Role models in movies: the impact of *Queen of Katwe* on students' educational attainment*

Emma Riley†

June 1, 2021

Abstract

This paper presents experimental evidence on the impact of a role model on secondary school student exam performance in Uganda. Students preparing to take their national exams were individually randomised to see either a movie featuring a female role model from a similar background succeeding at chess, *Queen of Katwe*, or to see a placebo movie. I find that treatment with the role-model movie leads to lower secondary school students being less likely to fail their maths exam a week later: 85% of those who watched *Queen of Katwe* passed the exam, whereas only 73% of those who didn't passed. This effect is strongest for female and lower ability students. For upper secondary school students, treatment with *Queen of Katwe* one month before their exams results in an increase in their total exam score of 0.13 standard deviations. Effects of treatment are strongest for female students in the compulsory mathematics paper. In both classes, female treated students are more likely to remain in education in subsequent years, closing the gender gap with their male peers. This study highlights the power of a movie role model as a way to improve secondary school students' educational attainment and close gender gaps.

JEL Codes: J24, J16, I25, I21, O12

Keywords: role models, uganda, education, secondary school, media

*I would like to thank Richard Sedlmayr for his contribution to the initial development of this study, and Edward Kisitu, Rose Kusingura and Rachel Proefke for their assistance with logistics in Uganda, particularly recruiting schools and transportation for the students. I would like to thank Margarita Diego for research assistance. I would like to thank the staff of the Initiative for Social and Economic Rights (ISER) in Uganda for their support recruiting schools and assistance at the cinema screenings. I would like to thank my supervisors, Climent Quintana-Domeque and Stefan Dercon for their comments, advice and support. I thank an anonymous donor for generous financial support of this research. The trial is registered at https://www.socialscienceregistry.org/trials/1832/history/14361 and the pre-analysis plan was uploaded there on 23rd February 2017 before endline data collection had finished and analysis begun. Three amendments to the PAP were also uploaded there before any additional data collection was finished. All analysis in this paper follows these pre-analysis plans unless clearly stated otherwise. Ethical approval was obtained from Oxford University on the 30th September 2016 with reference ECONCIA1617-002.

†Department of Economics, Manor Road Building, Oxford OX1 3UQ, UK. Email: emma.riley@economics.ox.ac.uk

1 Introduction

Worldwide, large gender gaps remain in education. While enrolment rate differences by gender are almost closed in developed countries, they still persist in developing countries, with 40% of men achieving some secondary school compared to only 29% of women (UNESCO Institute for Statistics, 2018). This is concerning, since each year of secondary school increases future earnings. What's more, gaps in maths performance between girls and boys are present across both the developed and developing world and girls are less likely to choose maths and sciences as subject choices (OECD, 2015; Dickerson et al., 2015). These gaps are problematic because studying mathematics or science subjects leads to higher paying jobs (Kirkeboen et al., 2016).

Role models have successfully been used to close gender gaps in mathematical subjects in a wide range of contexts (Carrell et al., 2010; Breda et al., 2018; Dennehy and Dasgupta, 2017; Kofoed and Mcgovney, 2016; Porter and Serra, 2018). How do role models affect behaviours? Firstly, role models can act as a powerful way to change beliefs about what is possible and raise aspirations (Beaman et al., 2012; Bernard et al., 2014). As a result, they can cause people to reassess and raise their goals, ambitions and effort level, breaking them out of an aspirations induced poverty trap (Dalton et al., 2016; Lybbert and Wydick, 2018). Secondly, stereotypes about women's ability in maths and science are widespread (Nosek et al., 2009; Carlana, 2019; Spencer et al., 1999), and role models have been successfully used to overcome these stereotypes by providing a counter example (Dasgupta and Asgari, 2004; Marx and Roman, 2002; Stout et al., 2011). Finally, role models have also been used to provide information about returns to schooling (Nguyen, 2008), which affects effort (Jensen, 2010, 2012). In addition to the evidence on real-life models, role models in the form of media have been also shown to impact beliefs, aspirations and behaviours (La Ferrara, 2016; La Ferrara et al., 2012; Chong and Ferrara, 2009; Jensen and Oster, 2009).

In this paper, I examine whether a movie featuring a potential role model can improve the educational outcomes of students, particularly female students. I do this through the randomised exposure of 1,500 secondary school students in Kampala, Uganda to a treat-

ment movie, *Queen of Katwe*, featuring a female role model excelling academically in a mathematical domain, versus a placebo movie. Students preparing to take important national exams at the end of lower and upper secondary school were individually randomised to see the treatment or placebo movie between 1 week and 1 month before their exams. Using administrative data, I examine impacts on both immediate exam performance 1 week- 1 month later as well as longer term outcomes related to continuation in education, allowing me to examine both short run and persistent effects of the role model treatment. The use of a placebo movie, here *Miss Peregrine's Home for Peculiar Children*, allowed me to exclude any beneficial effect to exam performance from the novelty of going to the cinema and media exposure in general (Bernard et al., 2015).

The treatment examined here, the movie *Queen of Katwe*, is based on the true story of a teenage girl from the slums of Kampala, Uganda striving to become a chess master through hard work and perseverance. Along the way she must overcome many obstacles to achieving her dream, including learning to read and write and getting into the top school in Uganda in order to play chess. The protagonist featured here is relatable to teenage students in Uganda, particularly female students and students who have done less well academically, for whom she is most similar. She may therefore act as a role model, similar and attainable to the audience (Ray, 2006). Through watching her story, student may change their beliefs about their ability to be successful, particularly in subjects they believe they are not good at, change their beliefs about returns to succeeding at school, or form higher aspirations, leading to behavioural changes towards increased motivation or effort.

I find that among students taking exams for lower secondary school, seeing the treatment movie results in a 0.11 standard deviation improvement in maths scores. This is similar in magnitude to another study that looked at the impact of a role model on exam performance in a developing country (Nguyen, 2008). Decomposing this improvement in maths score into the effect on the probability of achieving each grade (A-F), I find the entire effect is coming from a 30% decrease in the probability a student fails maths. I find no effect of the treatment movie on the probability of achieving any particular grade

in other compulsory subjects than maths or on a student's aggregate exam score.

When examining these findings by pre-defined subgroups, I find that it's female students and students performing the worst prior to the exam who benefit most from treatment. Female students go from failing their maths exam 32% of the time to 18% of the time after viewing *Queen of Katwe*, a 44% decrease in the probability of failing. This closes the gap with their male peers, who fail the maths exam 22% of the time. When looking at prior ability as measured by a mock exam taken in the summer before the study began, the benefit from seeing the treatment movie is experienced by low ability students. Students whose scores in the mock exam were below the median increase their maths scores by 0.23 standard deviations when exposed to the treatment movie and decrease their probability of failing maths by more than 50%, from 54% to 21%. Students who scored above the median in their mock exam experience no effects on their maths scores or probability of failing maths from seeing the treatment movie. These effects are largest for female students from the bottom of the ability distribution. These results suggest role models could be a particularly effective intervention for low-ability female students in maths.

Amongst students taking their finishing exams from upper secondary school, I find an overall improvement in their performance of 0.13 standard deviations in their chosen subject papers. Again, it is women who benefit from seeing the treatment movie and see the largest improvement in their overall exam scores. The largest benefits are seen amongst women in the compulsory maths and science paper who are not in the science track[1]. Female students in the placebo group scored 0.28 standard deviations worse in the compulsory science paper than male students in the placebo group, but treatment with the role model more than counteracts this gender gap. These results are similar to those seen in the S4 class; that it is female students who are struggling in a maths-focused paper who are most helped by exposure to a role model.

I examine long term outcomes by looking at the decision to continue in education. For the lower secondary school class, I look at whether they go onto and finish upper

[1]Students already selected into a science or humanities subject track before treatment

secondary school. For the upper secondary class I look at whether they apply and are admitted to university. I find that lower secondary school students who saw *Queen of Katwe* are more likely to continue in schooling at upper levels. This is particularly true for female students, where treatment closes the 10 percentage points gender gap in continuing in school. Two years later, treated students do as well in the final leaving exam as placebo students. Students in upper secondary school are also 6 percentage points more likely to apply to public university. Again, the effects are driven by female students, who are 15 percentage points more likely apply to university, entirely closing the gender gap in applications with boys. These students go on to be admitted to university at higher rates, showing application is a key constraint for whether or not a student attends university. The treatment therefore seems particularly effective at encouraging girls to remain in education at the same rates as boys, closing the gender gap. Both of these results also suggest the treatment could have longer term beneficial effects on human capital acquisition beyond performance in a single exam.

I also consider potential mechanisms that the role model could work through, focusing on those highlighted in previous literature. I discuss whether the role model could work through raising aspirations, providing information about returns to education or overcoming stereotypes for female students about their ability in maths subjects. However, given I am limited to administrative records, and do not have survey data that could allow me explore potential channels through which the role model worked, I am not able to rule out one mechanism over another. Future work could look to explore the mechanisms through which role models work further.

This paper contributes to the literature on the impact of role models on educational attainment, but shows that the role model does not have to be available in person to have an impact. Beaman et al. (2012) finds that random exposure to female role models on village councils in India closed the gender gap in aspirations, particularly for education and occupation-related aspirations. In Madagascar, Nguyen (2008) used a randomised experiment to compare information about returns to schooling to a role model. She finds large impacts on test scores from being exposed to a role model but only if the

role model is from a similar poor background to the student, with the largest effects for poor students. In developed countries, role models have been shown to affect maths performance and high school and college decisions to study mathematical subjects (Marx and Roman, 2002; Breda et al., 2018; Porter and Serra, 2018). My study complements these by showing that the role model does not have to be available in real life to have a positive effect on students educational choices and performance, significantly raising the scalability of this type of intervention.

This paper shows that impacts on both short and long term educational outcome are possible after a brief (2 hour) exposure to a role model. While there is evidence from both developing and developed countries that long-term exposure to role models in TV soap operas can change behaviours and shift norms, particularly around fertility norms, La Ferrara et al. (2012); Chong and Ferrara (2009); Jensen and Oster (2009); Kearney and Levine (2015), there is relatively little evidence on the impacts of short, one-off exposures, particularly over the longer term. My study complements work which has looked at the impact of brief media exposure to role models and found large behavioural changes. Bernard et al. (2014), in Ethiopia, invited people to watch short films about how people from similar backgrounds to them had improved their socio-economic position. Six months later, the treated group had higher aspirations and displayed forward looking behavioural changes. My study takes this type of one-off intervention into a new setting, student educational attainment, and shows there are likewise strong educational effects from brief role model exposure in both the short and long term.

This paper also contributes to a growing literature on the impact of media on economic behaviours (La Ferrara, 2016). The intervention used in this paper is closest to "edutainment" interventions, where information is presented through a narrative with an explicit policy change goal. Banerjee et al. (2018) use randomised screenings of a TV show called *Shuga*, with the aim of reducing risky sexual behaviour. They find striking changes in behaviour and knowledge about HIV, and present evidence that the entertainment component of the intervention was a key reason for its impact. Likewise, Paluck and Green (2009) randomized exposure to a soap opera in Rwanda and find effects on

behaviours and social norms. This paper is unique, however, in using as the intervention a narrative with the aim of purely being entertaining to bring about a behavioural change, rather than one explicitly designed for a social purpose.

