This analysis solely looked at Silver CREST Awards, but our results suggest that this type of programme might also have an impact at Bronze and Gold level. Moote's quasi experimental study on Bronze CREST Awards, which found an increase in motivation and self-regulated learning, supports this conclusion³⁶. Finding an impact on GCSE attainment suggests that a similar result may be found at Key Stage 5 attainment (i.e. A levels) as a result of participating in Gold. Gold is a longer-lasting project than Silver, so we may expect to see a larger impact on attainment.

If similar results are found using more robust methods e.g. in a Randomised Control Trial then there may be policy implications. Schools might want to consider whether CREST could be part of a solution to boost GCSE science results. The Department for Education might want to consider what this suggests for the importance of project-based work for both increasing attainment and for encouraging the uptake of STEM subjects in sixth form. This research has taken a broad definition of STEM and included all science and technology GCE AS level subjects. The Department for Education have indicated that they define STEM more narrowly, including only maths and core sciences. We additionally ran PSM for the narrow definition and found a significant but somewhat smaller result. See Annex B for results under each definition.

There are two main issues that complicate the above policy implications. The first is about the intervention itself, and the second is about how it is used in schools.

CREST in this research is treated as if it is a single intervention, always carried out in the same way. However, different teachers

may approach CREST with very different levels of enthusiasm. The literature suggests that some teachers find it time consuming and complicated: for instance the CREST expansion evaluation found that "CREST teachers who have made CREST work for them have experienced strong personal benefits. For teachers with a more low-key involvement, it has seemed to be a lot of hard work with less dramatic results"³⁷. This means that the results we have found may not be replicable if teachers were compelled to run CREST projects.

The decision to do CREST can be made at school, class or pupil level. In the sample of Silver CREST students 27% were recorded as having been delivered CREST during lesson time, with 18% in a club and 44% unknown. At the moment CREST is voluntary, so someone at school level, teacher level, or pupil level has made a decision to do it. We cannot rule out that some of our impact results are due to this self-selection. This is a common issue with evaluating educational interventions that are to some extent extracurricular and is cited as an issue in other research, for example Marsh³⁸. It means that wider roll out, especially if it was mandatory, may reduce the impact in comparison with the results in this report.

It should also be noted that our results cannot be applied to independent schools as these were excluded from this analysis.

36. Moote, J. (2014) 'When Students Lead: Investigating the Impact of the CREST Inquiry-Based Learning Programme on Changes in Self-Regulated Processes and Related Motivations Among Young Science Students', PhD University of Edinburgh

37. Fisher S, Tifford R & Gammon B (2011) CREST Expansion Scheme 2009-2011, London: British Science Association

38. Marsh, H. (1992) Extracurricular activities: Beneficial extension of the traditional curriculum or subversion of academic goals? *Journal of Educational Psychology*, 84(4), 553-562.



This analysis solely looked at Silver CREST Awards, but our results suggest that this type of programme might also have an impact at Bronze and Gold level



RECOMMENDATIONS



7 RECOMMENDATIONS

There are seven key recommendations for future research into CREST and three wider recommendations that stem from this study. The recommendations into future analysis are:

1) This research should be replicated with a randomised control trial (RCT)

An RCT would satisfy the third condition of causality - that all other variables which might have caused the result had been ruled out - and would allow the estimation of the Average Treatment Effect. This means that the findings would be applicable to the wider student population. It would also represent a higher level of Nesta evidence standards. Thought would need to be given to the research design, including at what level to randomise (pupil, class, school) and how to account for the degree of voluntarism in the project. The RCT may be best focusing on CREST delivered through lesson time to whole classes, rather than the after school club type of delivery.

2) A full cost-benefit analysis of running CREST programmes should be conducted for schools

A full cost-benefit analysis would take into account both in terms of the direct financial costs, but also the costs of staff time versus the benefits as identified in this research and other qualitative studies. This would help schools make a judgement about the relative value of investing time and resources into CREST compared to other science interventions.

3) Similar research should be carried out to assess the impact of Discovery, Bronze and Gold CREST Awards

The Gold Award would be of particular interest due to the higher number of hours students are on the programme. It could be interesting to link the students A Level and Higher Education data to see if there was an impact on the choice of university subjects. It would also allow for the use of GCSE results as a more robust proxy for academic ability, as opposed to the limited range of KS2 results that were used in this research.

