Impact of an active learning physics workshop on secondary school students' self-efficacy and ability

Jessie Durk,[1,*] Ally Davies,[2] Robin Hughes,[1] and Lisa Jardine-Wright[1]
[1]*Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue,*
Cambridge CB3 0HE, United Kingdom
[2]*Institute of Physics, 37 Caledonian Road, London N1 9BU, United Kingdom*

(Received 22 April 2020; accepted 9 September 2020; published 22 October 2020)

Female students and those with a low socioeconomic status (SES) typically score lower in assessments of self-efficacy and ability in science, technology, engineering, and mathematics (STEM). In this study, a cohort of over 200 UK students attended an intensive, active learning, physics workshop, with pre- and postassessments to measure both physics self-efficacy and physics ability before and after the workshop. Our control took the form of material that was closely related but not covered during the workshop. Students benefited from attending the workshop, as self-efficacy and ability increased significantly in the post-test, with the material not covered showing the smallest increase as expected. A significant socioeconomic attainment gap in ability was completely alleviated for questions on material covered at both secondary and upper secondary level, but not for questions on material seen at upper secondary only. In contrast, although no overall significant initial gender gap in ability was found, despite female students having a lower mean score than male students, a gender gap was alleviated for material seen only at upper secondary level. Female and low SES students' physics ability improved more than male and high SES students' physics ability, respectively. The workshop particularly benefited students from a mildly underperforming demographic tackling the hardest questions, or students from a significantly underperforming demographic tackling intermediate questions but not the hardest questions. The already high levels of confidence in their abilities felt by the cohort (which was boosted further by the workshop) meant that none of the demographics considered were less self-efficacious than their peers; however, the self-efficacy of female students improved more than male students, but of high SES students more than low SES students. This study provides a valuable contribution toward understanding the interaction between the extent of underperforming and question difficulty, and the features from the Bootcamp can be easily transferred to other STEM subjects.

DOI: 10.1103/PhysRevPhysEducRes.16.020126

I. INTRODUCTION

A. Ability and performance

It is known that gender and socioeconomic status (SES) largely determine a student's academic progress and achievement at school, across all subjects. Physics is one of the least diverse subjects, despite being one of the three main sciences, and sees certain demographic groups underperform or less likely to pursue it post-16, particularly female students and those with a low SES.

A student's ability in a subject is straightforward to measure, commonly in the form of assessments or examinations which assess a variety of skills such as problem solving (which we will take to be our definition of ability in this study), practical skills, content knowledge, and interpreting graphs and diagrams. In terms of gender, Table I shows a selection of the performance of male and female students for different grades in both of the main qualifications sat by students in England. The first of these is the General Certificate of Secondary Education (GCSE) qualification, taken by secondary school students aged 15–16 years old in England, Northern Ireland, and Wales. Until 2017, the top grade that could be achieved was an A*, the lowest a U (ungraded or unclassified). Numerical grades from 1 to 9 have since replaced the traditional letter grades, with 9 being approximately equivalent to an A*. The top two grades, A* and A, are now mapped completely onto the top three numerical grades, 7, 8, and 9. The second qualification is the upper secondary, post-16 Advanced Level (A Level) qualification. A typical A Level course lasts two years, with students taking their final exams when they are 17–18 years old. This grading system uses

---

[*] jd871@cam.ac.uk

TABLE I. Student performance by gender for different qualifications and grade in 2019. A grade 9 (highest grade) is equivalent to an A*, and grades 7, 8, and 9 are altogether equivalent to grades A and A* (highest two grades).

Qualification (age) and grade	% of male students	% of female students
GCSE physics (15–16 years old)		
9 [2]	14.2	10.7
7, 8 and 9 [2]	45.7	41.8
A Level physics (17–18 years old)		
A* [1]	8.8	8.5

traditional letters, with the highest grade being an A*. A Levels are taken by students in England, Wales, and Northern Ireland, with alternatives offered in Scotland. As can be seen in Table I, the gender gap in attainment in physics does continue at A Level, although considerably smaller. This can be explained by the fact that so few female students opt to continue studying physics post-16, (female students made up 23% of all A Level physics entrants in 2019 [1]), meaning those who do so are likely to have done well in physics, thus reducing any initial attainment gap.

In terms of SES, however, the attainment gap is much wider, as can be seen in Table II. This gap in performance is widely documented and has been the subject of many reports, particularly in science [3,4]. In the academic year 2005–2006, just 0.1% of all GCSE entrants achieving the top A* grade did so in physics and were eligible for free school meals (FSM), compared with 1.1% for non-FSM, while for A Level these figures were 0.4% and 1.2%, respectively [5]. Free school meal eligibility is commonly used as an indicator of low SES. A study of 6000 students in comprehensive secondary schools found that between year 6 (the final year of primary school, for pupils aged 10–11 years old) and GCSE level students eligible for free school meals were behind their non-FSM peers by almost one-third of a grade in science, with similar findings for English and math [6].

A student's SES is indicated by a number of variables, such as family background, free school meal eligibility, and various geographical indexes [7]. Measures referring to family or parental background include parental occupation, level of education, and income [5]. There are many reasons as to why low SES students perform worse than their peers, especially in science and physics. Some of these include a lack of specialist science teachers, poor career advice, low parental engagement, low aspirations toward pursuing a science, technology, engineering, and mathematics (STEM) career, an unawareness of what jobs are available by studying science or physics, and low science capital [3,7,8].

While much has been done to map out the current SES and attainment landscape, very few studies have looked at interventions to raise the science attainment of low SES students (for a comprehensive list, see Ref. [4]). Of these studies, the majority report a general effect on all participants, rather than analyzing any differences between FSM and non-FSM students. A large proportion of the studies are also based in the U.S. An intervention designed to develop science writing skills in elementary age pupils in the U.S. saw the attainment gap between FSM and non-FSM pupils decrease [9]. UK-based studies have often focused on analyzing primary school students, or sociocultural interventions such as informal science settings. Examples of such studies include a group work focused intervention for primary science lessons, which saw a small decrease in attainment when the percentage of FSM pupils increased [10], as well as a professional development intervention for primary science teachers, in which students were given a pre- and post-test to measure their science attainment before and after the intervention (alongside a control group) [11]. The impact of this intervention was reported in terms of the effect size. This quantifies the magnitude of the difference between the two sets of results, in this case between the FSM pupils' pre- and post-test scores, as well as between the non-FSM pupils' pre- and post-test scores. The effect sizes in the primary science professional development intervention for FSM and non-FSM were found to be 0.38 versus 0.22, respectively, showing that FSM pupils made greater progress than their peers. Most studies report positive effects for the low SES participants—several meta-analyses show mean effect sizes ranging from 0.25 [12] to 0.88 [13]; however, a two year

TABLE II. Student performance by free school meal (FSM) status in 2015. GCSE science includes physics, chemistry, and biology.

Qualification (age) and grade	% of non-FSM students	% of FSM students
GCSE science (15–16 years old)		
A* to C [4]	70	42

intervention in the U.S. for three different age groups saw the SES attainment gap actually increase in the post-test [14].

In terms of physics, there are even fewer studies. A specialist school for low SES pupils in the U.S. received an intervention focusing on improving student's self-regulation and metacognitive strategies. Low SES students traditionally lack the ability to take control of their own learning and evaluate their progress, often resulting in poor academic performance [15]. As a result of the intervention, their attainment in a physics test increased with an effect size of 0.42. Metacognitive instructional approaches were again shown to benefit low achieving students, more than their higher achieving peers in Ref. [16]. This is also a positive result as we know that low SES students are more likely to score low on tests, so any interventions designed to increase the attributes they typically lack will boost their attainment.

B. Self-efficacy, performance, and participation

Self-efficacy is an individual's belief that they can complete a given task and is related to self-perception, confidence, and motivation [17]. Students with high science capital are more likely to have high self-efficacy and therefore more likely to pursue science post-16, and be male or from a high SES background [8]. Typically, female students have lower STEM self-efficacy than male students [18–20]; however, this difference disappears when science anxiety is controlled for [21], with some studies showing the opposite. A study in the U.S. showed female students have higher science self-efficacy than male students [22]. The authors point out that this may be because the children in the study attended middle schools where science is taught in a more integrated, language based way, which would appeal to female students. Positive correlations between SES and self-efficacy have been found in math [23] and physics [24].

Self-efficacy is a strong predictor of academic performance and ability in science, with students reporting higher self-efficacy achieving better results [18,22,25]. Students with high self-efficacy have the confidence to tackle more challenging material, and so progress more than their low self-efficacy peers [26]. Low self-efficacy and underperforming in physics (and science) at secondary level has implications for students' post-16 participation in these subjects; see Ref. [27] for a comprehensive review of other factors. Low SES students are underrepresented in the sciences at A Level, with physics having one of the lowest representations [5,28]. Furthermore, despite there being only a small attainment gap in gender at secondary level, see Table I, a significantly low number of female students continue physics post-16—this result has remained unchanged for the past three decades [29].

Studying physics is important in our increasingly scientific and technological society [30]. There is a growing demand for STEM skills, particularly physics, as the UK government's industrial strategy identifies key areas for growth which require a STEM-skilled workforce [29]. It is important that such a workforce, currently experiencing a shortfall of 400 000 STEM graduates in the UK each year, is diverse and opportunities are open to all.