In terms of policies to improve performance in school in developing countries, this intervention was extremely costs effective, costing only $5 per student for a cinema screening and transport and so could easily be scaled up through screenings in schools. My findings demonstrate that a low cost, one-off and brief exposure to a role model can have as powerful effects on education outcomes as larger and more complex programmes, such as teacher incentives or instructional materials (Evans and Popova, 2015). It also shows that more costly media-based interventions designed to specifically affect certain behaviours are not needed to achieve the desired effect, suggesting wide potential to repackage existing materials for new aims.

The rest of this paper is organised as follows: Section 2 discusses the study design and data. Section 3 contains the results. Section 4 discusses potential mechanisms. Section 5 discussions the cost effectiveness of the findings and section 6 concludes.

2 Experiment Design and Data

2.1 Intervention

The study involved randomised exposure to either a treatment or a placebo intervention:
The treatment intervention involved a cinema screening of *Queen of Katwe*, the inspirational story of a young girl, Phiona Mutesi, from the slums of Kampala's rise out of poverty to become a world chess champion. The film is based on a true story.
The placebo intervention involved a cinema screening of *Miss Peregrine's Home for Peculiar Children*, a fantasy story about children with paranormal abilities.

2.1.1 Treatment movie

The (true) story of Phiona Mutesi is an inspiring rags-to-riches tale; Phiona goes from nothing, living in the slums and selling corn to passing drivers, to getting into the top school in Kampala, playing international-level chess and achieving her dreams. The real life Phiona becomes one of the first two women in Uganda to become a titled chess player. The movie version of Phiona's story, *Queen of Katwe*, was produced by Disney and ESPN and directed by Mira Nair. It received widespread acclaim from critics[2], being both nominated for and winning multiple awards.

The movie begins with a quote form Ellen Johnson Sirleaf "The size of your dreams must always exceed your current capacity to achieve them". This idea of having bold goals and fighting to achieve them in the face of obstacles is the central theme of the story. The movie sets up the story by showing the poverty and daily struggles of Phiona and her family to survive. Phiona only discovers chess after she approached a children's chess club because they were offering free food. When Phiona believes she does not belong at the club, after the other children make fun of her smell and tatty clothes, the club's coach tells her "Sometimes the place you're used to is not the place you belong. You belong where you believe you belong." Phiona returns to the chess club the next day.

One of the first things Phiona learns about chess is the idea that you can become

[2]The movie was scored 73/100 by metacritic and 7.4/10 by IMBD in their aggregates of critics scores. The New York times scored the movie 90/100

bigger than you are "In chess, the small one can become the big one", meaning that even the lowest piece, a pawn, can become the most powerful, a Queen. The story then charts Phiona's own metaphorical rise from pawn to Queen.

The film uses chess as a metaphor for life: it doesn't matter how strong, intelligent or wealthy you are, you can learn to strategize your way to a better life. As their coach tells his class "Use your minds. Make a plan. There you will find safety." The concepts of sacrifice and winning and losing are repeated frequently throughout the film. Failing is shown as a key part of life, with their chess coach telling his class "Losses happen to everyone. But then you reset the pieces and play again". However the students are strongly encouraged to never give up, being told "Do not be quick to tip your king. You must never surrender." and "This is a place for fighters."

A key narrative of the film is the fact that Phiona can only play in chess tournaments if she can get into a top school with a chess programme. Phiona never learnt to read and write, so she first has to learn in order to go to school, spending hours a day studying with her mentor, and taking time away from playing chess. Central messages of the story are therefore that education can be a means of achieving other goals and that intelligence is not fixed but can be gained by learning (Dweck, 2000).

For someone to be a role model, they must firstly be similar enough to the audience for them to relate to, and secondly achieve something extraordinary (Chung, 2000). Phiona, has many characteristics which have been shown in a large psychology literature to make her a meaningful role model with whom secondary school students in Kampala could identify: Phiona is similar in multiple dimensions to many of the students in my sample, such as age, geographical, academic and socio-economic background, and hence relevant to them and easy for them to relate to (Lockwood and Kunda, 1997).

Phiona achieves something exceptional. She begins from a position of extremely low educational attainment. She has dropped out of school and is only qualified to sell vegetables on the side of the road. She rises from this low position to get into a top private school in Uganda so that she can pursue her love of chess. Phiona might therefore be a particularly relevant role model for student at the bottom of the ability distribution and

those from a poor background.

Phiona is also a counter-stereotype in that she is a woman doing well at what is traditionally a male dominated game, chess (Dasgupta and Asgari, 2004). Exposure to a counter-stereotype has been shown to change attitudes and "inoculate" those exposed against applying stereotypes to themselves (Stout et al., 2011). Phiona might most appeal as a role model to female students in subjects they experience negative stereotypes about, the STEM subjects (science, technology, engineering and maths) and for those students who the stereotype is most close to the truth to, those already struggling.

The form of exposure in an entertaining movie could also magnify any impact of Phiona as a role model. A movie allows the narrative of the role model to be presented in an engaging and immersive way, causing the viewer to experience vicarious cognitive and emotional responses to the story as it unfolds (Green and Brock, 2000). Involvement with the characters and the storyline to allow the individual to feel 'transported' into the plot have been shown to be key determinants of the persuasive effects of edutainment programmes (Moyer-Gusé, 2008; La Ferrara, 2016). Video-based media has been shown to be particularly effective at tailoring information to individuals in a way that individuals relate to (Bernard et al., 2015). Exposure to a successful role model through a movie also allows a wider group of people to be exposed to a role model who may lack one in their immediate environment.

2.1.2 Placebo movie

Going to the cinema is an affluent activity in Uganda, reserved for the middle classes for a special occasion. Most of the students in the study would have never been to the cinema before, or been very few times. The placebo movie was therefore important to remove any potential effects simply from going to the cinema. For example, the very act of going to the cinema may have made students want to do well academically so they could get good jobs and afford to go to the cinema!

The placebo movie was chosen carefully to be appealing to this age group. The movie was entertaining and suitable for the students, containing characters of a similar

age but without a Ugandan background. Note that due to the need to select a movie for the placebo arm that was also currently showing at the cinema, it was not possible to choose a movie more similar in context to the treatment movie. The content of *Miss Peregrine's Home for Peculiar Children* was purely an adventure story focused on overcoming monsters threatening the characters. There was no educational content.

2.2 Setting

Secondary school in Ugandan last 6 years and is split into lower and upper. Lower secondary school lasts 4 years and at the end of this period, in S4, the Uganda Certificate of Education (UCE) is sat. Advanced level exams, the Uganda Advanced Certificate of Education (UACE), are taken after a further 2 years of study by the S6 class.

Girls in Uganda increasingly drop out of schooling as they progress through the grades. While nationally 34% of boys continue from lower to upper secondary school, only 24% of girls do (MoES Uganda, 2015). This compares to 40% of boys and 36% of girls completing lower secondary school. Only 40% of those sitting final secondary school finishing exams at S6 are female. In Kampala, Only 12% of all girls go to university, compared to 15% of boys. Nationally, less than 10% of female students take mathematics and science whereas 30% of boys do (Uganda National Examinations Board, 2018).

Girls also consistently perform worse than boys, especially in maths and science. While 49% of male students in secondary school are judged proficient in mathematics, only 33% of female students are. In contrast, in English both sexes are judged equally competent, with 49% meeting proficiency thresholds. Of those who finish secondary school, 89% of male students receive the required grades to enter university, whereas only 85% of girls do. Of the 8,677 female candidates who sat Mathematics at UACE in 2017, only 50.9 per cent scored E or above (and so passed the exam), compared to 60% of boys.

2.3 Sample

The sample consists of secondary school students in Kampala who were preparing to sit their national exams in 2016 (the S4 and S6 classes). Secondary schools were approached

during September 2016 in the urban Kampala area. The outreach to schools was done by an NGO, the Initiative for Social and Economic Rights (ISER), that was connected to the study via the funder. ISER approached 22 schools whom they had previously worked with. There were no criteria for a school being recruited into the study except for being known to ISER, being within 1 hours drive of the cinema (in normal traffic) and consenting to provide student records and later exam data. 13 schools agreed to participate in the study.

The study was pitched to schools as looking at the impact of film on exam performance. Schools were not told that the study was looking at the *Queen of Katwe* movie in particular. Schools were given a list of 4 possible movies, including the treatment and placebo movie, so they could assess their suitability for their students to see, but not told which of them their students would be seeing. The students were unaware of which movies they would be seeing until they arrived at the cinema. The schools provided their entire cohort of S4 or S6 students[3], such that the only untreated students in the year group were students who were absent from school on the day of the screening.

2.4 Randomisation

The movie screenings began on the day that both *Queen of Katwe* and *Miss Peregrine's Home for Peculiar Children* were released in Uganda, Friday 7th October 2016. Two sessions, each screening both movies, were run per day, one at 11am and one at 2pm, for 5 days, finishing on Tuesday 11th October. Consenting schools were allocated to one of five consecutive screening days in the second week of October, and either a morning or afternoon session. This was based on their geographical proximity to each other, the number of students at the school and the capacity of the cinema screens. Schools with less than 100 students were combined into a screening session with another school nearby.

The students were collected by mini vans hired for the study, which arrived at the cinema 1 hour before the screening to allow time for the randomisation. Students were individually randomised into the treatment or placebo movie upon arrival at the cinema

[3]I confirm this using registration lists, only 10 students were absent on the day

for a screening. This was done by students lining up outside the cinema and one by one entering, upon which an assistant picked a ticket out of a bag without looking and handed it to the student. The bag was opaque and the tickets identical except for the name of the movie printed in small print at the bottom of the ticket. An assistant was chosen to actually pick the ticket to further reduce any probability that a student might try and pick a particular ticket.

Immediately after getting a ticket, students went to the designated registration desk for that movie, before proceeding into the theatre. Students with tickets for different movies were kept separate the entire time, even using different bathrooms. I am therefore confident that all students saw the movie they were assigned to.

The cinema had 3 screens which could be use for screening the movies, two screens of 100 person capacity and one screen of 300 person capacity. Due to the difference in the sizes of the cinema screens, students within individual schools did not have an equal probability of seeing the treatment and placebo movie. School fixed effects will be used to control for this difference in treatment probability within a school.

2.5 Data

Data was collected from the schools, exam board and ministry of higher education. This provided background characteristics of the students and outcomes on both short and longer term educational attainment. A timeline showing the key schooling choices, the intervention and data collection is shown in the Appendix in Figure A1.

2.5.1 Student and school data

Limited information about the students was collected upon registration at the cinema. This was their name, age, gender, class (S4 or S6), school, and unique index number along with which movie they saw. This data was entered into excel using double data entry with any discrepancies checked. This resulted in a data set of 1,500 students who saw a movie and were due to take a national exam. Information on the schools themselves was also collected and is shown in the Appendix in Table A4.

Mock exam results data was collected directly from the schools for all students. Students sat a mock exam during the summer before their national exam, in August 2016, two months before treatment took place and one month before schools were approached about taking part in the study. This mock exam was administered by the schools and based on previous exams. Students in the S4 class sit mock exams in maths and English. Students in the S6 class sit the mock exam in the subjects they are registered for in the national exam. Schools were requested to provide the complete subject-by-subject mock results. However, some schools only provided the aggregate score across all subjects. The mock exam results will be used here as a baseline test score.