4) Analysis should be extended to Scotland, Northern Ireland and Wales

Due to the coverage of the NPD, this analysis has looked only at England, but CREST is available across the UK (and internationally, for example in Australia).

Education is a devolved policy area in the UK, so the education systems can vary. However, administrative data on students is still collected, so this sort of research using statistically matched control groups should be possible outside of England too.

5) The BSA should consider collecting consistent outcome data on awards

There is a certification process once students have finished their CREST project. This is undertaken by someone external to the school, usually the local coordinator. However, the information on certification is sometimes held locally, and not always supplied back to the BSA. The BSA may wish to collect this information on a more consistent basis. This would provide useful information for future evaluators, as they would be able to look at the impact of CREST on those who start and those who complete. It may also highlight schools with particularly high rates of students who failed to complete, indicating the potential need for more support. BSA will roll out a Digital CREST platform in 2017 and should consider how they can use this to maximise accurate and complete data collection.

6) The BSA may want to consider some standardisation of CREST

Students can undertake CREST Awards in a variety of different ways, including through schemes linked to industry - and the BSA argue there are considerable benefits to this approach. If the BSA wanted to focus the awards on increasing attainment or future STEM subject choice then they may want to consider standardising elements of the approach in order to be able to test the effectiveness of different models.

7) Research into the current delivery of CREST

The BSA already collects information about students and schools doing CREST. Whilst this information is not always complete, the BSA may want to consider how they can use the data they are already collecting to track information about the types of schools which are taking up CREST, and how this is changing over time. Qualitative research, which could be quite light touch, may also shed light on who is making decisions to do CREST and why (teachers and pupils). This understanding may help the BSA increase take up of the awards.

7 RECOMMENDATIONS

Our report also leads us to some broader recommendations. On the basis of our research we recommend that:

8) Charities should ensure their data collection is as complete as possible and their Data Protection statements allow for research using their data

If charities wish to use their datasets for research or evaluation purposes, linking to other datasets can be crucial. This analysis was only possible because it was possible to link the BSA data with the NPD. In order to link across datasets it is important to collect certain identifiers. Examples include: full name, date of birth, address and unique identifiers such as national insurance numbers (NINOs) or UPNs. Charities should ensure that they collect these and that their data protection statements support the linking and analysis of data for research purposes.

9) Young people consider participating in project/inquiry-based learning programmes such as CREST as part of their education

This report supports the findings of Philip Adey, Alicia Cotabish and others that project or inquiry-based science interventions can improve attainment outcomes. As such, students wishing to develop their skill set should consider participating in a programme like the Silver CREST Award.

10) The BSA consider the case for targeting CREST at students from low income families

This research suggests that Silver CREST Awards have the largest impact on attainment and uptake of STEM subjects for students who have at one time been eligible for free school meals. One tenth of students who participated in Silver CREST between 2010 and 2013 either were currently, or had previously been eligible for free school meals. This is about half as many as in the wider student population (just over one fifth). If the BSA were interested in maximising the impact that Silver CREST Awards had on attainment or STEM participation, assuming the effects of CREST estimated in this study apply more generally to those students not included in the study, then they may want to consider targeting this cohort.

To download a full copy of this report, please visit: bsa.sc/CRESTImpact



This research suggests that Silver CREST Awards have the largest impact on attainment and uptake of STEM subjects for students who have at one time been eligible for free school meals



BIBLIOGRAPHY & ANNEXES



BIBLIOGRAPHY

Adey, P. S. (1995) 'The Effects of a Staff Development Program: The Relationship between the Level of Use of Innovative Science Curriculum Activities and Student Achievement.' <http://files.eric.ed.gov/fulltext/ED383567.pdf>, ERIC

Anderson, DM. (2013) 'The Impact of HIV Education on Behavior among Youths: A Propensity Score Matching Approach', *Contemporary Economic Policy*, 31: 503–527

Blackwell et al (2009) 'Coarsened exact matching in Stata', *The Stata Journal*, Vol. 9, No. 4, pp. 524–546.