Many studies regarding self-efficacy involve sociocultural interventions, and so analyze a variety of domains, including beliefs, attitudes, and perceptions. 33 000 pupils in the UK were provided with residential science fieldwork, and data from 2706 of these students over five years showed improved self-perception as well as cognitive, interpersonal, and behavioral gains [31]. The majority of the pupils were low SES. Similarly, a two year STEM ambassador program in the UK for secondary age pupils saw them develop a science identity and relate to being a student at university [32]. Social and emotional learning practices, such as recognizing that the social curriculum is as important as the academic curriculum, and that how children learn is as important as what they learn, have also been implemented to improve self-efficacy [33]. A U.S. study analyzed ethnically and socioeconomically diverse 10-year-old pupils, and found that in classrooms using more social and emotional learning approaches, the students had higher science self-efficacy [21].

C. Active learning intervention studies

A number of studies employ interventions that involve active learning techniques. Active learning is a method of learning which involves the learner directly instead of passively as in the case of a traditional lecture. General characteristics associated with active learning include students doing more than simply listening, a greater emphasis being placed on developing student's skills rather than relaying information, higher order thinking takes place, as well as reading, discussing, and writing, and students may explore their own ideas and attitudes [34]. Active learning has been shown to be effective in increasing students' performance [35] as well as attendance and engagement [36]. Active teaching and team-based learning strategies can reduce a self-efficacy gender gap in physics [37,38], although others report an increased gap [39], with Ref. [40] reporting a general decrease in female students' self-efficacy after enrolling in physics courses. One study found students in classes using active engagement methods had better attitudes and approaches toward problem solving in physics than those in traditional lectures [41]. However, a more complex relationship between performance and self-efficacy is suggested in Ref. [42]. The authors studied introductory university-level physics courses and found students undertaking active learning learn more but actually perceive themselves to have learned less than their traditional learning peers. This suggests the increased cognitive effort in the active learning environments may initially have a detrimental effect on motivation and engagement; thus, care must be taken to remedy this in the long term, so that students can fully benefit from active learning.

II. METHODOLOGY

Within the current literature, studies that focus on physics do so by investigating undergraduates, particularly those based in the U.S., while studies that analyze secondary school students typically do so by looking at science as a whole, as we have seen in Sec. I. As a result, there are very few UK-based studies that have either focused on physics, or both ability and self-efficacy, or secondary school students, thus highlighting a significant gap in the literature surrounding intervention studies and active learning. In this study we have investigated both physics ability and self-efficacy, with an active learning based instruction in the form of an intensive physics workshop weekend. The effect of this on the self-efficacy and ability of approximately 150 students was analyzed through pre- and postassessments, as well as how certain demographics performed before and after the instruction. The physics workshop investigated in this study provided an ideal setting to measure ability and self-efficacy of UK secondary school physics students and underrepresented demographics, and provides a much needed contribution to the literature surrounding active learning intervention studies in physics.

A. Isaac Physics

The workshop was run as an integral part of Isaac Physics, a project based at the University of Cambridge and funded by the Department for Education and The Ogden Trust [43]. It is a free educational resource for secondary school and university students in the UK, consisting of an open platform for active learning, face-to-face events, online mentoring, and printed materials. The platform itself contains a wide range of physics (as well as math and chemistry) questions from GCSE to first year university level, that students can attempt and then receive immediate feedback, while monitoring their own progress. The ethos of the project is that by attempting and solving physics problems, students develop a deeper understanding of concepts and become more confident physicists. Since its inception in 2013, Isaac Physics has had over 265 000 registered students and 9000 registered teachers, and is currently used in over 3600 schools. Schools where more than 50% of the cohort use Isaac Physics see 40% of students achieve one grade higher than before, and students are statistically more likely to apply to, get an offer from, and attend higher tariff universities [44].

The focus of this study was on the annual A Level workshop event, the Isaac Physics Bootcamp, which takes place at the University of Cambridge, just before year 13, the final year of A Levels. It involves a single, intense weekend attending revision-style lectures and answering Isaac Physics questions, and members of staff provide guidance and support. The workshop incorporates aspects of active learning in the form of problem solving, group work, and discussion between students and their peers and group leaders.

The workshop ran from the evening of Friday, August 30, 2019 (with the first session at 7 p.m.) until midday on Sunday, September 1. The timetable for Saturday ran from 9 a.m. to 8:30 p.m., with meals and breaks scheduled throughout the day. Each session was devoted to a different topic, lasting typically 90 minutes each. During a session, students were given a 10–15 minute recap of a concept by an experienced physics teacher, and the remaining time to discuss and apply their conceptual understanding of the topic to a selection of multistep synoptic questions. The students therefore played an active role in their learning. They worked in groups of 6, with group leaders each responsible for helping 4 groups, and these group leaders consisted of members of the Isaac Physics team, experienced teachers, and undergraduates. The group leaders directed discussion and answered questions that arose both within smaller groups of 2 or 3 students within a group of 6 and also the entire group more generally. The students were asked exploratory questions such as, “What other way could you solve that question?” and “What happens if this aspect is altered?” The overwhelming majority of students engaged with their group discussion. Finally, no more than 3 students from the same school attended the Bootcamp, and the groups of 6 were mixed in terms of school background and gender.

Attendance at the Bootcamp is permitted only if the student meets one or more widening participation criteria and attends a state school or college in England. The criteria include being the first in their family to attend higher education, being eligible for free school meals or means-tested government bursaries during secondary school, attending a school or college with a below average A Level point score, and attending a school or college or having a home address in an area defined as having a low progression of students to higher education. The latter criterion is known as the participation of local areas (POLAR) quintile of a school or address in the UK, and the most recent version is known as POLAR4. Each school or college and home address is assigned a ranking from 1 to 5, with 1 representing an area with the lowest progression to higher education and 5 the highest. If a student attends a school in or lives in an area in a POLAR4 1, 2, or 3 quintile they are eligible to attend the Bootcamp.

Figure 1 shows the proportion of students that attended the Bootcamp in each POLAR4 quintile. 44% of students attended a school in POLAR4 1, 2, or 3 and 59% have a home address in POLAR4 1, 2, or 3. The values from Fig. 1 do not add up to these percentages as the figure values are given to the nearest integer for brevity. The reason that the percentages for school and home addresses differ is that there are many students who attend a school with a different POLAR4 ranking than their home address, as a result of them attending a school far away from where they live.

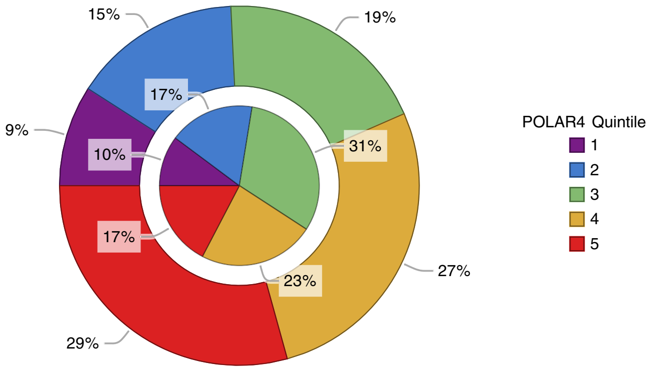

FIG. 1. Percentage of students that attended the Bootcamp in each POLAR4 quintile. The outer ring shows the percentages for the school address POLAR4 quintiles, and the inner ring shows the percentages for the home address POLAR4 quintiles. The total numbers are $N = 181$ for the schools data and $N = 172$ for the home data.

From Fig. 1 it is more likely that a student attends a school in a higher POLAR4 quintile than their home address.

Of the approximately 200 students who attended the workshop, 171 gave demographic data permission and reported their FSM status. Of the 171, 38 were eligible for FSM (22%), while the remainder, 133, were not eligible. Eligibility for free school meals, either at an individual level or at school level, in other words the percentage of children in a school that are eligible, is widely used as a measure of socioeconomic deprivation. A potential drawback is that it divides students into the very poorest and “everyone else” [5]. In the UK in January 2020, 16% of all secondary school pupils were eligible for and receiving FSM [45]. We decided to use individual eligibility for free school meals as our socioeconomic status indicator over the other criteria, such as school POLAR4 ranking, as FSM eligibility is a simple binary indicator and has been shown to be a good indicator for socioeconomic disadvantage [46,47]. We expect it to be a strong measure of any socioeconomic differences, as it reflects an individual’s own demographics and does not assume the school or home is representative of the student.

B. Research questions

We identify our research questions as follows:

[RQ1:] What is the impact of the workshop on the physics ability and self-efficacy of the students?

[RQ2:] What is the impact of the workshop on the physics ability and self-efficacy of the students, in terms of demographics such as gender and SES?

C. Measuring ability and self-efficacy

Each student was randomly assigned to complete only one of the two assessments—either a survey to measure physics self-efficacy or a physics test to measure physics ability. Each student then completed both a pre- and post-version of their assigned assessment. A study that measured students’ self-efficacy before and after a test showed that their initial self-efficacy ratings decreased after the test, particularly for lower ability students [48]. This suggests self-efficacy ratings are affected by taking part in a test, and therefore we decided to measure these two attributes separately. We further cannot be sure that measuring self-efficacy beforehand will not also affect test performance. As the students had been randomly allocated to these assessments, we are reasonably confident that there are no between group differences and that any effect on self-efficacy can be linked to the effect on physics ability. This meant, however, our sample sizes would by definition be smaller for each of the two assessments, and so poses a potential limitation for there being no differences between the randomly allocated groups. This is discussed further in Sec. III.