I also obtained further characteristics on the majority of the students from their school admissions records[4]. This information covered the student's previous exam outcome (PLE or primary leaving exam for S1 entry, UCE for S5 entry) as well as parental occupation, address, whether both parents were alive and if the student had applied for and was granted a bursary. At S4 20% of students are from Kampala, 34% have lost at least one parent, 18% have parents with a professional occupation and 16% receive a financial bursary. At S6 level 25% of student receive a bursary, reflecting the higher cost of upper secondary school, otherwise the averages are very similar to S4. These are shown for both treatment groups in Appendix Tables A1 and A2.

2.5.2 Short term outcomes: Test score data

Data on national exam results was collected in February 2017 once the exam data of both the S4 and S6 classes had been released. Results were collected directly from schools in the form of printouts of all the student's results provided by the exam board. These were double entered into Excel. In the case of a few schools not wanting to provide us with the exam results of their students, an SMS exam results collection system was used.[5]

[4]Not all the schools allowed this. At S6 admissions records were provided by 6 schools with 564/711 students. At S4, admissions records were provided by 5 schools with 565/735 students

[5]All the schools agreed to provide exam results as part of the study. Some schools, particularly those with poor results, later changed their minds about providing us with copies of results. However they were all aware that since we had the index numbers of the students we could obtain the results directly from the exam board. Results were collected for all missing students exam results and entered into Excel. A random sample of results were audited to ensure they had been entered correctly.

The exams sat by the students had already been chosen and registered for well before the intervention occurred and so neither the subject choices nor the number of subjects could be changed as a result of the intervention. They are pre-determined with respect to treatment. More detail on the test score calculation is given in Appendix A2.

Standard 4 Exam

UCE exams began on the 19th October 2016, 1 week after the last movie screening. These test score outcomes therefore capture the immediate impact of treatment. The UCE comprises six mandatory subjects administered in English; these are mathematics, English language, biology, chemistry, physics, and a choice of either geography, or history. Two other optional subjects are also chosen from subjects such as music and business. Candidates must register for a minimum of 8 and a maximum of 10 subjects. I look at the following outcomes at UCE:

Exam score aggregate: aggregate score composed of exam scores across all subjects taken by a student

Core exam score: composed of exam score in the six mandatory subjects taken by all students

Individual subject grade: Standardised score achieved in maths and English subjects

Standard 6 Exam

UACE exams began on the 14th November 2016, 1 month after the intervention. The UACE is taken in five subjects, three are optional papers taken in principle subjects, and two are compulsory: one in a subsidiary subject out of mathematics or computer and one in a general paper. The following outcomes were examined at UACE:

Total exam score: aggregate exam score composed of exam scores across all principal subjects taken plus 1 point for each pass in the two subsidiary subjects.

Principal score: aggregate score in the 3 principal papers only.

Maths/computer paper score: standardised score on the general and subsidiary papers in maths or computer[6].

[6]Note that students taking maths already as a principle subject have to take computer

2.5.3 Long term outcomes: Continuation in Education

Continuation in school

Three long term outcomes were specified relating to choices by students originally in the S4 class to continue in education. These are 1) An indicator for whether the student entered upper secondary school, 2) An indicator for whether the student chose maths as a principal subject, 3) An indicator for whether the student sat the final leaving exam (UACE). Additionally, for students who sat the final UACE leaving exam, I examine their exam scores in the same way as specified for the original S6 class.

An alternative route after taking UCE exams for students is to go into vocation training. Vocational training in courses such as mechanic, plumber, tailor and hair and beauty carries minimum entry requirements in terms of UCE grades. The minimum requirement for an accredit vocational course is 3 passes at UCE, including maths and English. I construct a dummy variable equal to one if a student meets this minimum.[7]

Admittance to university

I examine whether students from the S6 class applied and are admitted to university[8]. I obtained the application information directly from school, though some schools were unable to provide this data[9]. I obtained the admittance information from the Ugandan National Council of Higher Education, which collects data on all students admitted to University in Uganda. Additionally I examine a dummy variable for whether a student achieves the grades to get into public university. Public University in Uganda requires passing grades in two principal subjects[10].

[7]Note this outcome is not pre-specified. I do not have direct information on whether students who didn't continue in school started a vocational training course.

[8]I also pre-specified to examine whether the student obtained a government scholarship but only 16 students in my sample obtained such a scholarship so I am unable to look at this outcome

[9]6/12 schools representing 559/708 students were able to provide data on whether each child in their school applied to University

[10]Note that this outcome was not pre-specified in the analysis plan as I was not aware of the common grade requirement for university entrance at this time.

2.6 Randomisation balance check and attrition

Tables 1 and 2 show balance tests by class for the individual and exam choice characteristics collected during the intervention and from the schools. No significant differences are found between the samples. Students in the S4 class were on average just over 17 years old, half of them were female and most were taking 10 subjects in the exams. At S6 level, students are now two years older, at 19 years old on average, half are female and one third are taking maths or science as an optional paper.

Table 1: Balance test S4 class

	Placebo			Treatment				
	n	mean	sd	n	mean	sd	difference	p-value
Age	342	17.28	1.25	391	17.25	1.23	0.03	(0.76)
Female	343	0.51	0.50	391	0.51	0.50	0.00	(0.61)
Number of subjects	344	9.73	0.62	391	9.68	0.60	0.04	(0.34)
Mock total score	342	0.01	0.98	389	-0.01	1.01	0.02	(0.74)
Mock maths score	313	0.00	1.00	341	-0.00	1.00	0.01	(0.92)
Mock English score	313	0.01	0.91	341	-0.01	1.09	0.02	(0.84)

Age refers to age in years, Number of subjects is the number of subjects the student had been entered for exams in. Mock total score is the standardised score achieved in the mock exam taken prior treatment. Mock maths and English scores are the standardised scores achieve in the mock exam taken prior to treatment

Table 2: Balance test S6 class

	Placebo			Treatment				
	n	mean	sd	n	mean	sd	diff	p-value
Age	341	19.09	1.24	370	19.00	1.13	0.09	(0.31)
Female	341	0.47	0.50	370	0.50	0.50	-0.03	(0.40)
STEM	341	0.33	0.47	370	0.30	0.46	0.02	(0.53)
Mock total score	340	-0.02	0.97	368	0.02	1.01	-0.06	(0.45)
Mock principal score	182	-0.02	1.02	234	0.02	1.09	-0.01	(0.57)
Mock maths/computer score	311	-0.01	1.06	339	0.01	0.98	-0.02	(0.82)

Age refers to age in years, STEM is a dummy if the student is taking maths, biology, chemistry or physics as one of their subject choices. Mock total score refers to the standardised test score in the mock exam taken prior to treatment. Mock principal score is the mock score in the 3 chosen principal papers. Mock maths/computer score is the mock score in the compulsory general and maths or computer papers.

Attrition occurred in the form of students not taking the national exam. Since I had the students' exam index numbers I could always obtain exam results if they existed. Attrition was balanced across the treatment and control groups, as shown in Table 3

Table 3: Attrition Balance Test

	Placebo		Treatment			
	mean	sd	mean	sd	difference	p-value
Attrition rate	0.03	0.17	0.04	0.20	-0.01	(0.22)
Observations	706		794		1500	

Differences in mean attrition between placebo and treatment.

below. 21 students in the placebo and 33 in the treatment group did not take their national exams, 3.6% of the sample. I examined whether student or school characteristics were correlated with attrition in Appendix Table A3. Students at Christian schools are more likely to take the exam, as are older students and students in the S4 class.

2.7 Empirical strategy

To examine the effect of treatment on exam outcomes, I estimate the following regression:

$$y_{is1} = \beta_0 + \beta_1 \text{QofK} + y_{is0} + \boldsymbol{x_i'} \cdot \gamma + \theta_s + \epsilon_{is}, \tag{1}$$

where i indexes student at school s, y_{is1} denotes the exam outcome of interest, QofK is an indicator variable equal to one for if the student saw the movie *Queen of Katwe*, $\boldsymbol{x_i'}$ is a vector of individual characteristics, θ_s is a vector of school fixed effects and ϵ_{is} is a random error. y_{is0} is the standardised mock exam result from before treatment.[11]

The parameter of interest is β_1, the average treatment effects of the *Queen of Katwe* movie on an outcome. The school fixed effects, θ_s, are included since the randomisation was done within schools, and hence stratified by school. Robust standard errors are calculated to allow for heteroskedasticity.

Individual characteristics, $\boldsymbol{x_i'}$, are included to improve precision. These are: a dummy for whether the student is female; the age of the student in years; the number of subjects taken (for S4 students); and whether the student choose to take any subjects in maths or science (STEM subjects) at S6 level

[11] If the equivalent mock result is not available for an outcome, the aggregate result constructed from the available mock papers will be controlled for instead and a dummy variable included to capture this.

3 Results

3.1 Short term results: Impact on test scores

3.1.1 S4 Class

Table 4 shows the impact of the treatment movie on the S4 exam outcomes defined in section 2.5.2. I show results both with and without individual control variables. Treatment assignment has no effect on the total score, core score or English standardised scores. However, treatment does result in an increase of 0.11 standard deviations in maths score, significant at the 5% level, and this does not change with the addition of controls. This is a large positive effect on the maths exam outcome.

Table 4: Impact of treatment assignment on S4 standardized test scores

Impact of treatment assignment on standardized test scores												
	(1) Total Score	(2) Total Score	(3) Core Score	(4) Core Score	(5) Maths	(6) Maths	(7) Fail maths	(8) Fail maths	(9) English	(10) English	(11) Fail English	(12) Fail English
Treatment	-0.02	-0.02	-0.02	-0.02	0.10**	0.11**	-0.11***	-0.11***	-0.01	-0.01	-0.02	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)	(0.06)	(0.06)	(0.02)	(0.02)
Age		-0.00		-0.01		-0.07***		0.02**		-0.07***		0.00
		(0.01)		(0.01)		(0.02)		(0.01)		(0.02)		(0.01)
Female		-0.04		-0.04		-0.01		0.04		-0.17***		0.04**
		(0.03)		(0.03)		(0.05)		(0.03)		(0.06)		(0.02)
No. subjects		-0.07**		-0.08**		-0.02		-0.02		0.08		-0.01
		(0.03)		(0.03)		(0.05)		(0.03)		(0.06)		(0.02)
Mock score	1.00***	1.01***	0.96***	0.96***	0.66***	0.66***	-0.13***	-0.12***	0.53***	0.52***	-0.08***	-0.08***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.04)	(0.03)	(0.01)	(0.01)
Constant	-0.28***	0.33	-0.48***	0.33	-0.56***	0.82	0.67***	0.42	-0.35**	0.22	0.11	0.07
	(0.10)	(0.33)	(0.11)	(0.36)	(0.08)	(0.59)	(0.10)	(0.30)	(0.14)	(0.64)	(0.09)	(0.23)
Control mean							0.27	0.27			0.11	0.11
Observations	729	729	729	729	729	729	729	729	729	729	729	729
R-squared	0.88	0.88	0.82	0.83	0.53	0.54	0.29	0.30	0.46	0.47	0.26	0.26

Total score refers to standardised aggregate score across all subjects taken in the exam. Core score refers to standardised aggregate score in the 6 mandatory subjects at S4 level. Standardized test scores composed of subject standardized scores and renormalised. All regressions include school fixed effects. Mock score refers to the standardised score achieved in the corresponding mock exam, unless missing in which case it is replaced by the overall mock exam score and a dummy variable including to capture this. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

To examine the effect of treatment on exam performance in more detail, I break down the maths and English exams into dummies by grade achieved. The exam is graded from 0, fail, to 8, the maximum result. I then look at the impact of treatment on the probability of failing maths and English. It can clearly be seen in Table 4 that seeing the treatment movie reduces the probability that a student obtains the bottom, failing, grade in maths by 11 percentage points. In the placebo group, 27% of students fail maths, so

this is a 40% decrease in the probability of failing maths. In the placebo group, only 11% of students failed English. I find no effect of treatment on the probability of failing English. The impact of treatment on a dummy for failing each of the core exam subjects can be seen in Appendix Table A26, along with multiple hypothesis tests in the form of sharpened q-values to adjust p-values to control for the false discovery rate.[12] The only subject for which the significant result is robust to multiple hypothesis testing is the maths result, which is still significant at the 1% level.

