Institute for Fiscal Studies

How can we increase girls' uptake of maths and physics A-level? Recommendations regarding next steps for the STEM Skills Fund

Rachel Cassidy
Sarah Cattan
Claire Crawford

ESRC
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

How can we increase girls' uptake of maths and physics A-level? Recommendations regarding next steps for the STEM Skills Fund

Rachel Cassidy
Sarah Cattan
Claire Crawford

Published by

The Institute for Fiscal Studies

1. Recommendations to SSF

Rolling out the scholarship vs. something else

- The evidence collected through the pilot suggests that **a scholarship program along the lines of the one we trialled could increase the likelihood that girls apply for maths or physics A-level**. However, for the reasons discussed in detail in the main report, we would expect the scholarship to have a smaller impact than that suggested by the girls' answers to the pupil questionnaire. Moreover, the main mechanism through which the scholarship operates is unclear. The focus group discussions suggested that it might be seen as **a reward for the extra effort of undertaking subjects that are perceived as being the most difficult**. But financial support is not an obvious response to this potential barrier, nor to any of the other potential barriers identified by significant numbers of girls in the study. While the scholarship might be effective, therefore, there may be alternative interventions that could be just as if not more effective, potentially at lower cost. **We would therefore recommend to the SSF that they consider trialling other such interventions in addition to the scholarship.**

- The evidence provided in the study suggests that helping schools to offer **more STEM work experience placements, more opportunity to interact with female role models working in STEM and boosting girls' confidence in STEM subjects** could help to raise the continuation rates to maths and physics A-level amongst high-achieving girls.

Improving the potential (cost-) effectiveness of the scholarship

- **We recommend that the scholarship is provided to both FSM and non-FSM girls.** The gender gap in maths and physics is similar across these two groups, and our field work does not suggest that the scholarship would be likely to have a greater impact on FSM girls than on non-FSM girls. Moreover, as discussed in detail below, offering the scholarship to FSM girls only would require a very large number of schools – more than 500 per arm – to participate in a trial in order to be able to detect statistically significant effect sizes, which seems unlikely to be feasible.

- The trial did not provide a strong sense that the larger payments offered in some treatment arms were more effective at increasing continuation rates than the smaller payments offered in other treatment arms. Nor did it appear that offering some of the money upfront to all girls predicted to achieve at least a grade 7 in maths, physics or combined science GCSE lead to higher continuation rates in those treatment arms in which payments were split in this way. This suggests that **offering the lower scholarship amount and making payments conditional on applying to study A-level maths or physics might be sufficient to reproduce similar effect sizes** to those seen across all arms of the trial.

- To increase the chances that the scholarship has an impact on girls' choices, **we recommend the SSF to consider the option of telling schools and potentially eligible students about the scholarship in Year 10 or even earlier**. Many girls said

that the scholarship came too late, at a time when their choices were already made. An earlier promise of payment could also incentivise greater effort be put into STEM subjects from an earlier age and hence increase girls' confidence about their ability to take maths or physics A-levels when it is time to make their A-level subject choices. If this was not deemed possible or desirable, then telling students about the scholarship as early as possible in Year 11 would still be an improvement.

Improving the administration of the scholarship and the trial

- During the pilot, we experienced great difficulty in recruiting schools into the control group and gathering information on eligible girls from these schools. Should the SSF want to implement a larger-scale trial, **it is very important to think about what incentives can be offered to the control group** so that enough schools can be recruited and encouraged to provide information about girls' choices.

- The implementation of the trial, even on the scale of the pilot, posed a very large administrative burden on the research team. There is a lot of paperwork involved (e.g. the distribution and collection of school and pupil consent forms), and a large amount of time was spent communicating and following-up with schools and keeping records of all the information provided by schools. **Should the SSF want to implement a larger-scale trial, it will be important that enough resources are allocated to hiring a project manager who can perform these tasks.**

- **We recommend that the scholarship is distributed as a one-off payment** or that the payments are made no more frequently than once a month. Several schools reported administrative difficulties in making weekly payments. Although we did not ask girls specifically about their preferences over the frequency of payment, we do not believe there is a strong argument for giving weekly payments over less frequent payments.