Upon arrival at the Bootcamp, each student was given their preassessment to complete, and the postassessment was given at the end of the Bootcamp weekend, after the physics sessions and activities. The self-efficacy survey consisted of 24 statements (see the Appendix), and students were asked to indicate their confidence for each statement on a 10-point Likert scale. We chose this over a 5-point Likert scale because participants have a tendency to avoid extreme positions, and thus having a smaller number of responses to choose from limits their overall response range, reducing overall survey reliability [49]. The post-self-efficacy survey had the same 24 statements but in a different order, to control for order effects in which the statements are met. This was done to eliminate any statement order bias, as the students could have felt more or less self-efficacious as they progressed through the survey. As we randomized the order of the subsequent postsurvey statements, we do not expect any change in self-efficacy to be a result of the new order of the statements—we expect this effect to be negligible, with the main effect arising from the Bootcamp itself.

The physics test consisted of 11 multiple-choice questions adapted from various sources such as Isaac Physics questions and the University of Cambridge Natural Science Admission Assessment questions. We show both versions of an example question, question 2, in the Appendix, as well as the preamble text written at the beginning. These questions predominantly assess problem-solving skills, which we take as our definition of ability. Each question was designed to be answered in roughly 90 seconds. The post-test questions were in the same order as the pretest, and the questions were paired such that question 1 in the post-test was on exactly the same topic and of the same difficulty level to question 1 in the pretest, and so on.

As we considered only students attending the Bootcamp, our experiment takes the form of a quasicontrolled experiment. Our control consisted of some statements or questions about material that was not covered in the

TABLE III. Topics, difficulty level, and coverage (whether that material was covered or not) for each of the 11 questions.

Question	Topic	Difficulty level	Coverage
1	Mechanics	GCSE and A Level	Covered
2	Mechanics	A Level	Covered
3	Circuits	GCSE and A Level	Not covered
4	General skills	GCSE and A Level	Covered
5	Waves	A Level	Covered
6	Mechanics	A Level	Covered
7	Circuits	GCSE and A Level	Covered
8	Waves	A Level	Covered
9	General skills	GCSE and A Level	Not covered
10	Circuits	GCSE and A Level	Covered
11	Waves	GCSE and A Level	Covered

Bootcamp. We do not expect the scores from these to increase significantly. Any increase in the noncontrol statements or questions can be attributed to the effect of the Bootcamp. For the self-efficacy survey, our control consisted of 8 statements about math topics, as math was not covered during the Bootcamp. We decided to use math statements rather than chemistry or biology, as the majority of students taking A Level physics also take A Level math, and therefore any math material would be familiar to the students. For the physics questions, material that was covered consisted of being shown the equation in the revision-style lecture, then answering a question directly involving that equation, for example, or being shown a topic but not a specific equation, and then needing to recall or use the equation. The questions that were covered were then divided into two difficulty levels—those whose material is introduced at upper secondary (A Level) only, and those whose material the students met at secondary (GCSE) and cover again at upper secondary (A Level). Table III lists the 11 questions for the physics test, their topic, difficulty level, and whether the material of that particular question was covered in the Bootcamp or not (labeled "coverage"). We do not expect the scores for both questions 3 and 9 to increase significantly in the post-test, as these questions were on material not covered during the Bootcamp. The physics test was implemented in a flipped format, where half of the cohort's pretest was the other half's post-test, and vice versa. This was to remove any bias from the two papers being nonidentical.

1. Validity

In order to ensure our self-efficacy survey had internal reliability, we used the Cronbach's alpha measure of consistency. This measures how closely related a set of results are, with higher values indicating more consistent results. We calculated Cronbach's alpha for the entire self-efficacy presurvey, as the postsurvey is the same but ordered differently, as well as for each of the topics (math, therefore not covered, and physics). It was found that removing one statement (statement 22; see the Appendix) produced a higher Cronbach's alpha value, and did not correlate with the remaining statements. We therefore removed this statement from the reliability test and any subsequent analysis. Table IV shows the Cronbach's alpha values for the presurvey. All values are above the accepted level of 0.7 [50]. The value of alpha for the entire survey is higher than the values obtained for the physics and math statements—this is a result of the entire survey naturally containing more items than its component surveys. Although Ref. [50] illustrates that calculating the Cronbach's alpha for an entire survey can be problematic, see Table 1 within this reference, we decide to include our entire-survey value for completeness, and for the fact that all values are above 0.7 and do not differ greatly.

TABLE IV. Cronbach's alpha for all of the 23 statements combined and for the 2 different topics during the Bootcamp.

Topic or coverage	Number of statements	α
All	23	0.88
Math	8 (statements: 1, 7, 9, 10, 14, 15, 16, and 21)	0.73
Physics	Remaining 15 statements	0.85

III. RESULTS

A. Physics ability test

There were 81 students who completed both a physics ability pre- and post-test. The results for these are shown in Fig. 2, which shows box plots for the number of students getting each question correct, for both the pre- and post-test. The datasets that form each of the box plots consist of

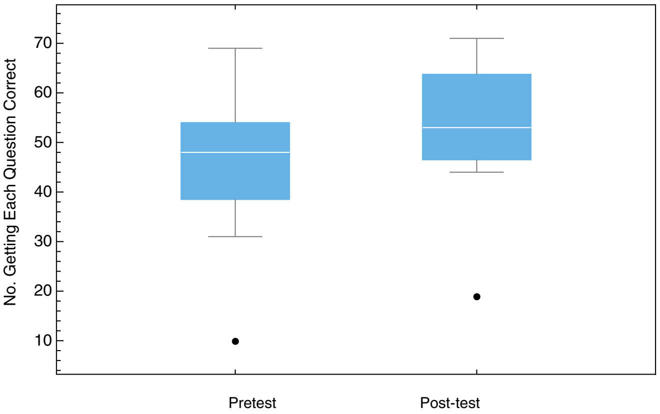

FIG. 2. Box plots for the 11 questions and the number of students getting them correct, for the pretest on the left and the post-test on the right. The bottom and top of the blue boxes represent the 25th and 75th percentiles, respectively, the white line in the center represents the median value (50th percentile), and the whiskers represent the highest and lowest values, excluding any outliers. The outlier (black dot) in both represents question 11.

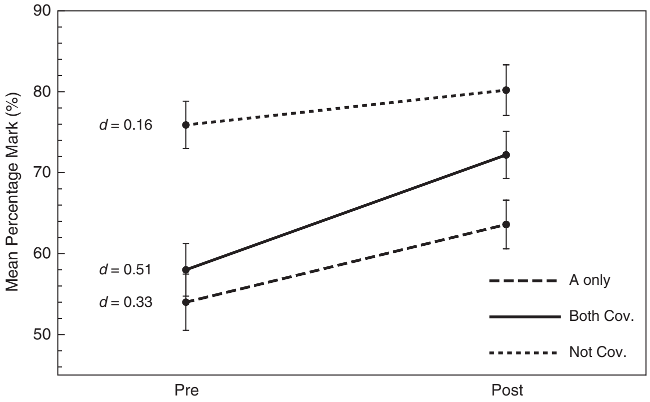

FIG. 3. Mean physics pre- and post-test scores for all 81 students, for each level. A Only (dashed line) refers to the questions that were A Level difficulty (questions 2, 5, 6, and 8). Both Cov. (solid line) refers to the questions that were both GCSE and A Level difficulty, and were covered (questions 1, 4, 7, and 10), while Not Cov. (dotted line) refers to those that were not covered at all (questions 3 and 9). The error bars represent the standard error of the mean.

11 data points, each data point corresponds to a question on the test, and its value is the number of students that answered it correctly. It is clear that in both tests there is one question that was answered correctly by significantly fewer students than the other questions, shown as an outlier. This corresponds to question 11, which we believe students did not have sufficient time to answer properly as it was at the end of the test. As a result, and in order to reduce any bias, we chose to remove this question from all subsequent analyses.

The results of the physics ability test for each of the difficulty levels are shown in the first row of Table V and graphically in Fig. 3 along with Cohen's d values for each of the levels. Cohen's d is a measure of effect size (how different two sets of data are) and is used when the two datasets are of equal size. It is calculated from the means of the two sets of data as well as the pooled standard deviation of the whole sample. Cohen's d values of 0.2, 0.5, and 0.8 are interpreted as small, medium, and large effects, respectively [51]. Our error bars for Fig. 3 and subsequent Figs. 4, 5, 6, and 7 represent the standard error of the mean, in this case the standard deviation of the sample divided by the square root of the sample size.

The control questions (questions 3 and 9) saw the smallest effect, and we can use a paired samples t-test to determine whether the small increase in scores was statistically significant or not. This type of statistical test assesses whether the mean scores from two sets of data are statistically different from each other. Our t-test for the control questions yields $[t(80) = 1.12, p = 0.265]$, where the value in brackets shows the degrees of freedom (d.o.f.), which represents the size of the sample; the test statistic $t(\text{d.o.f.})$ is related to the p value, and a larger value of t indicates a smaller probability that the results occurred by chance; and the p value indicates the actual probability that the results occurred by chance. In this case, as the p value is above 0.05, we conclude that the pre- and post-test scores for the control questions are not statistically different.