I also look at the effect of treatment on the probability of getting each maths grade using an ordered logit regression. Since the largest number of students getting a particular grade is students failing (27%), I have most power to detect an effect here. An ordered logit will allow me more power to see effects elsewhere in the grade distribution. The marginal effect of treatment on each grade is shown in Appendix Table A5. Here the coefficients shown give the probability of achieving each grade level as the treatment indicator goes from 0 to 1. Seeing the treatment movie reduces the probability of achieving the lowest 2 scores, particularly the probability of getting the failing score declines by 5.2 percentage points. The treatment movie also increases the probability of obtaining higher grades, with the effect significant at the 1% level for all scores above 2, though the magnitudes are small. I also do the same ordered logit for the other subjects in the core exams and find no effect of seeing the treatment movie on the probability of getting any particular grade (shown in Appendix Table A6).

To further understand where on the grade distribution the treatment effect is I plotted histograms by subject. Plots of the entire distribution of results for both treated and placebo students are shown in the Appendix in Figure A2. The histograms of total score, core score and English show no statistically significant impact of treatment in the distribution. To formally test this I perform a Kolmogorav test. For total score, core score and English the p-value on the test of equality of the distributions are 0.37, 0.35 and 0.64 respectively. Hence I cannot reject equality of the distributions. However, in

[12]The false discovery rate controls for the expected proportion of rejected hull hypotheses that are false (incorrectly rejected). It therefore controls for the rate of type I errors when testing many hypotheses. The method used here is Benjamini et al. (2006) sharpened q-values as used in Anderson (2008).

the histogram of maths score it can be seen that the histogram is shifted to the right, particularly at the lower end to just above the mean. The p-value for the Kolmogorav test is 0.006, so I can reject equality of the distributions at the 1% significance level and confirm that treated students achieve higher maths scores.

It is also possible that any improvement in maths scores and reduction in failing maths came at the expense of other subjects. This could be the case if effort was directed away from other subjects and towards maths rather than increased overall. While I see no significant negative on English or total or core scores, it is possible there are small decreases in other subjects across the board that can't be seen when aggregated with the maths improvement. I test this by looking at the impact of treatment on the aggregate score excluding maths. Results for this are shown in the Appendix in Table A7. I find no significant effect of treatment on the total exam score excluding maths.

3.1.2 S6 class

Table 5 shows the impact of assignment to see the treatment movie on the S6 exam outcomes defined in section 2.5.2. Results are shown both without and with individual control variables, but I will discuss only the results with control variables for brevity. I include the pre-specified control variables: age in years, a female dummy and a dummy for if at least 1 subject out of maths, biology, physics or chemistry were taken (STEM) and the baseline mock score for that outcome[13].

Seeing the treatment movie results in the overall exam score being 0.13 standard deviations higher, significant at the 1% level, a large effect on an education outcome. The score on the principal exam papers increases by 0.13 standard deviations, significant at the 5% level. The score achieved on the maths/computer papers is 0.13 standard deviations higher for treated students, significant at the 5% level.

I also look at the effect of treatment assignment on different parts of the results distribution. I do this by looking at the impact of treatment in ordered logit regres-

[13]One school only provided the overall mock results, not the individual subject paper results. In this case the overall standardised mock score is included in the regressions for the principal and maths/computer outcomes instead of those specific mock scores, with a dummy variable capturing this.

Table 5: Impact of treatment on S6 standardized test scores

	(1) Overall score	(2) Overall score	(3) Principal score	(4) Principal score	(5) Maths/ computer score	(6) maths/ computer score
Treatment	0.17**	0.13***	0.16**	0.13**	0.10	0.13**
	(0.07)	(0.05)	(0.08)	(0.05)	(0.06)	(0.05)
Age		-0.03		-0.02		-0.04
		(0.02)		(0.02)		(0.02)
Female		0.06		0.13**		0.03
		(0.05)		(0.05)		(0.06)
STEM		0.56***		0.22***		0.46***
		(0.06)		(0.06)		(0.07)
Mock score		0.76***		0.70***		0.39***
		(0.03)		(0.03)		(0.03)
Constant	1.26***	0.72*	1.15***	0.72*	1.43***	1.53***
	(0.09)	(0.41)	(0.10)	(0.42)	(0.09)	(0.47)
Observations	708	708	708	708	708	708
R-squared	0.19	0.62	0.15	0.59	0.28	0.49

Overall score refers to the aggregate score in the principal and maths/computer and general papers. Principal score refers to the standardised score on the 3 chosen subject papers. Maths/computer score subjects refers to the standardised score on the two mandatory maths or computer and general papers. Standardized test scores composed of subject standardized scores and renormalised. All regressions include school fixed effects. Age is age in years. Female is a dummy for being a female student. Stem is a dummy capturing students in the science track. Mock score is the score in the mock exam prior to the intervention for that outcome. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

sions of the raw scores and by examining histograms of adjusted scores by treatment assignment. Firstly, the impact on the raw points score by overall, principal paper and maths/computer paper, shown in Appendix Table A8. The overall score is out of a maximum of 20, the principal out of a maximum of 18 and maths/computer out of a maximum of 16[14] so these regressions shows how treatment affects the likelihood of getting any particular score. As seen in column (1), treatment results in a student's overall score being significantly less likely to be at the lower end of 2-7 out of 20 and significantly more likely to be at the higher end of 9-15 out of 20. However, this is only marginally significant at the 10% level, and of a small magnitude, suggesting improvement in the score from treatment are throughout the score distribution. A similar pattern in seen for

[14]remembering that the maths/computer papers enter the overall score as 1 point for a pass in each of the two papers, where one paper is a general knowledge paper and the other is either maths or computer

the principal paper in column (2), though nothing is significant. For the maths/computer paper a stronger pattern is seen; Treated students are less likely to fail and get scores below 5, and more likely to get scores of 6-13. Overall these results show a general shift as a result of treatment from the lower half of the grade distribution to the upper half.

I also plot histograms by these three outcomes to see where the treatment effect is shifting the distribution. Plots of the entire distribution of results for both treated and placebo students are shown in the Appendiex Figure A3. The histograms all show a shift to the right for students assigned to treatment throughout the distribution, particularly around the middle of the distribution. Kolmogorov test for equality of the distributions reject equality for all three outcomes (p-values of 0.07, 0.073 and 0.012 for the overall score, principal score and maths/computer score respectively). Overall, these results suggest that for the S6 class, students see improvements throughout the distribution, rather than being concentrated in the bottom end, as in the S4 class.

3.2 Heterogeneity

In the following, I examine heterogeneous effects by background characteristics of the student that make them more or less similar to the role model[15], as well as capturing if they came from a more deprived background, prior academic ability and gender. At S6 level I also examine heterogeneity by whether the student was on the STEM (sciences) track.

3.2.1 S4 class heterogeneous effects

Heterogeneous treatment effects by gender and prior ability on maths exam performance for the S4 class are shown in Table 6. I only show here heterogeneity by maths score and failing maths as there are no heterogeneous effects for the total score, core score or English score (see these in Appendix Table A9).

[15]These are whether the student comes from Kampala (where the role model comes from), whether the parents had a professional occupation (a mark of higher socio-economics status), whether the student was raised only by their mother (a signal of socio economic status, since widows are poorer, but importantly the role model was raised only by her mother) and whether the student receives a bursary to attend school (also measuring financial status of the household).

Looking first at gender in columns (1) and (2), though the point estimate for the interaction of female and treatment for maths score is positive and relatively large, it is not significant either using conventional p values or sharpened q values, and so I do not find a significant difference between girls and boys. However the overall effect of treatment for women is a positive and significant 0.18 standard deviations for maths score, and the treatment effect for maths score for boys is no longer significant (insignificant +0.03). This suggests that girls benefit more from seeing the treatment movie than boys, though I can't reject that the effects are equal.

Looking at students who scored below the median in their mock exam, there is a large and significant heterogeneous effect of treatment. Those whose mock scores were below the median see a 0.23 standard deviation increase in their maths exam score compared to those who scored above the median in the mock. Likewise, they see an additional 0.33 percentage point decrease in the probability of failing maths. These findings are also robust to multiple hypothesis testing. The coefficients on treatment for those above the median in the mock exam are very small and insignificant; these students are not seeing improvement in their scores or a decrease in failing maths from treatment. Overall the probability of failing maths falls by 28 percentage points. Compared to a mean of 54% of students with below median mock scores failing maths, this means treatment reduced by 50% the probability that a poorly performing student failed maths. These are very large effects, and suggest that those students struggling are more able to improve their maths scores from treatment.[16]

I also look at the interaction of being female and scoring below the median in the mock in columns (5) and (6). The largest benefits of treatment are seen by female students at the bottom of the ability distribution, who experience an additional 0.23 standard deviation increase in their maths score, and are 11 percentage points less likely to fail maths, almost entirely compensating for the increased likelihood that both female

[16]I breakdown the treatment effect by mock exam performance further by interacting the treatment with each decile of mock score. These are shown in Appendix Table A10. The results also suggest it is those in the bottom of the distribution who are benefiting from the treatment. Note however that this pattern of results could also be explained by it being easier to improve scores at the bottom of the distribution in a short amount of time.

students and students below the median in the mock fail the maths exam. Note however that while significant at the 10% level, these findings are not robust to multiple hypothesis testing.

I also examine heterogeneity by student background characteristics in Appendix Table A13. Again I see that effects are largest for students who scored below the median in the primary school leaving exam, PLE, matching those by mock exam performance. I see no effects by whether the student was from Kampala or raised by only their mother. I see that treatment effects are larger for those whose parents do not have professional occupations or who receive a financial bursary. These results suggest that being financial less well off is an important source of heterogeneity. Likewise I also see effects concentrated in students from disadvantaged schools[17] in Appendix Table A22. However, just being similar to the role model (being from Kampala and raised only by their mother) is not a source of heterogeneity. Note however that these sources of heterogeneity by background characteristics do not survive a multiple test correction.