Blanden, J., Greaves, E., Gregg, P., Macmillan, L., & Sibiet, L., (2015) 'Understanding the improved performance of disadvantaged pupils in London', *Social policy in a Cold Climate: Working Paper 21*, September 2015

Bryson, A., Dorsett, R. & Purdon, S. (2002) 'The use of Propensity Score Matching in the Evaluation of Active Labour Market Policies: A study carried out on behalf of the Department for Work and Pensions' DWP

Conlon, G. & Patrignani, P. (2015) 'The earnings and employment returns to A levels: A report to the Department for Education', *London Economics*, <http://london-economics.co.uk/wp-content/uploads/2015/03/London-Economics-Report>Returns-to-GCE-A-Levels-Final-12-02-2015.pdf>

Cotabish, A., Dailey, D., Robinson, A. and Hughes, G. (2013) 'The Effects of a STEM Intervention on Elementary Students' Science Knowledge and Skills', *School Science and Mathematics*, 113: 215–226. doi: 10.1111/ssm.12023

Crown, W. (2014) 'Propensity-Score Matching in Economic Analyses: Comparison with Regression Models, Instrumental Variables, Residual Inclusion, Differences-in-Differences, and Decomposition Methods', *Applied Health Economics and Health Policy*, 12: 7–18.

Dehejia and Wahba, (1999) 'Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs', *Journal of the American Statistical Association*, Vol. 94, No. 448. (Dec., 1999), pp. 1053–1062

Finkelstein, F. Hanson, T. Huang, C. Hirschman, B. & Huang, M. (2010) 'Effects of Problem based Economics on high school economics instruction'

Fisher S, Tifford R & Gammon B (2011) *CREST Expansion Scheme 2009-2011*, London: British Science Association

Grant, L. (2007) 'CREST Awards evaluation: Impact study', *University of Liverpool Science Communication Unit*

Hendry, E. (2013) 'CREST Awards and Employability Skills: A project for the British Science Association', <http://www.techfestsetpoint.org.uk/uploads/files/CREST%20Employability%20Report%20-%20Nov13.pdf>, emhconnect

Ji-Wei, W & Tseng, J. (2015) Gwo-Jen, H. 'Development of an Inquiry-Based Learning Support System Based on an Intelligent Knowledge Exploration Approach'. *Journal of Educational Technology & Society*. 18, 3, 282–300, July 2015

Lui, Lee & Lin (2010) 'Multifaceted Assessment of Inquiry-Based Science Learning', *Educational Assessment*, 15, 69–86.

Marsh, H. (1992) *Extracurricular activities: Beneficial extension of the traditional curriculum or subversion of academic goals?* *Journal of Educational Psychology*, 84(4), 553–562.

Metsäpelto, R., & Pulkkinen, L. (2012) *Socioemotional Behavior and School Achievement in Relation to Extracurricular Activity Participation in Middle Childhood*. *Scandinavian Journal Of Educational Research*, 56(2), 167–182.

Moote, J. (2014) 'When Students Lead: Investigating the Impact of the CREST Inquiry-Based Learning Programme on Changes in Self-Regulated Processes and Related Motivations Among Young Science Students', PhD University of Edinburgh

Nagengast, B., Marsh, H. & Hau, K. (2013) 'Effects of single-sex schooling in the final years of high school: a comparison of Analysis of Covariance and Propensity Score Matching.' <http://link.springer.com/journal/11199/v46/n10/sex-roles>, October 2013, Volume 69, <http://link.springer.com/journal/11199/69/7/page/1> \h Issue 7, pp 404–422

Pingault, J., Côté, S., Petitclerc, A., Vitaro, F. & Remblay, R. (2015) 'Assessing the Independent Contribution of Maternal Educational Expectations to Children's Educational Attainment in Early Adulthood: A Propensity Score Matching Analysis', *PLoS One*. 2015; 10(3)

Puttick, R & Ludlow, (2013) 'Standards of evidence: an approach that balances the need for evidence with innovation', https://www.nesta.org.uk/sites/default/files/standards_of_evidence.pdf, Nesta

Rubin, D. B. (2001) *Using propensity scores to help design observational studies: Application to the tobacco litigation*. *Health Services and Outcomes Research Methodology*, 2, 169–188.