Another useful statistical test in determining the significance of the results is the analysis of variance (ANOVA) test, which can be thought of as a generalization of the t-test. ANOVA tests look for statistically significant differences between the means of groups of data, often with more than one independent variable (e.g., gender, ethnicity, age group), and/or more than one condition within these independent variables (e.g., binned age groups of 0–18, 18–24 years, etc.). "Independent" ANOVA tests look for between group differences, such as between male and female students, while "repeated measures" ANOVA tests will look for differences within groups, such as pre- and post-test score. For repeated measures tests, participants take part in all the conditions, unlike an independent ANOVA where each participant can only take one condition of the independent variable. Naturally our analysis involves both of these comparisons (investigating demographic differences, and investigating pre- and post-test or math

TABLE V. Physics pre- and post-test scores for each of the demographics and by level. Marks are shown as a percentage for that level group. We display the mean, the standard deviation of the sample in brackets, and the Cohen's d between pre- and post-test scores. N indicates the number of students in each sample.

		A level only (covered)			GCSE and A level (covered)			GCSE and A level (not covered)		
Demographic	N	Pre	Post	d	Pre	Post	d	Pre	Post	d
All	81	54.0 (31.2)	63.6 (27.1)	0.33	58.0 (29.3)	72.2 (26.2)	0.51	75.9 (26.4)	80.2 (28.2)	0.16
Gender										
F	24	44.8 (30.4)	61.5 (26.6)	0.58	57.3 (23.9)	74.0 (21.5)	0.73	79.2 (25.2)	77.1 (25.5)	0.08
M	41	59.2 (32.5)	67.7 (26.3)	0.29	60.4 (29.8)	75.0 (28.1)	0.51	74.4 (27.6)	81.7 (31.1)	0.25
FSM										
Y	13	30.8 (25.3)	53.9 (20.0)	1.01	42.3 (27.7)	75.0 (28.9)	1.15	57.7 (27.7)	65.4 (37.6)	0.23
N	48	61.5 (30.9)	69.8 (26.8)	0.29	65.6 (25.6)	75.5 (23.9)	0.40	82.3 (24.2)	85.4 (25.2)	0.13

and physics differences, etc), so we perform two- or three-way "mixed" tests to take this into account. Two- or three-way refers to the number of independent variables, mixed refers to the fact we will compare *between* our independent variables, as well as *within* them, as all students sit a pre- and postassessment.

For the physics ability test for all 81 students, we perform a two-way mixed ANOVA, with test and level as the two independent variables, and score as the dependent variable. The independent variables then take two (pre- and post-test) and three (A Level only; GCSE and A Level covered; GCSE and A Level not covered) conditions, respectively. Our ANOVA for pre- and post-test scores yields $[F(1,80) = 16.67,\ p \sim 0]$, where the values in brackets represent the independent variable and residual error degrees of freedom, respectively; the test statistic, or F ratio F is related to the variance, and a larger F means that it is more likely the effect being investigated is significant, and p represents the p value. In this case, there were very statistically significant differences between pre- and post-test scores, as expected, and can also be seen at a naive glance from Fig. 2. There was also a statistically significant difference between the three levels, $[F(2,160) = 37.88,\ p \sim 0]$. This shows how the three levels do not have equal mean values, which can be seen in Fig. 3, as the questions that were not covered see higher marks than the other two levels. The A Level only questions are performed least well. We deduce that this is because these would have been harder questions, on material introduced at A Level only. Our ANOVA test then analyzed the "interaction" effect between the two variables of pre- and post-test and level, and found this to be nonsignificant, $[F(2,160) = 2.16,\ p = 0.119]$. Again this can be seen in Fig. 3; although the three levels do experience different effects from the Bootcamp (shown also by the differing Cohen's d values), with the GCSE and A Level covered question scores increasing the most, the three increases in marks are not sufficiently different enough to be significant. We expect this result to change when looking at students' demographics.

1. Student demographics: Gender

Of the 81 students who completed a physics test, 24 female and 41 male students consented for their scores to be matched to their demographic data. This gives a sample with 37% female students, which is well above the national percentage of physics A Level students who are female (23%). The female students had a lower overall mean pretest score than the male students [male 62.7%(24.2) versus female 56.7%(20.1)], where the quantities in brackets denote the standard deviations of the sample. However, an independent t-test found there to be no statistically significant gender gap before instruction $[t(55.8) = 1.07, p = 0.144]$. These findings align with existing results described in Sec. I [1,2], as there is a small but not significant gender gap, with male students achieving slightly better results, and suggests that the female students attending the workshop are as capable as the male students, as far as significant differences are concerned.

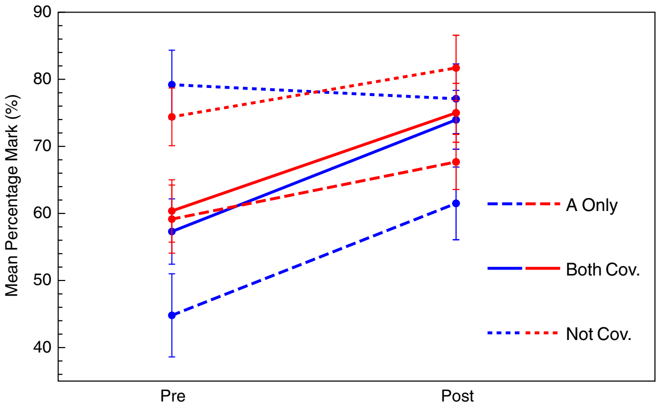

FIG. 4. Mean physics pre- and post-test scores for female (blue lines) and male (red lines) students, for each level. A Only (dashed lines) refers to the questions that were A Level difficulty (questions 2, 5, 6, and 8). Both Cov. (solid lines) refers to the questions that were both GCSE and A Level difficulty, and were covered (questions 1, 4, 7, and 10), while Not Cov. (dotted lines) refers to those that were not covered at all (questions 3 and 9). Errors bars denote the standard error of the mean.

The main results for the physics test in terms of gender are shown in the second two rows of Table V, where we display the mean and standard deviation of the sample for the three levels for male and female students, across the pre- and post-tests, as well as the number in each sample. Figure 4 displays these results graphically.

Both genders see the smallest change for the questions that were not covered as expected. We find a significant gender gap between the male and female students scores from the A Level only questions, while no such gap appears for questions from the GCSE and A Level questions. Figure 4 shows how this initial A Level gender gap is alleviated at the end of the Bootcamp. This suggests that the slight initial but nonsignificant gender gap in results seen in Table V is exaggerated when considering harder topics—those introduced only at A Level. The female students experience a larger effect size than the male students for both types of questions that were covered (A Level, and GCSE and A Level), in particular for the A Level only questions (0.58 for female students versus 0.29 for male students), which allows this initial gap to be reduced.

We perform a three-way, mixed ANOVA with gender, level, and pre- and post-test as the independent variables, and score as the dependent variable. Gender is a between-participants variable, while level and pre- and post-test are within-participants variables. Levene's test confirmed that the assumption of homogeneity of variance had been met for the pretest scores for the A Level only questions; both GCSE and A Level questions; and those not covered $[F(1,63) = 0.58,\ p > 0.05;\quad F(1,63) = 2.26,\ p > 0.05;$

$F(1,63) = 1.27$, $p > 0.05$], respectively, and the same for the post-test scores [$F(1,63) = 0.01$, $p > 0.05$; $F(1,63) = 2.26$, $p > 0.05$; $F(1,63) = 0.09$, $p > 0.05$], respectively.

The ANOVA revealed statistically significant ($p \sim 0$) main effects of pre- and post-test on score, as expected and seen previously, and of level on test score, again as seen previously (the A Level only questions see the lowest marks, not covered questions the highest). There was no significant main effect for the between-participants variable of gender (averaging both pre- and post-test scores) [$F(1,63) = 0.66, p = 0.421$], indicating that despite the female students having slightly lower marks overall than the male students, these scores are not significantly lower. Furthermore, there were no significant interaction effects between gender and pre- and post-test score; gender and level; and gender, level, and pre- and post-test score. The first of these suggests that despite the female students experiencing a larger effect size than the male students (shown by the Cohen's d values) for all three levels, the ANOVA test tells us these are not significantly larger. Finally, we find a statistically significant interaction between level of question and pre- and post-test score [$F(2,126) = 3.24, p = 0.042$], where the increases in score for the three levels are significantly different. To summarize, although the female students have slightly lower marks than the male students, overall this gap is not large enough to be significant. However, when considering the hardest questions only, those at A Level, a significant gap is revealed, which the Bootcamp is able to alleviate. We expect these results for gender to be a watered-down version of those for FSM.

2. Student demographics: SES

Of the 81 students who completed a physics test, 61 students (13 FSM and 48 non-FSM) gave FSM data and consented for their scores to be matched to their demographic data. For all pretest questions overall, there is a statistically significant initial gap in attainment between the FSM and non-FSM students, with the FSM eligible students performing much worse [FSM 40.8%(18.0) versus non-FSM 67.3%(20.5)]: [$t(21.23) = 4.57, p \sim 0$].

Their results for the physics test for each of the three levels are shown in the final two rows of Table V, where we display the mean and standard deviation of the sample across the pre- and post-tests for students eligible and not eligible for FSM, as well as the number in each sample. Figure 5 displays these results graphically.

The FSM eligible students experience the greater effect from the workshop for all levels, shown by the Cohen's d values, compared with the non-FSM students. It is clear that there is an attainment gap for all three levels of question type, with the largest gap occurring for the harder, A Level only questions. For the post-test, the attainment gap for the GCSE and A Level covered questions is reduced completely, and marginally for the A Level only questions. This is due to the FSM students experiencing large effect sizes for these (1.15 and 1.01, respectively) compared with their non-FSM peers. However, despite a large effect size for the A Level questions, it is not enough to close the gap completely. The attainment gap still persists with the questions that were not covered, and these questions see the smallest increase in marks for both types of FSM eligibility as expected. We attribute the effect for the covered material to the Bootcamp.