3.2.2 S6 class heterogeneous effects

For the S6 class, heterogeneous effect by gender are shown in Table 7. The coefficient on the interaction of being female with treatment is always large, but insignificant for the overall score and principal subjects score. The interaction is significant for the maths/computer score, with female treated students scoring 0.26 standard deviations higher in these papers. There is no effect of treatment for male students for the maths/computer score (the coefficient on treatment is 0.01), meaning only female students see an improvement in their scores on this paper from treatment. This suggests that the beneficial effects from treatment are going to female students, and in the compulsory, mathematical papers. In the control group, male students score on average 0.13sd whereas female students only score -0.15sd in the maths paper . Hence treatment closes the gender gap in the maths/computer paper score.

I further break down these effects by looking at the interaction of gender with whether

[17]Note this outcome was not pre-specified and should be considered exploratory only

Table 6: Heterogeneity in treatment effect for S4 by gender and prior ability

	(1) maths	(2) fail maths	(3) maths	(4) fail maths	(5) maths	(6) fail maths
(A) Treatment	0.03	-0.08**	0.01	0.05	0.02	0.03
	(0.07)	(0.04)	(0.07)	(0.03)	(0.10)	(0.02)
	[0.98]	[0.10]*	[0.99]	[0.28]	[0.99]	[0.99]
(B) Treat * female	0.15	-0.07			-0.02	0.03
	(0.10)	(0.05)			(0.13)	(0.05)
	[0.42]	[0.29]			[0.99]	[0.99]
Female	-0.09	0.07*	0.03	0.02	-0.02	0.04
	(0.07)	(0.04)	(0.05)	(0.03)	(0.06)	(0.04)
(C) Treat * below mock			0.23**	-0.33***	0.11	-0.27***
			(0.10)	(0.05)	(0.12)	(0.06)
			[0.05]**	[0.00]***	[0.99]	[0.00]***
Below median mock			-0.81***	0.33***	-0.80***	0.33***
			(0.08)	(0.04)	(0.09)	(0.05)
(D) Treat* female* below mock					0.23*	-0.11*
					(0.14)	(0.06)
					[0.64]	[0.38]
Mock score	0.66***	-0.12***	0.51***	-0.09***	0.51***	-0.09***
	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)
Constant	1.54**	-0.05	1.77***	-0.25	1.69***	-0.21
	(0.67)	(0.33)	(0.62)	(0.32)	(0.61)	(0.33)
Overall treatment effect						
Female (A)+(B)	0.18**	-0.14***				
	(0.07)	(0.04)				
Below Mock (A)+(C)			0.24***	-0.28***		
			(0.07)	(0.04)		
Female above mock (A)+(B)					-0.00	0.06
					(0.10)	(0.04)
Female below mock (A)+(B)+(C)+(D)					0.33***	-0.32***
					(0.07)	(0.05)
Male above mock (A)					0.02	0.03
					(0.10)	(0.02)
Male below mock (A)+(C)					0.13	-0.24***
					(0.08)	(0.06)
Mean female	-0.11	0.32				
Mean male	0.12	0.22				
Mean low ability			-0.70	0.54		
Mean high ability			0.62	0.03		
Observations	729	729	729	729	729	729
R-squared	0.54	0.30	0.61	0.36	0.61	0.36

Maths is a standardized maths score. Fail maths is a dummy for whether a student failed the maths exam. All regressions include school fixed effects, student age and the number of subjects that student is taking. Below median mock is a dummy variable for if the student performed below the median in a mock exam. Mock score is the standardised score achieved in the maths mock exam taken prior to treatment. The bottom panel shows the overall treatment effect for each sub-group. The mean shows the control mean for that sub-group. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sharpened q values in square brackets.

the student was in the science track or not. While the role model may appeal more to those taking science already, she may have a larger effect on those not taking maths or

sciences in the compulsory sciences papers, by having a larger effect on those who struggle with maths. I examine this by looking at the interaction of gender and science stream on the exam paper outcomes[18].

While I see no significant heterogeneity by gender and science track interactions for the overall score and principal paper score, I see large and significant heterogeneity for women who are in the non-STEM track on the maths/computer paper (column (6)). Female students not in the STEM track see a significant 0.44 standard deviation increase in their maths/computer paper score. This is robust to multiple hypothesis testing.

At the bottom of Table 7, I calculate the overall treatment effect for different groups, calculated by adding together the treatment effects that apply to each group. Overall, female STEM, male STEM and Male non-stem students have no significant treatment effect from seeing *Queen of Katwe*, whereas female non-STEM students see an overall significant increase in their compulsory maths/computer paper score of 0.34 standard deviations. Using F-tests, I can reject that the overall effect for treated female non-STEM students is equal to that of female STEM students and male non-STEM at the 10% level (results not shown), but I cannot reject that the effect of treatment for female non-STEM students is equal to that of male STEM students. I cannot reject that female STEM, male STEM and male non-STEM are equal to each other, and to zero.

This impact of treatment for non-STEM female students exceeds the 0.18 standard deviation less that female non-STEM placebo treatment students achieve on their maths/computer paper, resulting in treated non-STEM female students achieving the same average score as male non-STEM students. Treatment therefore corrects the lower scores that female non-STEM students achieve on the maths/computer paper, taking them to the same level as their male non-STEM peers[19]. Treatment with a role model therefore has the potential to bringing female student's grades back in line with their male peers.

I also examine heterogeneity by background characteristics, shown in Appendix Table

[18]See table A11 for heterogeneity by STEM for all outcomes. I see no effects of being on a STEM track for any outcomes

[19]Those taking STEM subjects always perform better on the maths/computer paper, highlighting the science focus of this paper

A14. I do not see heterogeneity by whether the student scored below the median on the prior UCE exam, or by whether the student was from Kampala. I see heterogeneity for the compulsory maths/computer paper, with larger effects by whether the student was raised by only their mother, not with parents with professional occupations and in receipt of a bursary. However, only being raised by mother passes a multiple test correction. Again, this pattern is consistent with treatment affecting the compulsory maths paper with larger effects for those from disadvantaged backgrounds[20].

I also look at heterogeneous effects by prior ability as measured in the mock score in Appendix Table A11. The only outcome area that shows heterogeneity is the maths/computer paper score, with students who performed below the median in the maths/computer mock paper seeing a larger impact from treatment, though this is only significant at the 10% level and doesn't survive multiple test correction. The compulsory maths/computer paper most similar to the maths paper at S4, and so to see heterogeneity here by prior ability is aligned with the results for the S4 class.

I breakdown the treatment effect by mock exam performance further by interacting the treatment with each decile of mock score. These are shown in Appendix Table A12 for the total score (there are no differences for principal score) and maths/computer score. I find no differential effects of treatment by mock decile at S6 for either outcome, though for the maths/computer paper the treatment interactions are large for deciles 1-3 and small or negative for deciles 4 and above, suggesting there might be larger effects here for lower ability students.

I do not find any evidence that effects differed depending on the length of time between seeing the treatment movie and the student's first exam (see Appendix table A21).

3.2.3 Differences between S4 and S6 results

At S4 level effects are concentrated in those in the lower half of the prior ability distribution. At S6, effects are seen more across the ability distribution, though there is some evidence that in the maths/computer paper those in the lower half benefited more from

[20]I also look at heterogeneity by school characteristics in Table A23 and see no effects

Table 7: Heterogeneity in treatment effects for S6 by gender and science track

	(1) overall	(2) overall	(3) principal	(4) principal	(5) maths/ computer	(6) maths/ computer
(A) treatment	0.07	0.07	0.07	0.12	0.01	-0.10
	(0.07)	(0.08)	(0.07)	(0.09)	(0.07)	(0.09)
	[0.44]	[0.99]	[0.58]	[0.99]	[0.99]	[0.58]
(B) treat*female	0.14	0.14	0.12	0.07	0.26**	0.44***
	(0.10)	(0.12)	(0.10)	(0.12)	(0.11)	(0.13)
	[0.44]	[0.99]	[0.46]	[0.99]	[0.04]**	[0.01]***
(C) treat*stem		0.01		-0.12		0.24*
		(0.13)		(0.14)		(0.14)
		[0.99]		[0.99]		[0.27]
(D) treat* female*stem		-0.01		0.11		-0.63**
		(0.21)		(0.22)		(0.25)
		[0.99]		[0.99]		[0.04]**
female* stem		-0.12		-0.26		0.27
		(0.15)		(0.16)		(0.19)
female	-0.02	0.02	0.06	0.15*	-0.10	-0.18*
	(0.07)	(0.09)	(0.07)	(0.09)	(0.08)	(0.09)
stem	0.61***	0.60***	0.34***	0.36***	0.48***	0.35***
	(0.08)	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)
mock score	0.76***	0.76***	0.71***	0.70***	0.39***	0.39***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	0.71*	0.71*	0.70*	-0.34	1.55***	1.55***
	(0.41)	(0.41)	(0.42)	(0.46)	(0.47)	(0.47)
Overall treatment effect						
female:	0.20***		0.19**		0.27***	
(A)+(B)	(0.07)		(0.08)		(0.08)	
female stem:		0.19		0.18		-0.05
(A)+(B)+(C)+(D)		(0.14)		(0.15)		(0.18)
female non-stem:		0.20**		0.19**		0.34***
(A)+(B)		(0.08)		(0.08)		(0.09)
male stem:		0.07		0.00		0.14
(A)+(C)		(0.10)		(0.11)		(0.11)
male non-stem:		0.07		0.12		-0.10
(A)		(0.08)		(0.09)		(0.09)
Control mean female	-0.03		0.02		-0.15	
Control mean male	0.03		-0.02		0.13	
Observations	708	708	708	708	708	708
R-squared	0.62	0.62	0.59	0.59	0.49	0.50

Standarised scores. All regressions include school fixed effects. Mock score is the standardised mock score for that outcome. STEM refers to choosing maths, biology, chemistry or physics as a principal subject. The bottom panel shows the overall effect of treatment from adding the treatment and interaction effects together. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sharpened q-value in square brackets.

treatment.

One reason for the effects across the ability distribution at S6 could be the longer time span between treatment and the exams. There were different time lags between seeing the treatment and the exam for S4 (one week) and S6 (one month). Having only one week means that there is only time to make a limited amount of improvement in the S4 class. It is easier to improve an exam score from fail to one above fail than from a B to an A, so it will be the students who would have failed if they hadn't seen the treatment movie whose increased effort would most easily be seen in an improvement in exam grade. For them, a very small amount of extra effort could translate into a higher grade. Hence the time constraint might have meant I could only detect effects at the bottom of the distribution. At S6, they had at least a full month between seeing the treatment movie and the exams, over which time frame longer term motivation and perseverance comes into play, allowing more general improvements.

Additionally, the class profiles are very different between S4 and S6. All schools have minimum requirements to go from S4 to S6. Lower ability students who were helped the most at S4 in terms of failing maths are significantly less likely to continue in school; Only 5% of those in decile 1 in the mock in the control group continue to S5, whereas 50% of those in decile 10 do. Any student who failed maths at S4 would not be able to continue to S6. Hence precisely the students I see the largest impact for at S4 would not usually be present in the S6 class. Additionally, the principal subjects at S6 are chosen by students, with only the maths/computer papers compulsory. At S4 maths and all the core paper are compulsory. Hence selection by both student ability and into the student's strongest subjects may explain the differences.

Importantly, I see a number of commonalities across both grades, most notably that effects are concentrated in compulsory maths/science papers and for girls.