Rubin, D.B. & Thomas, N. (1996) *Matching Using Estimated Propensity Scores: Relating Theory to Practice*, *Biometrics*, Vol. 52, No. 1 (Mar., 1996), pp. 249–264

Shoring, N. (2000) 'Evaluation of the CREST Award programme in Australia', *Australian Science Teachers' Journal*, v46 n2 p24–27 Jun 2000.

Smith and Todd, (2005) 'Does matching overcome LaLonde's critique of nonexperimental estimators?' *Pier Working Paper No. 01-035*. Available at SSRN: <http://ssrn.com/abstract=286297> or <http://dx.doi.org/10.2139/ssrn.286297>

ANNEX A - Nesta Standards for Evidence

Nesta Standards for Evidence are based on the framework of standards for evidence-based medicine and are now used in most areas of social policy (e.g. the Maryland scale). Nesta's framework is a version that is particularly adapted for helping charities understand their impact³⁹.

Level	Our expectation	How the evidence can be generated
At Level 1	You can give an account of impact. By this we mean providing a logical reason, or set of reasons, for why your intervention could have an impact and why that would be an improvement on the current situation.	You should be able to do this yourself, and draw upon existing data and research from other sources.
At Level 2	You are gathering data that shows some change amongst those receiving or using your intervention.	At this stage, data can begin to show effect but it will not evidence direct causality. You could consider such methods as: pre and post-survey evaluation; cohort/panel study, regular interval surveying.
At Level 3	You can demonstrate that your intervention is causing the impact, by showing less impact amongst those who don't receive the product/service.	We will consider robust methods using control group (or another well justified method) that begin to isolate the impact of the products/service. Random selection of participants strengthens your evidence at this Level, you need to have sufficiently large sample at hand (scale is important in this case).
At Level 4	You are able to explain why and how your intervention is having the impact you have observed and evidenced so far. An independent evaluation validates the impact. In addition, the intervention can deliver impact at a reasonable cost, suggesting that it could be replicated and purchased in multiple locations	At this stage, we are looking for a robust independent evaluation that investigates and validates the nature of the impact. This might include endorsement via commercial standards, industry kitemarks etc. You will need documented standardisation of delivery and processes. You will need data on cost of production and acceptable price points for your (potential) customers.
At Level 5	You can show that your intervention could be operated up by someone else, somewhere else and scaled up, whilst continuing to have positive and direct impact on the outcome, and whilst remaining a financially viable proposition.	We expect to see use of methods like multiple replication evaluations; future scenario analysis; fidelity evaluation.

39. See: www.Nesta.org.uk/publications/Nesta-standards-evidence

ANNEX B - Further tables

For the full set of tables please see the 'Detailed Tables' file available at www.bsa.sc/CRESTImpact

Table B1 OLS regression of average GCSE points on KS2 points

Average GCSE points	Coefficient	Standard Error.
KS2 English point score	0.64*	0.0014
KS2 Maths point score	0.47*	0.0014
KS2 Science point score	0.33*	0.0017
Constant	0.26*	0.042
R-squared	0.40	
Observations	1,922,700	

Notes: * statistically significant at the 5% significance level.

Table B2 Comparison between CREST and non-CREST students (for all students with data available at KS4)

Characteristic	Non-CREST	CREST
	% (unless stated)	
Highest GCSE science points (mean)	38	50
GCSE points (mean)	42	49
5 A*-C GCSEs (incl. Maths & English)	70	95
Triple science	24	56
KS2 science level		
2	0	0
3	6	1
4	44	17
5	50	83
KS2 maths level		
2	1	0
3	15	3
4	48	26
5	36	71
KS2 English level		
2	0	0
3	12	2
4	54	33
5	33	65

ANNEX B - Further tables

Table B2 Comparison between CREST and non-CREST students (for all students with data available at KS4) - CONT.