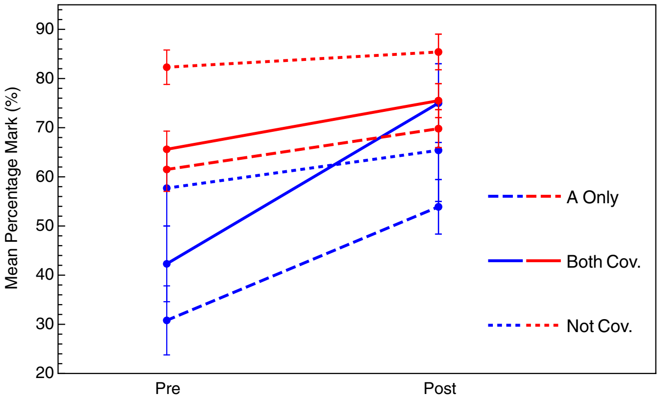

FIG. 5. Mean physics pre- and post-test scores for FSM (blue lines) and non-FSM (red lines) students, for each level. A Only (dashed lines) refers to the questions that were A Level difficulty (questions 2, 5, 6, and 8). Both Cov. (solid lines) refers to the questions that were both GCSE and A Level difficulty, and were covered (questions 1, 4, 7, and 10), while Not Cov. (dotted lines) refers to those that were not covered at all (questions 3 and 9). Errors bars denote the standard error of the mean.

As there was a significant difference between the FSM and non-FSM students' pretest scores, shown by the previous t-test, we perform an ANCOVA with FSM eligibility as the independent variable, post-test score as the dependent variable, and pretest score as the covariate. ANCOVA stands for analysis of covariance, and is an extension of an ANOVA test as it looks for statistically significant differences between adjusted means. This type of test is used when controlling for a variable, in this case, the pretest score, known as the covariate. We perform separate ANCOVA tests for each of the difficulty and coverage levels. For the A Level only questions, Levene's test for homogeneity of variance was not significant [$F(1,59) = 3.45, p > 0.05$], and the ANCOVA revealed no significant difference in FSM and non-FSM students' post-test scores, when controlling for pretest scores [$F(1,58) = 0.50, p = 0.483$]. Similar results were found for the GCSE and A Level covered questions—Levene's assumption was not violated [$F(1,59) = 0.60, p > 0.05$], and the ANCOVA also showed no significant difference in FSM and non-FSM students' post-test scores, when controlling for pretest scores [$F(1,58) = 0.95, p = 0.334$]. For the control questions, those that were not covered, the assumption of homogeneity of variance was not met

TABLE VI. Self-efficacy pre- and postsurvey ratings for each of the demographics and by topic. The marks shown are the mean self-efficacy rating per statement for that topic, and are out of 11. We display the mean, standard deviation of the sample in brackets, and the Cohen's d between pre- and post-test scores. N indicates the number of students in each sample.

		Physics			Math		
Demographic	N	Pre	Post	d	Pre	Post	d
All	79	7.89 (1.02)	8.93 (0.71)	1.19	8.21 (1.18)	8.59 (1.10)	0.33
Gender							
F	29	7.79 (0.98)	8.93 (0.70)	1.34	8.15 (0.95)	8.59 (0.85)	0.49
M	41	7.86 (1.09)	8.93 (0.68)	1.19	8.15 (1.29)	8.44 (1.24)	0.23
FSM							
Y	11	7.73 (1.05)	8.80 (0.81)	1.14	8.20 (1.15)	8.40 (1.49)	0.15
N	53	7.84 (1.07)	8.98 (0.67)	1.28	8.23 (1.15)	8.58 (0.99)	0.33

$[F(1, 59) = 6.31, p = 0.015]$; therefore our finding that there is still a significant difference at the 10% level between the FSM and non-FSM post-test scores in material they did not cover is to be met with caution, $[F(1, 58) = 2.83, p = 0.098]$. This is to be expected given a relatively small sample size for the FSM students, and the fact that the not covered category of questions contains only two questions. An ANCOVA test for the post-test as a whole (incorporating all three difficulty and coverage levels) revealed that there is indeed no significant difference between the two groups' post-test scores, with the pretest as a covariate $[F(1, 58) = 0.83, p = 0.367]$ (Levene's test was not significant $[F(1, 59) = 0.15, p > 0.05]$). As the t-test found an initial significant difference, we conclude that students eligible for FSM show a greater improvement in physics ability than their non-FSM peers, and that the initial attainment gap is reduced significantly.

We further conclude that as there is no overarching gender gap but there is one in terms of socioeconomic status, which is prevalent for all difficulty levels, that SES is a more powerful contributing factor to attainment.

B. Self-efficacy

There were 79 students who completed both a self-efficacy pre- and postsurvey. For each student, the ratings they gave for each statement were totaled, then divided by the number of statements to obtain an average rating for that student, for each topic. Their results are shown in the first row of Table VI and graphically in Fig. 6, as well as the Cohen's d for the physics and math statements. Although self-efficacy was initially quite high, this increased after attending the Bootcamp. It is clear that despite receiving no instruction in math, the students reported an increase in their math self-efficacy. However, this was still a much smaller change than for the physics statements. Both of these increases were significant [math $t(78) = 5.596, p \sim 0$; physics $t(78) = 13.147, p \sim 0$]. Students can feel like they have learned, and therefore report an increase in self-efficacy, despite having received little instruction [42]. We can therefore attribute the larger effects seen in the physics self-efficacy to the students receiving direct instruction in this topic at the Bootcamp.

We performed a two-way mixed ANOVA, where our independent variables were pre- and postsurvey and topic (math or physics), and average rating as the dependent variable. There were statistically significant differences between pre- and postsurvey ratings $[F(1, 78) = 130.97, p \sim 0]$ but not between the two topics themselves $[F(1, 78) = 0.001, p = 0.974]$. This is clear from Fig. 6, where the mean overall physics rating—averaging the pre- and postsurvey value—is comparable to the equivalent rating but for math. However, we found a significant interaction effect between the topic and pre- and postsurvey rating $[F(1, 78) = 68.04, p \sim 0]$, indicating that the physics statements saw a significantly larger increase in self-efficacy than the math statements. Both topic's increases were significant, but even more so for the physics statements.

It is interesting to investigate whether students who presented a lower presurvey self-efficacy rating see a greater increase than those students who already ranked

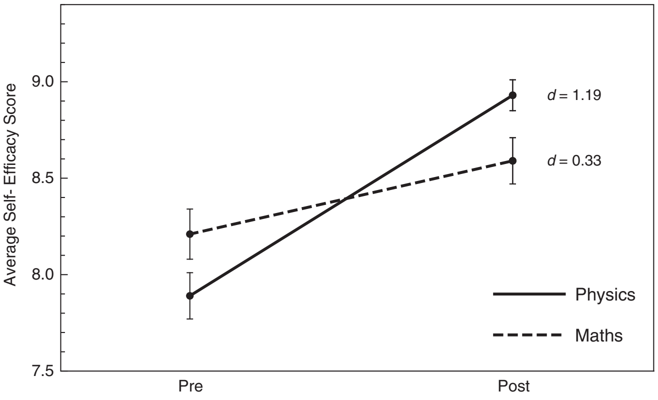

FIG. 6. Mean self-efficacy pre- and postsurvey ratings for all 79 students, grouped by statement topic (physics, shown by the solid line, and those not covered, in this case math, shown by the dashed line). Marks shown represent the average rating per statement, on a 10-point Likert scale. The error bars represent the standard error of the mean.

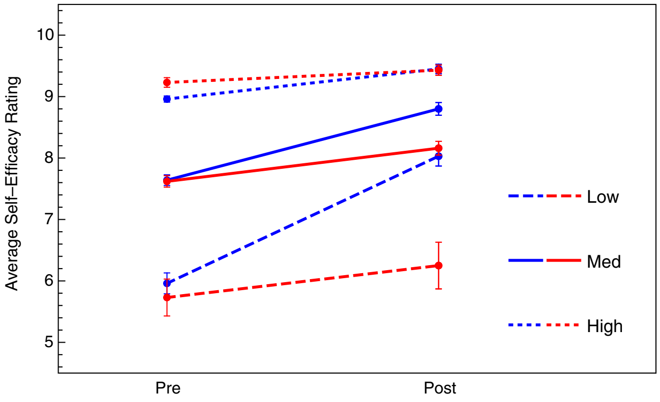

FIG. 7. Mean pre- and postsurvey self-efficacy ratings for students categorized as having low (< 6.5), medium (< 8.5), or high (≥ 8.5) initial self-efficacy for physics (blue lines) and math (red lines). These low, medium, and high rankings are shown by the dashed, solid, and dotted lines, respectively. Error bars represent one standard error of the mean.

themselves highly. We define students who have a "low" initial rating as < 6.5, "medium" as < 8.5, and "high" as ≥ 8.5, and calculated each group's mean pre- and postsurvey ratings. We did this for both the math and physics statements separately, and show it graphically in Fig. 7. There is relatively moderate variation from pre- to postsurvey ratings for the math self-efficacy statements across the low, medium, and high groups. The Cohen's d values for these three categories for math are $d = 0.62$, 0.86, and 0.41, respectively. However, for physics the change from pre- to postsurvey ratings was much greater. The students with the lowest initial self-efficacy see the largest change, $d = 5.14$, with the values for medium and high ratings being $d = 1.98$ and 1.25, respectively. We conclude that the greater change in physics self-efficacy compared with math self-efficacy is due to the students receiving actual instruction in this topic.