3.3 Long term outcomes: Continuation in education

3.3.1 Continuation to upper secondary school

After finishing S4, those who wish to continue academically can go onto S5 in upper secondary school, and hopefully continue to take the leaving exam at the end of S6 and go to University. Looking at these outcomes gives an indication of if student's ambitions for continuing in schooling are changing as a result of treatment.

Results for the impact of treatment on whether students from S4 sat the final leaving exam at the end of S6 are presented in Table 8. I find that treated students are 6 percentage points more likely to take the final leaving exam at the end of S6, significant at the 10% level. Since 27% of the control group continue to secondary school, this is a 22% increase in the probability of continuing to upper secondary school[21]

I look at heterogeneity by gender and whether a student scored below the median in the mock exam taken prior to treatment. I see that the treatment interaction with gender is positive and relatively large. There is no heterogeneity by prior mock exam performance. Female students who saw *Queen of Katwe* are 12 percentage points more likely to take the final S6 leaving exam, on a mean of 21% of the female sample in the control group taking the final leaving exam, an increase of over 50%. In fact, 32% of male students in the placebo (and treatment) group took the leaving exam at the end of upper secondary school[22], so seeing *Queen of Katwe* brings a female student from being 12 percentage points less likely to take the S6 exam to being as likely as her male peers. This suggests that seeing an inspirational movie featuring a positive female role model can correct the decline in female school completion at higher school levels.

I check whether students induced to stay in school by treatment then struggled to perform as well in the final leaving exams. This would suggest that students were motivated to stay in school who may have been better off doing vocational training instead. I look at the score on the final leaving exam at S6 in Appendix Table A16. I find no

[21]Results for continuing to S5 are shown in Appendix Table A15. The results as very similar due to the low drop out rate during S5 and S6, with only 2% of the control group dropping out, something seen nationally (MoE, 2015).

[22]This pattern is seen nationally, with 33% of male students continuing to S5 from S4 and 32% continuing to S6. Only 25% of female students continue from s4 to s5 and 22% to S6 (MoE, 2015)

evidence that students treated in S4 perform any differently on the final leaving exam at S6 than untreated students, nor that treated female students in particularly do any differently. If anything, the coefficients on the treated variable and treatment interacted with being a female student are positive, suggesting these students do slightly better. However, I have very low power to detect any effects on the final exam scores at S6 for the class than was in S4 at the time of treatment, as only 221 of the 730 original students take the final leaving exam. I also look at whether treated students who stayed in school were any more likely to take maths as a subject at S6 then untreated students. Results are shown in Appendix Table A17. I do not see any effects of treatment on the likelihood of taking maths, though the interaction of female with treatment is positive and large.

Finally, I look at whether treated students induced to stay in school were less likely to apply to University. These results are shown in Appendix Table A18. I find no impact of treatment on whether students apply to university, though the coefficient for treated girls is negative and, with only 185 observations, I have very little power to detect impact here. Overall these results suggest that treatment induced female students to stay in school longer and they went on to be just as successful academically as their peers.

3.3.2 Admittance into a vocational programme

For students who do not continue onto upper secondary school after finishing S4, the main route they follow is to undertake vocational training in programmes such as mechanic, hair dressing and beauty, electrician, tailor etc. These vocational courses also have minimum entry requirements of three passes at UCE, including English and Maths[23].

I show results for whether seeing the treatment movie impacted the probability of meeting minimum standards for vocational training in Appendix Table A19. I see that students who saw *Queen of Katwe* are between 10 and 11 percentage points more likely to get the minimum of 3 passes, including maths and English, needed to continue onto a vocational course. This means that more students who see *Queen of Katwe* will be able

[23]The ministry of education and sports runs a portal listing all accredited vocational courses http://www.btvet-uganda.org/

Table 8: Impact of treatment assignment on continuation in upper secondary school

	(1)	(2)	(3)	(4)
Outcome variable: Stays in school until the end of S6				
Treatment	0.06*	0.05	-0.01	0.06
	(0.03)	(0.03)	(0.05)	(0.05)
Treat* female			0.12*	
			(0.07)	
Female		-0.03	-0.11**	-0.04
		(0.04)	(0.05)	(0.03)
Treat* below mock				-0.03
				(0.07)
Below median mock				0.11
				(0.07)
Age		-0.04**	-0.04**	-0.04**
		(0.01)	(0.01)	(0.01)
No. subjects		-0.06*	-0.06*	-0.07*
		(0.04)	(0.04)	(0.04)
Mock score		0.07***	0.07***	0.10***
		(0.02)	(0.02)	(0.03)
Constant	0.33***	1.42***	1.42***	1.38***
	(0.07)	(0.39)	(0.39)	(0.40)
control mean	0.27	0.27		
control mean female			0.21	
control mean low ability				0.22
Observations	729	729	729	729
R-squared	0.03	0.06	0.06	0.06
Overall treatment effect				
female			0.11**	
			(0.05)	
below median mock				0.04
				(0.05)

No. subject is the number of subjects taken at UCE. Age is the age in years. Female is a dummy variable for female student. Below median mock is a dummy variable if the student scored below the median mark on the mock exam. Mock score is the standardised aggregate mock score. The control mean shows the mean for the control group, in a particular subgroup for heterogeneity. The bottom panel shows the overall treatment effect for each sub-group. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

to continue in education[24], potentially raising their lifetime earnings.

[24]I do not have data on whether students actually went on to do a vocational course, but without meeting these minimums they would not be able to start an accredited course.

3.3.3 Admittance to University

At S6 level, students can apply to University for the next academic year immediately after getting their UACE results. I therefore look at both whether students applied for university and whether they actually took up a place at university, which requires both being offered a place and accepting it.

In Table 9, I report whether the student was admitted to university and whether they applied to University[25]. I see that treated students are 6 percentage points more likely be admitted to University, significant at the 10 percent level, with a similar sized but insignificant coefficient for applied to university.

There is strong heterogeneity for girls in admittance to university, with treated female students being 13 percentage points more likely to be admitted to university. Female placebo students are 13 percentage points less likely to go to university than male placebo students, so treatment entirely removes the gender gap in admittance to university. There is also heterogeneity for applying to University, with female treated students being 15 percentage points more likely to apply to university, again closing the gender gap with their male peers. Note however that these coefficients are very large compared to the means (representing a 50% increase), and only significant at the 10% level, and so so should only be taken as suggestive that women are more likely to continue to university.

3.4 Robustness

I use permutation tests to compute exact test statistics which do not depend on asymptotic theorems. I do this for the primary S4 and S6 outcomes using 10,000 permutations and without individual control variables, only school fixed effects. These are reported in Appendix Tables A24 and A25 underneath the robust p-values. At the S4 level, treat-

[25]Note that whether a student applied to University was obtained from the schools themselves, whereas admittance to university was obtained from the Uganda National Council for Higher Education. Only 6 of the 12 schools provided statistics on applications to University. I also examine if students were more likely to get the minimum grades necessary to attend public university in Appendix Table A20, finding treatment increases by 4 percentage points the probability a student achieves the minimum grades. Note this outcome was not pre-specified. I did pre-specify whether seeing the treatment movie increase the odds that a student obtained a government scholarship. However only 16 students obtained scholarships from my study sample, and so I do not include the results.

Table 9: Impact of treatment on University admittance

	(1) Admitted to University	(2) Admitted to University	(3) Applied to university	(4) Applied to university
Treatment	0.06*	-0.01	0.06	-0.01
	(0.03)	(0.05)	(0.04)	(0.06)
Treatment * female		0.13*		0.15*
		(0.07)		(0.08)
Female	-0.07*	-0.13***	-0.05	-0.13**
	(0.04)	(0.05)	(0.04)	(0.06)
Mock score	0.08***	0.08***	0.07***	0.07***
	(0.02)	(0.02)	(0.02)	(0.02)
Age	-0.00	0.00		
	(0.01)	(0.02)		
STEM	-0.12***	-0.12***	-0.09*	-0.09*
	(0.04)	(0.04)	(0.05)	(0.05)
Constant	0.35	0.36	0.43***	0.46***
	(0.31)	(0.31)	(0.08)	(0.08)
Observations	708	708	559	559
R-squared	0.14	0.14	0.06	0.06
Control mean	0.30		0.45	
Control mean female		0.25		0.39
Overall tretament effect				
Female		0.12***		0.14**
		(0.05)		(0.06)

Admitted to University refers to obtaining a space at a public University, and was obtained directly from the UNCHE which maintains a record of all students. Applied to University refers to making an application and was obtained directly from the schools, with some schools declining to provide it. All regressions include school fixed effects. Age is age in years. Female is a dummy for being a female student. Stem is a dummy capturing students in the science track. Mock score is the overall score in the mock exam prior to the intervention. Control mean is the mean in the control group. Overall treatment effect is calculated by adding together the coefficient on treatment and on the treatment interaction for that group. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

ment still only has a positive impact on the maths exam, significant at the 5% level. At the S6 level, for all outcomes, treatment has a positive effect, significant at the 5% level.

4 Potential Mechanism Discussion

The role model in the treatment movie could have affected student schooling outcomes through a number of different channels (Chung, 2000), all of which have been documented from role models in prior literature. Firstly, the role model may raise aspirations broadly by encouraging students to have high goals and work hard to achieve them (Beaman et al., 2012; Bernard et al., 2014). Secondly, the role model could have provided information about the returns to schooling. Wilson (1987) argued that individuals form their beliefs about returns to education from individuals 'like them' and Nguyen (2008) finds a role model impacts educational outcomes through providing information about returns to schooling. This may raise their effort levels, improving educational outcomes (Jensen, 2010, 2012). Lastly, the movie could have provided a counter-stereotype role model, a woman from a deprived background succeeding at a man's game, that may have overcome stereotype threat amongst those groups subject to it, predominantly low-ability female students (Stout et al., 2011; Dasgupta and Asgari, 2004).[26] In this section, I examine the potential mechanisms and discuss the impacts I'd expect to see, sources of heterogeneity and intermediate outcomes needed to test each mechanism (see Table 10 for a summary). I will then discuss what can and can't be ruled out about the mechanism given my results.

If aspirations increasing was the dominant mechanisms by which the treatment movie improved schooling outcomes, I would expect to see increases in performance across subjects and impacts on long term outcomes, such as continuation in schooling and attending university, as students both raised their education goals and increased effort to achieve those goals. I would not expect to see improvements only in certain subjects[27]. I might see larger increases for those more similar to the role model (Bernard et al., 2014; Genicot and Ray, 2017; Ray, 2006), including female students, and those from a deprived background. However, female role models have also been shown to raise men's aspirations (Lockwood, 2006), and so it is possible I would also see some impacts on men. Measures

[26]There are other potential mechanisms, such as increased hope or inspirational effects due to seeing a film about local people but these are harder to separate out from the other mechanisms, and there is less research on them generally, and so I do not discuss them here.

[27]though if some subjects are more responsive to effort than others I may only see impacts on some outcomes

of aspirations and goal would allow me to determine if aspirations were the dominant mechanism.

If information about returns to schooling was the dominant mechanism through which the role model operated, I would also expect to see improvements in academic outcomes across the board due to increased effort[28]. The film only featured information about secondary school, not further schooling, so I would not expect to see any changes in university applications and attendance. The information would be most relevant to students most similar to the role model, so I would expect larger increases for those from a similar socio-economic background to the role model (Nguyen, 2008). I would expect to see differences by gender, since returns to education are gender specific (Attanasio and Kaufmann, 2014). Beliefs about returns to schooling would allow me to determine if returns to education were the dominant mechanism.