Characteristic	Non-CREST	CREST
	% (unless stated)	
Year of GCSEs		
2010	19	10
2011	21	12
2012	21	23
2013	21	29
2014	18	27
Female	50	50
IDACI (mean)	0.21	0.14
Ever FSM (FSM eligibility in the past 6 years)	22	10
Non-English first language	10	10
Ethnicity		
White	83	82
Asian	7	9
Black	4	3
Chinese	0	1
Mixed	4	3
Any other ethnic group	1	1
Unclassified	1	1
Special educational needs	16	7
Type of School		
Community	36	19
Voluntary	15	18
Foundation	20	23
Academy	29	39
Other	0	0
Region		
London	12	13
South East	15	11
South West	10	18
East Midlands	7	13
West Midlands	11	7
East	12	11
Yorkshire & Humber	10	3
North East	5	5
North West	14	15
Observations	2,366,151	3,774

ANNEX B - Further tables

Table B3 PSM results for effect of CREST participation on highest GCSE science score broken down by different groups

Population		Number of observations	Difference		Standard error		Mean Bias
			Unmatched	ATT	Unmatched	ATT	
Total	Total	2,367,453	12.08	3.32*	0.26	0.16	0.5
	CREST	3,774					
Male	Total	1,186,394	12.47	3.48*	0.37	0.22	0.4
	CREST	1,877					
Female	Total	1,178,310	11.65	3.23*	0.36	0.24	0.6
	CREST	1,897					
Club	Total	2,327,032	12.21	3.94*	0.60	0.35	0.6
	CREST	683					
Lesson time	Total	2,351,021	11.45	3.24*	0.50	0.36	0.5
	CREST	1,003					
Off timetable & Discovery	Total	2,327,032	13.05	3.47*	0.75	0.34	1.1
	CREST	438					
FSM	Total	527,788	13.65	4.18*	0.89	0.65	0.6
	CREST	382					
Non-FSM	Total	1,830,996	10.60	3.22*	0.25	0.16	0.6
	CREST	3392					

Note: *statistically significant at the 5% significance level. "Unmatched" is the difference between the CREST students and non-CREST students in terms of highest GCSE science points score. "ATT" is the difference between the treatment and the statistically matched control.

Table B4 OLS regression for the impact of CREST on GCSE science attainment

Dependent variable: Highest science GCSE point score	Coefficient	Standard error
CREST	3.22*	0.13
R Squared	0.46	
Sample size	2,369,925	

Notes: * = statistically significant at the 5% significance level. Control variables are included in all models but are not shown.

Table B5 OLS regression for the impact of CREST on GCSE science attainment - including CREST-FSM and CREST-gender interaction variables

Dependent variable: Highest science GCSE point score	Coefficient	Standard error
CREST	3.08*	0.18
CREST FSM interaction	1.63*	0.52
CREST female interaction	-0.04	0.26
R Squared	0.46	
Sample size	2,369,925	

Notes: * = statistically significant at the 5% significance level. Control variables are included in all models but are not shown.

ANNEX B - Further tables

Table B6 PSM results for the effect of CREST participation on STEM AS level participation broken down by different groups

Population	Number of Observations		Difference in percentage points		S.E		Mean Bias
			Unmatched	ATT	Unmatched	ATT	
Total (main STEM definition)	Total	975,922	26	14*	0.01	0.01	0.3
	CREST	2,332					
Total (narrow STEM definition)	Total	972,922	28	13*	0.01	0.01	0.3
	CREST	2,332					
Female	Total	529,455	27	15*	0.01	0.01	1.0
	CREST	1,118					
Male	Total	446,467	24	13*	0.01	0.01	1.1
	CREST	1,214					
Club	Total	971,443	31	18*	0.03	0.02	1.0
	CREST	358					
Lesson	Total	947,661	22	10*	0.02	0.02	1.4
	CREST	650					
FSM	Total	133,357	30	21*	0.04	0.03	1.2
	CREST	193					
Non-FSM	Total	842,565	25	13*	0.01	0.01	0.3
	CREST	2,139					

Notes: *=statistically significant at the 5% significance level. "Unmatched" is the difference between the CREST students and non-CREST students in terms of highest GCSE science points score. "ATT" is the difference between the treatment and the statistically matched control.

Table B7 OLS & Logit regression for the impact of CREST participation on STEM AS level subject take-up

	OLS				Logit			
	Broad Definition		Narrow Definition		Broad Definition		Narrow Definition	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
CREST	0.08*	0.01	0.06*	0.01	0.63*	0.06	0.40*	0.06
Sample size	977,288		977,288		977,288		977,288	
R-squared	0.43		0.4		0.22		0.33	

Notes: * = statistically significant at the 5% significance level. Logit coefficients are not marginal effects. Control variables are included in all models but are not shown. Broad definition of STEM includes: biology, human biology, chemistry, physics, maths, further maths, statistics, additional maths, pure maths, electronics, environmental science, geology, public science, statistics, computer studies, graphics, textiles, 3D studies, food technology, design systems, and production design
Narrow definition of STEM includes: biology, human biology, chemistry, physics, maths, further maths.