1. Student demographics: Gender

The results for the self-efficacy survey in terms of gender are shown in the second two rows of Table VI, where we display the mean and standard deviation of the sample for both topics for male and female students, across the pre- and postsurveys, as well as the number in each sample. It is clear that for both topics, and for pre- and postsurveys, the female students have nearly identical self-efficacy to the male students, but do experience a slightly larger effect size. A three-way, mixed ANOVA reveals no significant interactions between gender and pre- and postsurvey rating, gender and topic, and gender, topic and pre- and postsurvey rating. Levene's test confirmed there was no violation of homogeneity of variance for the physics and math presurvey ratings $[F(1,68) = 0.19, p > 0.05]$ and $[F(1,68) = 2.20, p > 0.05]$, respectively, and for the physics and math postsurvey ratings $[F(1,68) = 0.24, p > 0.05]$ and $[F(1,68) = 2.65, p > 0.05]$, respectively.

Studies investigating the difference in male and female self-efficacy have shown mixed results and highlight that the context of the study is important [22]. We conclude that it is likely that as students self-select to attend the Bootcamp, only those who feel confident in their abilities will consider attending, and therefore it is not surprising that the female students have similar self-efficacy to the male students. However, the female students did experience a slightly greater effect.

2. Student demographics: SES

The results for the self-efficacy survey in terms of SES are shown in the final two rows of Table VI, where we display the mean and standard deviation of the sample for both topics across the pre- and postsurveys for students eligible and not eligible for FSM, as well as the number in each sample. As with gender, but perhaps more surprisingly, the FSM eligible students have very similar initial self-efficacy ratings in both topics compared with their non-FSM peers. Again, this suggests students who have opted to attend the workshop have high self-efficacy initially, which may be a factor in them taking part. Additionally, the FSM eligible students see a smaller effect than their non-FSM peers from the workshop. This is explored in detail below.

The students were randomly assigned to complete either a physics test or a self-efficacy survey, and we therefore expected there to be no differences between these two cohorts of students. However, our results from the self-efficacy survey, as discussed above, are unexpected, and may be a result of the two groups not being homogeneous; i.e., the self-efficacy FSM students may have been particularly self-efficacious and had they completed a physics test, they would have scored equally highly. This is compounded by the fact that the sample sizes for the two FSM cohorts are quite small (13 for the physics test, 11 for the self-efficacy survey). In order to provide evidence that the two groups can be seen as homogeneous, aside from being randomly allocated to their assessments, we analyzed their online Isaac Physics data for the students eligible for free school meals. Of the 13 FSM students that completed the physics test, 12 gave permission for their Isaac Physics online progress to be analyzed, and all of the 11 FSM self-efficacy students gave their permission. We looked at the number of questions they had answered from when they joined Isaac Physics to the day before the Bootcamp, and the number of these that were correct, to give a percentage of correct answers. An independent samples t-test with unequal variances revealed there to be no significant difference between the mean percentage score for the physics test cohort and the self-efficacy cohort $[t(21.0) = 0.41, p = 0.689]$. However, this does not take into account the difficulty of the questions answered by the students, as it may be the case that one cohort answered more difficult questions than the other and was a more able cohort. As the majority of the questions answered by the students are ones set by their teacher, which supplement their school work and therefore the curriculum, we can be relatively sure that both cohorts have answered very similar questions. The Isaac Physics data show no difference between the cohorts in terms of ability, pointing toward the two groups being similar, and any differences are therefore negligible. This is further supported by the fact that the cohorts were randomly allocated to their assignments. Further analysis would be needed to assess this.

However, we do see a clear discrepancy, as the successes the FSM students gain in their physics tests scores do not translate into a greater effect size for their self-efficacy. We believe ultimately this is due to the fact that the students attending the Bootcamp were self-selecting and taking A Level physics beyond what is compulsory in secondary school; therefore a degree of self-efficacy is needed to feel confident enough to attend the Bootcamp. This is shown by their relatively high mean self-efficacy ratings, which are between "moderately certain I can do" and "highly certain I can do." As mentioned previously, although we have provided preliminary evidence to show the two cohorts are homogeneous, this is not conclusive and the FSM students were small sample sizes; therefore there may be very small, undetected, between group differences.

Finally, a three-way, mixed ANOVA reveals no significant interactions between FSM eligibility and pre- and postsurvey rating, FSM eligibility and topic, and FSM eligibility, topic, and pre- and postsurvey rating. Levene's test confirmed there was no violation of homogeneity of variance for the physics and math presurvey ratings $[F(1,62)=0.03, p>0.05]$ and $[F(1,62)=0.01, p>0.05]$, respectively, and for the physics and math postsurvey ratings $[F(1,62) = 0.46, p > 0.05]$ and $[F(1,62) = 1.94, p > 0.05]$, respectively.

C. Isaac Physics data

We can further analyze in detail the students' Isaac Physics usage to investigate whether there is a relationship between their Isaac Physics performance and scoring highly in terms of ability or self-efficacy. As students consented to match their scores to their Isaac Physics data, our subsequent analyses may be biased by the fact that those giving permission may have answered more question parts than those who did not give permission.

We again analyzed the students' Isaac Physics data from the date they joined Isaac Physics to the day before the Bootcamp. For simplicity, we assume the rate at which students answer question parts on a daily basis is uniform, and we found a moderate, positive relationship between the number of question parts answered per day and the pretest score $[r_s = 0.430, p \sim 0, N = 83]$, where r_s is the Spearman's rank correlation coefficient, which is significant. However, the fact that there is a correlation between the number of question parts attempted and pretest score does not necessarily imply that any success in the pretest is simply due to the number of question parts attempted. Furthermore, our assumption of students answering questions uniformly on a daily basis imposes a limitation, as students answer more questions during the school term than during the school holidays, for example.

We found no significant relationships between the number of question parts answered per day and the presurvey self-efficacy rating $[r_s = 0.119, p = 0.152, N = 76]$ or between the percentage of question parts a student answers correctly and their presurvey self-efficacy rating $[r_s = -0.058, p = 0.309, N = 76]$. We note that in the literature, students with higher self-efficacy take higher level math courses, leading to a higher failure rate [52]; however, our results are not statistically significant enough to reject the null hypothesis.

Finally, all students who are eligible for FSM complete on average 0.6 question parts per day (standard deviation 0.7) compared with 1.0 question parts per day completed by those not eligible (standard deviation 1.1). An independent, unequal variances t-test reveals that this is statistically significant $[t(80.2) = 2.19, p = 0.016, N = 32 \text{ (FSM)}, N = 115 \text{ (non-FSM)}]$. These results suggest that students eligible for FSM are doing significantly fewer questions on average than their non-FSM peers, but we cannot infer that this is the reason why they perform poorly in the physics test.

IV. SUMMARY AND DISCUSSION

This study is the first to consider UK-based, secondary school physics students and both their self-efficacy and ability after an active learning-style intervention, with particular emphasis on the demographics of gender and SES. Our study complements existing results and adds to the literature base of intervention studies designed to raise attainment and self-efficacy. Particularly promising is the complete alleviation of a socioeconomic attainment gap for GCSE and A Level questions, and a gender gap for A Level only questions, as we believe the transferable features from the Bootcamp can be applied to other subjects, and that given the right intervention, students can make significant progress.

We performed a quasicontrolled experiment to measure the impact of the Isaac Physics Bootcamp on a cohort of students, and its effect on female students and low SES students. Free school meal eligibility was used as a proxy for low SES. The Isaac Physics Bootcamp consisted of an active learning-style intervention, and we measured students' self-efficacy and ability before and after. Our control for both assessments took the form of material that would be familiar to the students, but was not covered during the Bootcamp sessions. The motivation for our study was that interventions have been shown to alleviate attainment and self-efficacy gaps, for gender or socioeconomic status. Gaps in self-efficacy, attitudes, and perceptions can be alleviated using either active learning [36,37,41], social and emotional learning [21,33], or sociocultural interventions

[31,32]. Attainment and performance gaps can again be alleviated using active learning [35] or interventions based around professional development [11], science literacy [9], group work [10], and instructional approaches [16].

The mean marks for all student's self-efficacy and ability increased significantly, as expected, given that the intervention involved practising physics questions and problem-solving skills. In terms of the control variables, the ratings for the math self-efficacy statements also increased significantly. This may have been because students perceive they have learned even when they receive little or no instruction [42]. A similar effect was seen for the physics test, but not to the same extent, where control questions scores increased slightly but nonsignificantly. This small increase may be a background effect in which the students become more fluent at answering questions and gradually recall how to answer them.

Female students had a lower mean pretest score than male students, as did the low SES students compared with their high SES peers; however, only the latter was statistically significant. On average, the female students' post-test scores increased by more than the male students', and the low SES post-test scores increased by more than the high SES students, implying that the Bootcamp had a greater effect on these two demographics. Again, however, only the FSM increase was statistically significant. The initial socioeconomic attainment gap was completely alleviated for the GCSE and A Level covered questions. The reason why low SES and high SES students experience a different effect may be that the style of the intervention was particularly suited to help lower achieving students; high achieving students already posses the necessary skills to do well and therefore will not demonstrate as much of an effect [15,16]. Further analysis revealed a gender attainment gap when considering the harder, A Level only questions, which the Bootcamp did manage to alleviate.