Required scale of a potential new trial

- If the SSF were to run a new trial to evaluate the effectiveness of the scholarship in a way that would stand a good chance of detecting a statistically significant effect, we would **recommend that a minimum of 176 schools is recruited to the trial** (78 to the treatment group and 98 to the control group), **to try to ensure a final achieved sample size of 56 schools per arm** after allowing for attrition. This assumes a baseline continuation rate of 46.8% and an effect size of 8.2 percentage points, which we still believe is likely to be an upper bound of the effect size that the scholarship could realistically achieve. **This would imply that scholarship payments of around £177,000 would need to be financed, with administration and evaluation costs on top of this**.

- **We also recommend that an additional treatment arm be added to the trial to test the effectiveness of an alternative (potentially lower cost) intervention**. This would necessitate the recruitment of an additional 78 schools to a separate treatment arm in order to detect an effect size of similar magnitude. Such an approach could contribute significantly to the body of available evidence on 'what works' in this area,

and – if a low or no cost intervention, such as the provision of information, were chosen – could do so at the cost of recruiting and retaining these schools alone.

2. Powering a potential new trial

The pilot trial of the scholarship did not include a large enough number of schools to allow us to make statistically robust conclusions with respect to the effect of the scholarship on the continuation rates of girls. However, as discussed in detail in the main report, the evidence we gathered from the pupil questionnaire does suggest that the scholarship could induce girls who are on the margin of applying for A-levels in maths or physics to actually make these choices.

Based on the estimates we obtained, we can perform power calculations that indicate how many schools and eligible girls we would need in a future randomised controlled trial in order to detect impacts of particular magnitudes. As with the pilot trial, we have assumed that this large-scale trial would randomise the treatment at the school level, so that every eligible girl in a treated school would be offered the scholarship and every eligible girl in a control school would not. Unlike the pilot trial, we are assuming that there would be only one treatment arm in which all girls would receive the same amount of money.

Based on the evidence from the pilot trial, our 'best guess' of the likely impact of the scholarship is for it to increase the probability of applying for maths or physics A-level by at most 8.2 percentage points, from a base of 46.8%. This takes the responses to the pupil survey at face value and assumes that all girls who did not respond to our survey but who applied for maths or physics A-level would have done so anyway, i.e. that none changed their A-level subject choices as a result of being offered the scholarship. We believe this is likely to represent an upper bound of the effect of the scholarship on girls' A-level choices.

We perform these calculations under two assumptions about the number of eligible girls in each school: a) that there are 24 eligible girls (which is the average number of FSM and non-FSM girls who are predicted to achieve at least a grade 7 in maths, physics or combined science GCSE in schools in the pilot trial), and b) that there are 2 eligible girls (which is the average number of FSM girls who are predicted to achieve at least a grade 7 in these subjects in schools in the pilot trial). All calculations are performed so that impacts are detected with 95% statistical confidence.

The top panel of Table 1 below reports the number of eligible girls per arm and the number of schools per arm that would be needed to significantly detect these effect sizes given the baseline application rates.

Table 1. Power calculations for trial to detect significant increase in application rates for maths or physics A-level

	(1) Baseline continuation rate*	(2) Minimum detectable effect size	(3) Achieved number of eligible girls per arm	(4) Achieved number of schools per arm	(5) Initial number of TREATMENT schools required assuming attrition rate of 28%**	(6) Initial number of CONTROL schools required assuming attrition rate of 43%***	(7) Estimated cost of scholarship payments (£240 per eligible girl) ****
	Recommended minimum sample sizes (effect size of 0.082; intra-cluster correlation of 0.03)						
FSM and non-FSM girls eligible for scholarship (avg. of 24 per school)	0.468	0.082	1,344	56	78	98	£177,408
FSM girls only eligible for scholarship (avg. of 2 per school)	0.468	0.082	802	401	557	704	£105,864
	More pessimistic scenarios (assuming FSM and non-FSM girls would be eligible, i.e. assuming an average of 24 eligible girls per school)						
Smaller effect size	0.468	0.0615	2,376	99	138	173	£301,942
Higher intra-cluster correlation (0.05)	0.468	0.082	1,704	71	99	125	£224,928

* proportion of eligible girls applying for maths or physics A-level in the absence of the scholarship

** assumes that 72% of the schools in the treatment group who agreed to participate in the trial return data on girls' A-level subject choices

*** assumes that 57% of the schools in the control group who agreed to participate in the trial return data on girls' A-level subject choices

**** assumes the percentage of girls in the treatment arm who go on to take maths/physics (and so receive payment) is equal to the baseline continuation rate plus the effect size.