ANNEX B - Further tables

Table B8 OLS & Logit regression for the impact of CREST participation on STEM AS level take up including CREST-FSM and CREST-gender interaction variables

	OLS		Logit	
	Coefficient	Standard error	Coefficient	Standard error
CREST	0.06*	0.01	0.72*	0.10
CREST FSM interaction	0.07*	0.03	0.41*	0.20
CREST female interaction	0.04*	0.01	-0.22	0.13
Sample size	977,288		977,288	
R-squared	0.27		0.22	

Notes: * = statistically significant at the 5% significance level. Logit coefficients are not marginal effects. Control variables are included in all models but are not shown. Broad definition of STEM includes: biology, human biology, chemistry, physics, maths, further maths, statistics, additional maths, pure maths, electronics, environmental science, geology, public science, statistics, computer studies, graphics, textiles, 3D studies, food technology, design systems, and production design

ANNEX C - A detailed look at our methodology choices

In this annex, we discuss why we chose PSM as our primary methodology for this study, the choices we made around how to match, and how we tried to avoid bias in our PSM methodology, and the limitations of PSM.

Why we chose PSM

In this study our preferred methodology is Propensity Score Matching. There are three main reasons we chose this over regression methods:

- 1) Fewer assumptions are required over the distribution of variables and error terms.
- 2) Ordinary least squares regression also relies on the assumption that the relationship between outcome variables and determinants are linear, when this may not necessarily be the case.
- 3) The potential problem of multi-collinearity in regression models (a high degree of correlation between explanatory variables which can lead to incorrect conclusions being drawn) is also removed when using PSM.

Results using regression methods are nonetheless reported in Annex B, to sense check against the primary methodology.

One further option we considered was creating a control group using exact matching methods, which involves matching treated observations (CREST students) to non-treated students with the exact same characteristics. This can lead to a better balance between treatment and control groups⁴⁰. However, due to the number of characteristics we want to balance over, exact matching would have led to very few matches and comparison groups too small to conduct robust analysis. We tested how balanced our treatment and control groups were following PSM and found them to be very closely balanced in all cases.

How we conducted our Propensity Score Matching

PSM involves selecting a comparison group of students who did not receive 'the treatment' (in this case CREST participation), but who have similar characteristics on average to the group who received

the treatment. It does this by assigning 'propensity scores' to each individual in the study, which measures the probability that the individual participated in CREST, given their set of characteristics.

More specifically, the propensity score for each person tells us the probability that they participated in CREST based on their past achievement, their ethnicity, their social background, and so on. Propensity scores are generated by defining binary choice models, with the treatment variable (CREST participation) as the dependent variable. Logit models were used to do this. Next, by taking the propensity scores of those we know participated in CREST, and matching them to a similar propensity score of somebody who we know didn't participate in CREST, the group of propensity scores which have been selected as matches (all of which will be from people who didn't participate in CREST) will be assigned to people with roughly the same characteristics on average as the group who we know did participate. Outcomes can then be compared between the two groups, with the difference being interpreted as the 'average treatment effect on the treated'. When generating propensity scores, we included as control variables all characteristics that had a statistically significant effect (at the 5% significance level) on the outcome variable in OLS regression models, regardless of whether they were significantly correlated with CREST participation in the binary choice models.

40. Blackwell et al (2009), 'Coarsened exact matching in Stata', *The Stata Journal*, Vol. 9, No. 4, pp. 524-546.

ANNEX C - A detailed look at our methodology choices

There are several different methods of matching which have been developed, each of which matches propensity scores in a different way, leading to different comparison groups in each case:

- **Kernel matching:** each treated observation is matched with several control observations. Each control observation is weighted inversely to the distance between treated and control observations.

- **Nearest neighbour matching:** for each treated observation, a control observation is selected that has the closest propensity score. Alternatively, a number of the closest observations can be selected.

- **Radius matching:** each treated observation is matched with control observations that fall within a specified radius of the treated observation's propensity score. This radius can be selected.