The initial socioeconomic attainment gap was eliminated completely for the GCSE and A Level questions, but not for the A Level only questions as these harder questions saw the largest gap in the pretest. Thus the Bootcamp provided FSM students with the skills to do well, but not for the hardest questions, suggesting further interventions or support would be needed. The corresponding results for gender were similar, but watered down, in the sense that there was no attainment gap for the GCSE and A Level questions, but there was for the harder A Level questions. This gap in A Level attainment by gender was not as large as the corresponding one for socioeconomic status; thus the Bootcamp was able to alleviate this completely for female students. This is to be expected as the overall gap for gender was smaller than that for socioeconomic status, which is a well-known result and described in detail in Sec. I. The Bootcamp can therefore alleviate moderate attainment gaps, but does not provide the necessary resources to reduce the largest discrepancies in attainment. We therefore identify an interaction between how much a demographic underperforms and the difficulty of the questions being attempted. The Bootcamp provides students from a significantly underperforming background with the skills to tackle intermediate-level questions, but not the hardest questions. Students from only a mildly underperforming background, however, can make significant progress in the hardest questions.

Students who initially had a low self-efficacy rating saw the largest increase in their rating for the physics statements, compared to students with a higher initial self-efficacy rating. Students build a picture of their own self-efficacy based on their experiences in four categories—mastery experiences (own success at similar tasks), vicarious learning (seeing others be successful), verbal persuasion (praise or judgement from peers and teachers), and their physiological state (feeling incapable, anxiety) [25,26]. Students with low self-efficacy therefore may have rated themselves in this way due to previous experiences, and the active learning, group work element of the Bootcamp, the feedback and support provided by staff, and the experience itself of tackling the problems would have allowed those students to build their self-efficacy. Students with higher self-efficacy may already have been exposed to the successes of others, verbal praise from teachers, and their own mastery of similar questions, so would not feel the benefit of these features to the same extent. In addition, students with lower self-efficacy typically tackle easier material, and as everyone at the Bootcamp was given the same problems, these students were able to see that they could solve the questions, resulting in a greater increase in self-efficacy than students who would already naturally tackle such questions.

No differences were seen in presurvey ratings for the physics and math statements for either male and female students or FSM and non-FSM students. These findings contrast with previous studies, although findings for gender based self-efficacy differences remain mixed. We conclude our cohort is initially quite confident, as shown by their high self-efficacy scores. For the math statements, the female students see a medium effect ($d = 0.49$), while the male students only see a small effect ($d = 0.23$). The effect of the Bootcamp on the self-efficacy of the two demographics was not as expected; on average, the female students' postsurvey ratings increased by more than the male students'; however, the high SES postsurvey ratings increased by more than the low SES students. Although the FSM eligible students see a larger change in terms of ability, this does not translate into the same larger change for self-efficacy, and suggests there are still barriers to improving their self-efficacy.

The findings of this particular study are limited by the fact that students self-selected to attend the Bootcamp. They may have perceived themselves as capable and confident, whereas other students may have been put off applying to the Bootcamp for a variety of reasons. In addition, our setup was quasiexperimental, in that we only considered students attending the Bootcamp, and consequently, our control variables saw small increases in self-efficacy and ability, which we believe are background effects. Furthermore, the self-efficacy survey contained a total of 8 control statements, while the physics test only had questions 3 and 9 as a control; therefore, this smaller set of 2 control questions imposes another limitation of our study.

Future work could be to follow up the progress of the students who attended the Bootcamp and assess whether attendance leads to increased usage on Isaac Physics. Isaac Physics also contains math and chemistry questions, and there is the sister site Isaac Computer Science, so there is potential for a more general STEM or computing Bootcamp which would increase participation and attainment [53]. Finally, future work could involve a dedicated experiment to investigate whether answering more Isaac question parts causes students to perform better in physics tests.

A related concept to self-efficacy is that of anxiety, with students having low self-efficacy reporting more anxiety about their school subjects. Most research has focused on math and/or science anxiety and their gender gaps [21,54–56]; therefore a possible avenue for future exploration would be to measure students' physics anxieties, and how this relates to ability and self-efficacy.

Finally, along with gender and socioeconomic status, ethnicity is also a factor in a student's educational achievement. These three variables do not play an equal role, however, as we have seen with gender and FSM status. It is suggested that the social class differences are 3 times larger than the ethnicity gap, and 6 times larger than the gender gap [57]. The ethnicities that perform above average are Indian and Chinese, while ethnicities with below average achievement include Black Caribbean, mixed White and Black Caribbean, and Pakistani [58], as well as White low SES when considering interacting socioeconomic factors [59]. Certain minority ethnicities are underrepresented among high SES students, but overrepresented among low SES students, suggesting a strong overlap between these two factors. As a result, future work for Isaac Physics could investigate the impact of the project on these potential differences in ability and self-efficacy.

If possible, we encourage teachers of A Level physics students to strongly consider the Bootcamp for their students, provided they are eligible, as we have shown it is particularly beneficial for students from underrepresented backgrounds in the physical sciences. For those teachers unable to send their students to the Bootcamp, we believe the features demonstrated in the workshop can be transferred to other subjects and other academic levels.

ACKNOWLEDGMENTS

The authors would like to thank James Sharkey and Ben Hanson for their help and contributions with data collection. This work was funded by the Department for Education and The Ogden Trust as part of the Isaac Physics project.

APPENDIX: ASSESSMENTS

1. Self-efficacy survey

For each of the 24 statements below (1–24), please indicate your confidence using a 10 point scale where: 1 = *highly certain* I **CANNOT** do; 5 = *moderately certain* I **can do**; 10 = *highly certain* I **CAN DO**.

1. For two independent events A and B, if I know the probability of A occurring and the probability of B occurring, I can calculate the probability of both events (A and B) occurring.
2. I can calculate the effective (total) resistance of a combination of resistors (e.g., 2 or 3 resistors in series and/or parallel).
3. I can apply equations for uniformly accelerated motion in one dimension.
4. I can calculate the refractive index of a material, using values of angle of incidence (i) and angle of refraction (r) for a ray entering the material from air (or a vacuum).
5. When a laser beam is pointed at a diffraction grating, I can use a formula to find the angles of the points of maximum brightness.
6. I can calculate the voltage output from a potential divider circuit.
7. I can calculate the acceleration of two connected particles (for example, two unequal masses hanging on a smooth pulley).
8. If I am given a formula of the form $a = bc^2 \div d$, I can find the percentage increase/decrease in a quantity (e.g., a) if I know the percentage change of the quantities in the formula (e.g., b increases by 20%, c doubles and d doubles).
9. I can calculate the standard deviation of a set of data.
10. I can calculate the sum of the first n terms of an arithmetic series.
11. I can apply the equations for uniformly accelerated motion to solve projectile problems in two dimensions.
12. I can use Pythagoras's theorem to find the resultant of two vectors (one horizontal and one vertical).
13. If I know the amplitude and intensity of a polarised electromagnetic wave, I can find the amplitude and intensity after it has passed through a second polariser (oriented at a different angle).
14. I can calculate the gradient of a polynomial function at a point (e.g., find the gradient of the function $y = 3x^5 + 7x^2$, when $x = 4$).
15. I can calculate a definite integral of a polynomial function (e.g., $y = x^4$, between $x = 3$ and $x = 7$).
16. I can calculate the sum of a convergent geometric series (one that has a finite sum).
17. I can calculate meaningful quantities from gradients and areas under graphs (e.g., displacement from a velocity time graph).

18. I can convert unit prefixes and powers of ten (e.g., $mm^2 \rightarrow$ x10^{-6} m).
19. I can change the subject of (rearrange) equations (e.g., $V = IR$, or $E = \frac{1}{2}mv^2$).
20. I can find the components of a vector (using trigonometry).
21. I can calculate the roots of a quadratic equation.
22. I can use a formula to find the drift velocity of the charge carriers in a piece of uniform conductor.
23. I can use observations of patterns of electromagnetic standing waves (e.g., microwaves) to find the frequency of the waves.
24. I can calculate the internal resistance and/or emf of a cell using measurements of current when two different resistors are connected to the cell.

2. Physics test

Show the correct answer by circling one letter for each question on the sheet. There is only one correct answer for each question.

1. [Pre-test] A plane at a height of 400 m flies horizontally at a speed of 340 $\mathrm{km\,h^{-1}}$. It releases a package which falls to the ground. Ignoring air resistance, what horizontal distance has the package travelled between being released and landing?

A. 88 m
B. 400 m
C. 600 m
D. 850 m
E. 3100 m

2. [Post-test] A child slides a wooden brick off a kitchen table which stands 70 cm above the floor. The brick slides at 1.2 $\mathrm{m\,s^{-1}}$ along the table. Ignoring air resistance, what horizontal distance has the brick travelled between the edge of the table and landing point on the ground?