To detect an 8.2 percentage point increase in the probability of applying for maths or physics A-level from an assumed baseline of 46.8%, with the scholarship being offered to an average of 24 high-achieving girls in each school, the first row of Table 1 shows that the trial would need to include 56 schools in each arm, i.e. 112 schools in total (split equally between one treatment group and one control group). This would cost SSF approximately £177,408 in scholarship payments (assuming that 55% (46.8%+8.2%) of the 1,344 eligible girls in the final achieved treatment group applies for maths or physics A-level, and that each of these students receives £240). Of course, it would be prudent to set aside additional budget as a buffer, in case the treatment proved more effective than our best estimates and hence more girls went on to take maths/physics and claim a scholarship payment.

If instead the scholarship was only offered to FSM girls and there was an average of 2 high-achieving FSM girls per school, the trial would need to include a much larger number of schools. Specifically, under the same assumptions regarding the baseline continuation rate (46.8%) and effect size (8.2 percentage points), the trial would require the participation of 401 schools in each arm, and would cost an estimated £105,864 in scholarship payments.

As outlined in the main report, the pilot suffered from attrition (at school level) between the time that the school initially agreed to participate in the trial and the time when the data on eligible girls' A-level subject choices was requested. It is the number of schools that provide this data which is relevant for the power calculations, so it would seem prudent to start a new trial with a larger number of schools in order to try to achieve the final sample sizes outlined in Table 1 above.

To provide a sense of how many schools would need to agree to participate in a new trial in order to achieve this aim, we have applied the attrition rates identified in the main report – calculated separately for the treatment and control groups – to the estimates of the final number of schools required in the table above. The results are provided in Columns 5 and 6 of Table 1. This suggests that to achieve a final sample size of 56 schools in the treatment group, 72 would need to initially agree to participate in the trial. This assumes that we will see the same attrition rate in the treatment group (22%) between the start and end of the trial as we did in the pilot. Similarly, to achieve the same number of schools in the control group, 98 would need to agree to participate in the trial. Again, this assumes that we will see the same attrition rate in the control group (43%) as we did in the pilot. Recruiting nearly 100 schools to be part of a control group in a trial will be extremely challenging to achieve, so consideration should be given to potential incentives that could be offered to schools in the control group to encourage them to join and remain in the trial.

Any decision regarding the size of a new trial must of course balance the desire to detect a significant effect with the need to design a trial that is both affordable and feasible to deliver (in terms of recruiting sufficient numbers of schools). **It is our view that the power calculations set out in the first row of Table 1 represent the minimum size of trial that should be considered by the SSF in order to stand a reasonable chance of detecting a statistically significant effect, while still being feasible to run from an administrative point of view and (hopefully) affordable from a fundraising perspective.**

However, power calculations are based on a number of assumptions. If one or more of the assumptions we have made are too optimistic, then a trial of this size may still not be sufficient to identify a statistically significant effect, even if the scholarship did encourage some girls to change their A-level subject choices. The bottom panel of Table 1 illustrates the importance of some of these underlying assumptions. The first row in this panel shows that if the effect size were to be three-quarters of the magnitude we have assumed in the top panel – i.e. if it were to be 6.15 percentage points rather than 8.2 percentage points – then the achieved number of schools in the treatment and control group would need to almost double, from 56 to 99 per arm, in order to be sure that this effect was significantly different from zero.

The second row shows what happens if we increase the 'intra-cluster correlation' – that is, the extent to which the A-level subject choices of high achieving girls are correlated within schools. Such correlation might arise if, for example, there was a particularly charismatic physics teacher in one school who encouraged all but one of the 10 high-achieving girls in their school to apply for maths or physics A-level. This is problematic when trying to identify the impact of an intervention designed to change these choices because it makes it more difficult to separate the effects of these factors that are common to all girls within a particular school from the potential effects of the scholarship. We have assumed an intra-cluster correlation of 0.03 in our preferred power calculations, as this is the degree of correlation estimated from historic data on continuation rates amongst the schools who expressed interest in participating in the trial. But this is very low, and if the true intra-cluster correlation were to be slightly higher –e.g. 0.05 rather than 0.03 – then we would need an extra 15 schools per arm in order to illustrate that an 8.2 percentage point effect size can be statistically distinguished from zero.