After the matching process has been completed, statistical tests can be conducted which compare how similar treatment and control groups are on included characteristics. These tests calculate the 'mean bias' across all control variables included as well as Rubins' B and R. The mean bias summarises the absolute biases across each variable. Rubin's B is the absolute standardised difference of the means of the linear index of the propensity score in the treated and (matched) non-treated group, whilst Rubin's R is the ratio of treated to (matched) non-treated variances of the propensity score index. Rubin recommends that for sufficiently balanced samples B is less than 25 and R is between 0.5 and 2⁴¹. Our results consistently fall well within these ranges.

Throughout the study a variety of matching methods were tried and tested. Results for the average treatment effect (on the treated) of CREST were stable across all matching methods, but the method which consistently achieved the lowest mean bias was radius matching, with a specified radius

of 0.1. Therefore, we reported all results using this method. This chosen specification followed a recommendation from the Ministry of Justice's Statistical Methods and Development team of the method for determining the optimal radius to set (0.1 multiplied by the standard deviation of the logit function of the propensity score).

PSM limitations

As previously discussed propensity scores were generated using binary choice models where all variables which were significantly related to outcome variables in regression models were controlled for, regardless of whether or not they were significantly correlated with CREST participation in the binary choice models. This approach minimises the risk of causing bias to estimates by omitting important variables by ensuring that treatment and control groups are balanced on a large number of characteristics we think are important for determining the outcome as possible. One limitation of this approach is that the variance of estimators will not be as low as they could be, which may make statistically significant variables look unimportant. However, we feel that avoiding bias to estimates is of higher priority. There is a dispute in the literature on the issue of trimming models in name of parsimony. For example, Bryson et al (2002) note that 'over-parameterised models should be avoided'⁴², whereas Rubin and Thomas (1996) argue that 'a variable should only be excluded from analysis if there is consensus that the variable is either unrelated to the outcome or not a proper covariate'⁴³.

41. Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2, 169-188.

42. Bryson, A., Dorsett, R. & Purdon, S. (2002) 'The use of Propensity Score Matching in the Evaluation of Active Labour Market Policies: A study carried out on behalf of the Department for Work and Pensions'

43. Rubin, D.B. & Thomas, N. (1996) Matching Using Estimated Propensity Scores: Relating Theory to Practice, *Biometrics*, Vol. 52, No. 1 (Mar., 1996), pp. 249-264.

ANNEX C - A detailed look at our methodology choices

PSM also has disadvantages in comparison to regression methods when assessing the impact of programmes such as CREST. One limitation is that it's not possible to test whether differences in effect between different groups are statistically significant. We cannot, for example, determine whether the impact of CREST is statistically different for males and females. This is possible using regression models by including interaction terms, which we used in this analysis to test the statistical significance of our breakdowns by gender etc.

A less important constraint of PSM is that the effect of additional variables cannot be determined. For example, when assessing the impact of CREST on science attainment we get no sense of the relationship between science attainment and factors such as prior attainment. For these reasons, and to provide a check against PSM results, regression methods are used as a secondary methodology.



Schools might want to consider whether CREST could be part of a solution to boost GCSE science results



STUDENT TESTIMONIALS

"A great teacher at my school ran a science club... so I got involved from there. I believe it really helped me to get into a good university and experience in industry. It definitely makes studying easier when you're doing something that you enjoy. The sense of achievement when the project was completed and you were presenting to others is a great feeling."

Rachel, Hertfordshire

(Age 15 when she undertook her CREST Silver project which explored different spot creams)

"CREST is a really challenging and exciting experience for anyone with an inquisitive mind. It is a rare opportunity to push beyond the boundaries of the curriculum and find the answers to your own questions. It helped me develop new skills like experiment design, teamwork, data presentation and web design.

"The best thing about the CREST Silver project was being able to develop our own idea and then share it with others. I also really enjoyed the challenge of presenting our work to experienced scientists.

"CREST definitely inspired me to work harder in STEM subjects and to aim high in further education."

Jonathan, Cheshire

(Age 16 when he undertook his CREST Silver project which investigated the sensitivity of people's taste buds)

British Science Association
Wellcome Wolfson Building
165 Queen's Gate
London SW7 5HD

www.britishtscienceassociation.org

Registered charity: 212479 and SC039236