A. 17 cm
B. 23 cm
C. 45 cm
D. 59 cm
E. 318 cm

[1] Joint Council for Qualifications, GCE A Level and GCE AS Level summer results 2019, https://www.jcq.org.uk/wp-content/uploads/2019/08/A-Level-and-AS-Results-Summer-2019.pdf.
[2] Joint Council for Qualifications, GCSE summer results 2019, https://www.jcq.org.uk/wp-content/uploads/2019/09/GCSE-Full-Course-Summer-2019-%E2%80%93-Outcomes-for-main-grade-set-by-jurisdiction.pdf.
[3] S. Gorard, B. H. See, E. Smith, S. Bevins, E. Brodie, and M. Thompson, Exploring the relationship between socioeconomic status and participation and attainment in science education, The Royal Society, SES and Science Education Report, 2008, ISBN: 978-0-85403-699-8.
[4] T. Nunes, P. Bryant, S. Strand, J. Hillier, R. Barros, and J. Miller-Friedmann, Review of SES and science learning in formal educational settings, University of Oxford, Education Endowment Foundation Report, 2017.
[5] S. Gorard and B. H. See, The impact of socio-economic status on participation and attainment in science, Stud. Sci. Educ. **45**, 93 (2009).
[6] J. Ireson, S. Hallam, and C. Hurley, What are the effects of ability grouping on GCSE attainment?, Br. Educ. Res. J. **31**, 443 (2005).
[7] Institute of Physics, Raising aspirations in physics, http://iop.cld.iop.org/publications/iop/2014/file_64463.pdf.
[8] L. Archer, E. Dawson, J. DeWitt, A. Seakins, and B. Wong, “Science capital”: A conceptual, methodological, and empirical argument for extending Bourdieusian notions of capital beyond the arts, J. Res. Sci. Teach. **52**, 922 (2015).
[9] B. Hand, L. A. Norton-Meier, M. Gunel, and R. Akkus, Aligning teaching to learning: A 3-year study examining the embedding of language and argumentation into elementary science classrooms, Int. J. Sci. Math. Educ. **14**, 847 (2016).
[10] E. Baines, P. Blatchford, and A. Chowne, Improving the effectiveness of collaborative group work in primary schools: Effects on science attainment, Br. Educ. Res. J. **33**, 663 (2007).
[11] P. Hanley, R. Slavin, and L. Elliott, Thinking, doing, talking science: Evaluation report and executive summary, http://eprints.hud.ac.uk/id/eprint/29812/.
[12] E. Furtak, T. Seidel, H. Iversen, and D. Briggs, Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis, Rev. Educ. Res. **82**, 300 (2012).
[13] A. Lazonder and R. Harmsen, Meta-analysis of inquiry-based learning: Effects of guidance, Rev. Educ. Res. **86**, 681 (2016).
[14] O. Lee, R. Deaktor, C. Enders, and J. Lambert, Impact of a multiyear professional development intervention on science achievement of culturally and linguistically diverse elementary students, J. Res. Sci. Teach. **45**, 726 (2008).
[15] J. Fouche, The effect of self-regulatory and metacognitive strategy instruction on impoverished students’ assessment achievement in physics, Ph.D. thesis, Liberty University, 2013.
[16] B. Y. White and J. R. Frederiksen, Inquiry, modeling, and metacognition: Making science accessible to all students, Cognit. Instr. **16**, 3 (1998).

[17] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychol. Rev. **84**, 191 (1977).

[18] S. Lau and R. Roeser, Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science, Educ. Assess. **8**, 139 (2003).

[19] T. Lyons and F. Quinn, Choosing science: Understanding the declines in senior high school science enrolments, https://eprints.qut.edu.au/68725/.

[20] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Damage caused by women's lower self-efficacy on physics learning, Phys. Rev. Phys. Educ. Res. **16**, 010118 (2020).

[21] M. Griggs, S. Rimm-Kaufman, E. Merritt, and C. Patton, The responsive classroom approach and fifth grade students' math and science anxiety and self-efficacy, Sch. Psychol. Q. **28**, 360 (2013).

[22] S. Britner and F. Pajares, Self-efficacy beliefs, motivation, race, and gender in middle school science, J. Women Minorities Sci. Eng. **7**, 10 (2001), https://ui.adsabs.harvard.edu/abs/2001JWMSE...7d..10B/abstract.

[23] D. Bakan Kalaycioğlu, The influence of socioeconomic status, self-efficacy, and anxiety on mathematics achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and the USA, Educ. Sci. Theor. Pract. **15**, 1391 (2015), https://files.eric.ed.gov/fulltext/EJ1101308.pdf.

[24] S. Yerdelen-Damar and H. Peşman, Relations of gender and socioeconomic status to physics through metacognition and self-efficacy, J. Educ. Res. **106**, 280 (2013).

[25] S. Britner and F. Pajares, Sources of science self-efficacy beliefs of middle school students, J. Res. Sci. Teach. **43**, 485 (2006).

[26] A. Bandura, *Self-Efficacy: The Exercise of Control* (W. H. Freeman, New York, 1997).

[27] Institute of Education, Subject choice in STEM: Factors influencing young people (aged 14–19) in education, https://discovery.ucl.ac.uk/id/eprint/10006727/1/wtx063082.pdf.

[28] M. Homer, J. Ryder, and I. Banner, Measuring determinants of post-compulsory participation in science: A comparative study using national data, Br. Educ. Res. J. **40**, 610 (2014).

[29] Institute of Physics, Why not physics? A snapshot of girls' uptake at A-Level, https://www.iop.org/sites/default/files/2018-10/why-not-physics.pdf.

[30] P. Finegold, P. Stagg, and J. Hutchinson, Good timing: Implementing STEM careers strategy in secondary schools, http://hdl.handle.net/10545/621262.

[31] R. Amos and M. Reiss, The benefits of residential fieldwork for school science: Insights from a five-year initiative for inner-city students in the UK, Int. J. Sci. Educ. **34**, 485 (2012).

[32] C. Gartland, Student ambassadors: 'Role-models', learning practices and identities, Br. J. Sociol. Educ. **36**, 1192 (2015).

[33] J. Zins and M. Elias, Social and emotional learning: Promoting the development of all students, J. Educ. Psychol. Consult. **17**, 233 (2007).

[34] C. Bonwell and J. Eison, Active learning: Creating excitement in the classroom, J-B ASHE, Higher Education Report Series (AEHE), 1991, ISBN-1-878380-06-7.

[35] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U.S.A. **111**, 8410 (2014).

[36] L. Deslauriers, E. Schelew, and C. Wieman, Improved learning in a large-enrollment physics class, Science **332**, 862 (2011).

[37] T. Espinosa, K. Miller, I. Araujo, and E. Mazur, Reducing the gender gap in students' physics self-efficacy in a team- and project-based introductory physics class, Phys. Rev. Phys. Educ. Res. **15**, 010132 (2019).

[38] B. Toggerson, S. Krishnamurthy, E. E. Hansen, and C. Church, Positive impacts on student self-efficacy from an introductory physics for life science course using the team-based learning pedagogy, arXiv:2001.07277.

[39] E. M. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?, Phys. Rev. Phys. Educ. Res. **14**, 020123 (2018).

[40] J. Nissen, Gender differences in self-efficacy states in high school physics, Phys. Rev. Phys. Educ. Res. **15**, 013102 (2019).

[41] M. Good, A. Maries, and C. Singh, Impact of traditional or evidence-based active-engagement instruction on introductory female and male students' attitudes and approaches to physics problem solving, Phys. Rev. Phys. Educ. Res. **15**, 020129 (2019).

[42] L. Deslauriers, L. S. McCarty, K. Miller, K. Callaghan, and G. Kestin, Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom, Proc. Natl. Acad. Sci. U.S.A. **116**, 19251 (2019).

[43] Isaac Physics, https://isaacphysics.org/.

[44] Isaac Physics, Impact and engagement summary, https://cdn.isaacphysics.org/isaac/publications/impact_summary_201804_v6.pdf.

[45] Department for Education, Schools, pupils and their characteristics: January 2020, https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2020.

[46] S. Ilie, A. Sutherland, and A. Vignoles, Revisiting free school meal eligibility as a proxy for pupil socio-economic deprivation, Br. Educ. Res. J. **43**, 253 (2017).

[47] C. Taylor, The reliability of free school meal eligibility as a measure of socio-economic disadvantage: Evidence from the millennium cohort study in Wales, Brit. J. Educ. Stud. **66**, 29 (2018).

[48] J. Barrows, S. Dunn, and C. A. Lloyd, Anxiety, self-efficacy, and college exam grades, Univ. J. Educ. Res. **1**, 204 (2013), https://files.eric.ed.gov/fulltext/EJ1053811.pdf.

[49] A. Bandura, in *Self-Efficacy Beliefs of Adolescents*, Adolescence and Education, edited by F. Pajares and T. Urdan (Information Age Publising, Inc., Greenwich, CT, 2006).

[50] K. S. Taber, The use of Cronbach's alpha when developing and reporting research instruments in science education, Res. Sci. Educ. **48**, 1273 (2018).

[51] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge, New York, 1988).

[52] H. Watt, The role of motivation in gendered educational and occupational trajectories related to math, Educ. Res. Eval. **12**, 305 (2006).

[53] Isaac Computer Science, https://www.isaaccomputerscience.org/.

[54] A. Dowker, A. Sarkar, and C. Y. Looi, Mathematics anxiety: What have we learned in 60 years?, Front. Psychol. **7**, 508 (2016).

[55] F. Hill, I. C. Mammarella, A. Devine, S. Caviola, M. C. Passolunghi, and D. Szũcs, Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity, Learning and individual differences **48**, 45 (2016).

[56] M. K. Udo, G. P. Ramsey, and J. V. Mallow, Science anxiety and gender in students taking general education science courses, J. Sci. Educ. Technol. **13**, 435 (2004).

[57] S. Strand, Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and ‘getting it’ for the white working class, Res. Papers Educ. **29**, 131 (2014).

[58] C. Parsons, Social justice, race and class in education in England: Competing perspectives, Cambridge J. Educ. **49**, 309 (2019).

[59] S. Strand, School effects and ethnic, gender and socioeconomic gaps in educational achievement at age 11, Oxf. Ref. Educ. **40**, 223 (2014).