If stereotype threat was the dominant mechanism I would expect to see effects only in outcomes where stereotypes dominate and the role model can act to counter them (Dasgupta and Asgari, 2004; Stout et al., 2011; Dennehy and Dasgupta, 2017). These would be maths and science papers for female students, particularly female students with below average ability and who did not select into maths or science but are required to study them in compulsory papers (Spencer et al., 1999). I may also see effects on long term outcomes where stereotypes also affect one group; this is likely to also be female students, who are less likely than male students to continue to upper secondary school and university (MoES Uganda, 2015). Stereotype threat has also been shown to affect students from a poor background (Croizet and Claire, 1998), and so may be stronger in those from more deprived backgrounds. Beliefs and norms about the abilities of different groups would allow me to determine if stereotype threat were the dominant mechanism.

I do not have survey data capturing beliefs, norms, aspirations and effort, nor variation in my treatment that would allow me to examine mechanisms. As a result, I am limited to examining mechanisms through looking at which outcomes are affected and through analysing heterogeneous effects to see which types of students were most affected

[28]though again, I may only see impacts in certain subjects if performance in these subjects is more elastic with respect to increased effort

Table 10: Summary of likely channels of different mechanisms

Mechanism	Outcomes affected	Heterogeneity	Intermediate outcomes
Aspirations	Exam performance, continuation in school	similar to role model - from Kampala, low socio-economic status, female	aspirations and goals, effort
Returns to schooling	Exam performance	similar to role model - from Kampala, low socio-economic status, female	beliefs about returns, effort
Stereotype threat	Exam performance in maths/science, continuation in school for stereotyped group	low ability/ low socio-economic status, female	norms and beliefs about ability

by the intervention. As seen in Table 10 though, many of the mechanisms affect similar outcomes and would be expected to have similar heterogeneous effects. The main differences between the mechanisms are on whether there are impacts on students continuing in education (aspirations and stereotype threat) and whether impacts are concentrated in maths and science subjects (stereotype threat). However, while seeing impacts only on maths and science papers could hint at stereotype threats being the dominant mechanism, if maths and science results are more easily improved during the time period available before exams, I might only see impacts on maths and science papers despite the mechanism being returns to education or aspirations. Any indication as to the mechanism should therefore be thought of as tentative, proving an indication of where the findings are consistent with one channel over another, without being able to rule out other mechanisms.

The heterogeneous results from both the S4 and S6 classes point to effects being concentrated in female students, either with low prior ability or not in the science track, in compulsory maths-based papers. Effects are also stronger in those likely to be poorer financially, but not in those similar to the role model in other dimensions (for example from Kampala). This alone does not allow me to rule out or find evidence in favour of any particular mechanism. The fact that effects are concentrated in compulsory maths-based papers across both year groups could be evidence in favour of the stereotype threat

mechanism, as effects across subjects would be more likely if either raised aspirations or increased beliefs about returns to schooling were the dominant mechanism. However, as mentioned, it could also be that mathematical subjects are more responsive to effort in the time students had between the intervention and exams. Overall, the heterogeneity results do not strongly suggest one mechanism over the others.

Turning to the long term effects of treatment, I see an increased likelihood of lower secondary school students who saw *Queen of Katwe* continuing to upper secondary school. If the role model operated through increasing aspirations, this could result in students being more likely to remain in school, and particularly female students, if they started with lower aspirations or their aspirations were raised more by the female role model. If the mechanism was higher beliefs about returns to schooling, then I may also see students being more likely to stay in school, and if information was gender relevant then these impacts may be larger for female students. Finally, if the mechanism was stereotype threat, I would also expect the role model to counter the stereotype that women do not achieve higher levels of education and hence increase female students enrolment at upper secondary school. I therefore cannot determine from continuation to upper secondary school which mechanism is more likely.

However, the movie makes no mention of university, so it is not clear why it would have changed beliefs about returns to university. The fact that I also see strong treatment effects for female students on university applications and admittance, and their lower prior rates of going to university, suggests that information about returns to education is unlikely to be the mechanism that the role model is working through here. Either stereotype threat deterred women from attending university, or low aspirations amongst women, could explain why the role models increases university admittance and applications.

Further evidence against the returns to schooling mechanism is apparent when examining the content of the movie: the film shows Phiona returning to schooling so that she can pursue her love of chess. Schooling is not seen as a contributor to Phiona's success, but as a means to an end. Phiona does not need a high school diploma in order to play

chess, and her income comes from winning chess, nor finding a higher paying job tied to her schooling. It therefore seems unlikely that students are learning about returns to schooling directly by watching the movie. It seems more feasible that their aspirations are being raised by the inspiration content of the story, or stereotypes being overcome by Phiona's example of success at what is perceived as a man's game.

Overall, the heterogeneous treatment effects and long run findings suggest that aspirations or stereotype threat, or both, are the more likely mechanism that the role model effect works through. However, it will take further work where the treatment varies along dimensions designed to test alternative mechanisms or including survey variables capturing measures of the channels to shed more light on the mechanisms that role models operate through.

5 Cost effectiveness and impacts on earnings

5.1 Cost effectiveness

The study was extremely costs effective, with the total cost of the intervention only $3 per student for the cinema screening and $2 per student for transport to the cinema. This means there was only a cost of $5 per student to improve education outcomes by 0.11 standard deviations in maths at S4 and 0.13 standard deviations overall at S6. To compare this to other education interventions, I use the method in Kremer et al. (2013) of comparing the standard deviation of impact that could be had for $100 of spending. In my study, you could improve test scores by 2.2 to 2.6 standard deviations (by raising the scores of 20 students by 0.11-0.13 standard deviations) for $100 of spending. This is comparable to a remedial education programme in India which generated a 3 standard deviation test score gain per $100 spent (Banerjee et al., 2007) or to teacher incentives in Kenya (Glewwe et al., 2010). This study also showed the *Queen of Katwe* movie at a cinema. If the study was scaled up, the movie could be shown in schools to entire classes, perhaps through a projector or a specially arranged screening for many schools, and this would lower the cost further.

5.2 Impact on earnings

I do not have data on the impact of treatment on earnings. However, I can calculate an estimate of the income gains associated with the gain in educational attainment based on the returns to education in Uganda. Treatment led to a 6 percentage point increase in the likelihood of being admitted to university, increasing enrolment from 30% to 36%. According to government statistics, 94% of students admitted to government universities complete their programmes (Basheka, 2013). Kavuma (2014) found that those finishing university earn 120% more than those finishing just secondary school, so there are sizeable potential income gains from helping students get into university. Treatment also led to a 6% percentage point increase in the likelihood of finishing S6, with the return to an additional year of secondary schooling an 11% increase in wages.

6 Conclusion

I find that exposing secondary school students to a movie featuring a potential role model improves national exam performance. Amongst S4 students completing lower secondary school, seeing the treatment movie increases maths scores by 0.11 standard deviations, with the effect coming from lower ability students, particularly girls, being 50% less likely to fail maths. Treated female students are also 12 percentage points more likely to continue in school, closing the gender gap with boys.

At S6 level, amongst students trying to achieve the grades to get into university, I find seeing the treatment movie improves overall exam performance by 0.13 standard deviations I see heterogeneity by gender, with girls' scores increasing more from seeing the treatment movie. The impacts are largest for female students not in the sciences track, that the role model is likely to have the largest effects on. I also see that treated female students are 13 percentage points more likely to be admitted to university, also closing the gender gap with boys.

An implication of these findings is that schools should place more emphasis on having appropriate role models in schools, whether through showing a movie or through having former students come in to tell their stories. This is likely to be particularly effective at closing gender gaps and in helping the worst performing students.

One limitation of this work is that there was no pure control group, so it is possible that the placebo movie reduced exam performance. This seems unlikely given the effort that went into finding and reviewing movies that would be appropriate for the age group. I discuss this in the Appendix section A9.1.

A second limitation is that, due to lack of time and money, no individual surveys were done with the students, preventing a deeper understanding of the mechanisms by which the treatment movie worked. This study is therefore best viewed at providing evidence on whether a role model in a movie can affect important economic outcomes, and who experiences the largest effect. Further work should examine potential mechanisms for how the treatment movie led to changes in behaviour, either by designing role model interventions that vary on specific dimensions or including measures to capture mechanisms.

References

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Attanasio, O. P. and Kaufmann, K. M. (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 109:203–216.

Banerjee, A., La Ferrara, E., and Orozco, V. (2018). The Entertaining Way to Behavioral Change: Fighting HIV with MTV. *Working paper*.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics*, 122(3):1235–1264.

Basheka, B. C. (2013). The State of Higher Education and Training in 2013/14. Technical report.

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girl: A Policy Experiment in India. *Science*, 334(February):582–586.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive Linear Step-Up Procedures that Control the False Discovery Rate. *Biometrika*, 93(3):491–507.

Bernard, T., Dercon, S., Orkin, K., and Taffesse, A. S. (2014). The future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia. *CSAE Working Paper Series*, 44(April):48.

Bernard, T., Dercon, S., Orkin, K., and Taffesse, A. S. (2015). Will Video Kill the Radio Star? Assessing the Potential of Targeted Exposure to Role Models through Video. *World Bank Economic Review*, 29:S226–S237.

Breda, T., Grenet, J., Monnet, M., Effenterre, C. V., Breda, T., Grenet, J., Monnet, M., Effenterre, C. V., and Van Effenterre, C. (2018). Can female role models reduce the gender gap in science? Evidence from classroom interventions in French high schools.

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias. *Quarterly Journal of Economics*, 134(3):1163–1224.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: how professor gender perpetuates the gender gap. *Quarterly Journal of Economics*.

Chong, A. and Ferrara, E. L. (2009). Television and Divorce: Evidence From Brazilian Novelas. *Journal of the European Economic Association*, 7(2-3):458–468.

Chung, K.-S. (2000). Role Models and Arguments for Affirmative Action. *American economic review*, 90(3).

Croizet, J.-C. and Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24(6).

Dalton, P. S., Ghosal, S., and Mani, A. (2016). Poverty and Aspirations Failure. *Economic Journal*, 126(590):165–188.

Dasgupta, N. and Asgari, S. (2004). Seeing is Believing: Exposure to Counterstereotypic Women Leaders and its Effect on the Malleability of Automatic Gender Stereotyping. *Journal of Experimental Social Psychology*, 40(5):642–658.

Dennehy, T. C. and Dasgupta, N. (2017). Female Peer Mentors Early in College Increase Women's Positive Academic Experiences and Retention in Engineering. *Proceedings of the National Academy of Sciences*, 114(23):5964–5969.

Dickerson, A., McIntosh, S., and Valente, C. (2015). Do the maths: An analysis of the gender gap in mathematics in Africa. *Economics of Education Review*, 46(7174):1–22.

Dweck, C. S. (2000). *Self-Theories: Their Role in Motivation, Personality, and Development*. Pscyhology press.

Evans, D. K. and Popova, A. (2015). What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. *World Bank Policy Research Working Paper*.

Genicot, G. and Ray, D. (2017). Aspirations and Inequality. *Econometrica*, 85(2):489–519.

Green, M. C. and Brock, T. C. (2000). The Role of Transportation in the Persuasiveness of Public Narratives. *Journal of Personality and Social Psychology*, 79(5):701–721.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *Quarterly Journal of Economics*.

Jensen, R. and Oster, E. (2009). The Power of TV: Cable Television and Women's Status in India. *The Quarterly Journal of Economics*, 124(3):1057–1094.

Jensen, R. T. (2010). The (Perceived) Returns to Education and the Demand for Schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Kavuma, S. N. (2014). Private Returns to Education for the Wage-employees and Self-employed in Uganda.

Kearney, M. S. and Levine, P. B. (2015). Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing. *American Economic Review*, 105(12):3597–3632.

Kirkeboen, L., Leuven, E., and Magstad, M. (2016). FIELD OF STUDY, EARNINGS, AND SELF-SELECTION. *The Quarterly Journal of Economics*, 131(3).

Kofoed, M. S. and Mcgovney, E. (2016). The Effect of Same-Gender and Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point.

Kremer, M., Brannen, C., and Glennerster, R. (2013). The Challenge of Education and Learning in the Developing World. *Science*, 340(6130):297–300.

La Ferrara, E. (2016). Mass Media and Social Change: Can We Use Television to Fight poverty? *Journal of the European Economic Association*, 14(4):791–827.

La Ferrara, E., Chong, A., and Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*, 4(4):1–31.

Lockwood, P. (2006). "Someone Like Me Can Be Successfull": Do College Students Need Same-Gender Role Models? *Psychology of Women Quarterly*, 30(1):36–46.

Lockwood, P. and Kunda, Z. (1997). Superstars and Me: Predicting the Impact of Role Models on the Self. *Journal of Personality and Social Psychology*, 73(1):91–103.

Lybbert, T. J. and Wydick, B. (2018). Poverty, Aspirations, and the Economics of Hope. *Economic Development and Cultural Change*, 66(January):0–46.

Marx, D. M. and Roman, J. S. (2002). Female Role Models: Protecting Women's Math Test Performance. *Personality and Social Psychology Bulletin*, 28(9):1183–1193.

MoES Uganda (2015). Education Statistical Abstract. Technical report.

Moyer-Gusé, E. (2008). Toward a Theory of Entertainment Persuasion: Explaining the Persuasive Effects of Entertainment-Education Messages. *Communication Theory*, 18(3):407–425.

Nguyen, T. (2008). Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar. *MIT Job Market Paper*.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Lix Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., Nika Somogyi, M., Akrami, N., Ekehammar, B., Vianello, M., Banaji, M. R., and Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *PNAS*, 106(26):10593–10597.

OECD (2015). The ABC of Gender Equality in Education.

Paluck, E. L. and Green, D. (2009). Deference, Dissent, and Dispute Resolution: An Experimental Intervention Using Mass Media to Change Norms and Behavior in Rwanda. *American Political Science Review*, 103(4):622–644.

Porter, C. and Serra, D. (2018). Gender Differences in the Choice of Major: The Importance of Female Role Models.

Ray, D. (2006). Aspirations, Poverty and Economic Change. *In Understanding Poverty. Oxford: Oxford University Press.*, pages 409–421.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. Technical report.

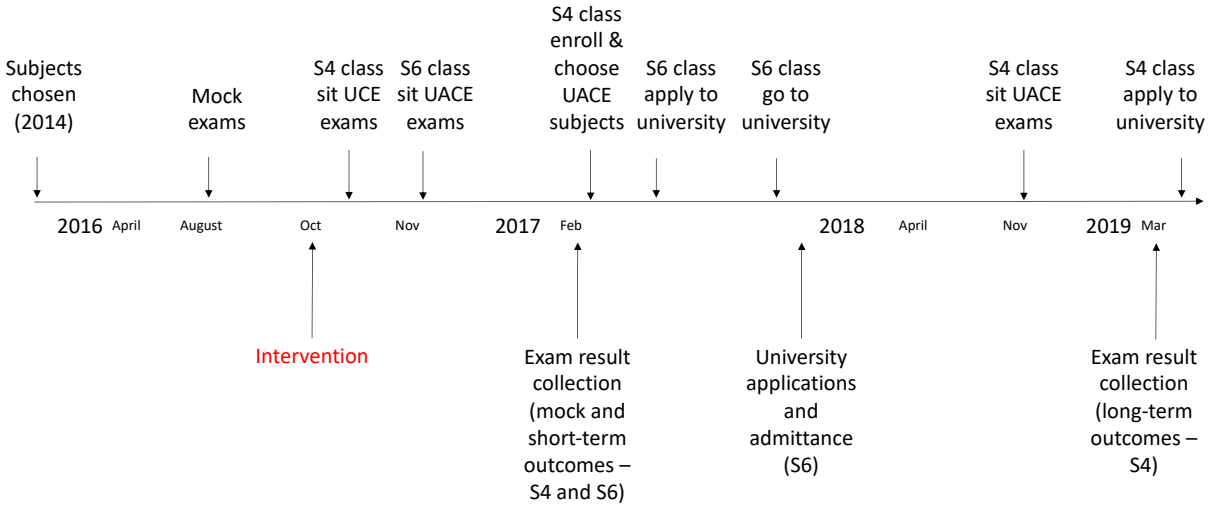
Stout, J. G., Dasgupta, N., Hunsinger, M., and McManus, M. A. (2011). STEMing the Tide: Using Ingroup Experts to Inoculate Women's Self-Concept in Science, Technology, Engineering, and Mathematics (STEM). *Journal of personality and social psychology*, 100(2):255–270.

Wilson, W. . (1987). *The Truly Disadvantaged*. University of Chicago Press.

Appendix

A1 Experimental design

Figure A1: Survey and exam timeline



Note that for both class all subjects had already been chosen before the intervention took place. The S4 class had not made their choice to continue in schooling to UACE, subject chosen at UACE if they continued, or university application or attendance choice. The S6 class had not yet applied to university or been admitted (applications are due by May 2017 to start August 2017)

Table A1: Balance test S4 class - additional background variables

	Placebo			Treatment				
	n	mean	sd	n	mean	sd	difference	p-value
Below median PLE	256	0.44	0.50	309	0.49	0.50	-0.05	(0.20)
From Kampala	256	0.21	0.41	309	0.20	0.40	0.01	(0.85)
Mother raised	256	0.12	0.33	309	0.13	0.34	-0.01	(0.68)
Parents professionals	256	0.16	0.37	309	0.19	0.39	-0.03	(0.40)
Bursary	256	0.18	0.38	309	0.15	0.36	0.02	(0.45)

Below median PLE means the student scored below the median in the primary school leaving exam. From Kampala means the student's home address is within Kampala. Mother raised means the father is deceased. Parents professional means at least one parent has a professional occupation. Bursary means the student receives financial support from the school.

Table A2: Balance test S6 class - additional background variables

	Placebo			Treatment				
	n	mean	sd	n	mean	sd	diff	p-value
Below median UCE	269	0.48	0.50	267	0.56	0.50	-0.08	(0.07)
From Kampala	274	0.22	0.41	286	0.17	0.37	0.05	(0.13)
Mother raised	274	0.14	0.34	286	0.16	0.36	-0.02	(0.46)
Parents professionals	274	0.17	0.37	286	0.17	0.38	-0.01	(0.83)
Bursary	274	0.24	0.43	286	0.26	0.44	-0.01	(0.70)

Below median UCE means the student scored below the median in the lower secondary school leaving exam. From Kampala means the student's home address is within Kampala. Mother raised means the father is deceased. Parents professional means at least one parent has a professional occupation. Bursary means the student receives financial support from the school.

Table A3: Individual and school characteristics correlated with attrition

	Attrition
Boarding	0.01
	(0.01)
High fees	-0.02*
	(0.01)
Christian	-0.05***
	(0.01)
Age	-0.02***
	(0.01)
Female	-0.01
	(0.01)
S4	-0.02*
	(0.01)
Observations	1,498
R-squared	0.05

This table shows attrition correlates with 3 school characteristics and 3 student characteristics. Boarding refers to whether the school only has boarding pupils, high fees if the fees charged are above the median in this sample, Christian is the schools religious affiliation. Age is the age in years, S4 is a dummy if that student is in the S4 class. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A2 School and test score data

A2.1 S4

The exams are graded with a score from 1-9 with 1 being the best score and 9 the worst. Passing grades are considered to be an 8 or lower. For a candidate sitting 10 exams, the best score is therefore 10 and the worst 0. For this analysis scores have been inverted so that a 9 becomes 0 and a 1 becomes 8. This is so that a higher score can be interpreted as a better performance, while a higher aggregate score can indicate better performance per paper or more papers taken.

Standardized test scores have been created for each subject by subtracting the mean and dividing by the standard deviation of the control group. An overall aggregate of exam performance was calculated by summing standardised test scores across all subjects and renormalising. A core index of exam performance was calculated by summing test scores across the six core subjects and renormalising.

A2.2 S6

Every student sits the general paper and one of two subsidiary papers in maths or computer. Those taking maths as a principle subject must take computer in the subsidiary paper. The subsidiary subjects and general paper are graded on a 1-9 scale, with 1 being the best and 9 the worst grade. Grades 7 and above are fails. For this analysis scores have been inverted so that a 9 becomes 0 and a 1 becomes 8. This is so that a higher score can be interpreted as a better performance. The subsidiary and general paper score is therefore out of a maximum of 18 and a minimum of 0.

Students select into 3 principal papers. The principal papers are marked on a A, B, C scale, with an A earning 6 points, a B 5 points etc. The maximum a student could score on the principal papers is 3 As, giving 18 points in total.

To calculate the total score, the subsidiary and principal papers are given different weights. Any student achieving a 6 or below (before inverting) on a subsidiary paper or the general paper gets one point. The maximum of 2 points earned on the subsidiary and general paper are added to the points earned on the principal papers to give the total score. This means the highest total score a student could earn is three As and passes on each of the subsidiary and general papers, giving 20 points in total.

Standardised test scores were constructed for each subject by subtracting the mean and dividing by the standard deviation of the control group. An overall index of exam performance was calculated by summing test scores across all subjects and renormalising.

Table A4: School Characteristics

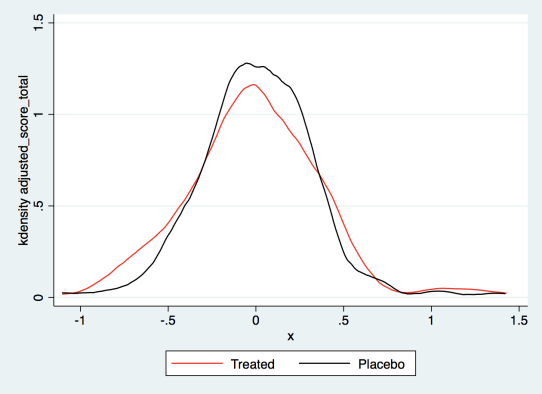
School	Religion	Boarding	Ranking		Fees				Class size	
			UACE Rank /1882	UCE Rank /3294	S4 board	S4 day	S6 board	S6 day	S4	S6
Hope	Islamic	Boarding only	7	94	650,000		650,000		93	65
Paul Musaka	Christian	Mixed day and boarding	220	199	680,000	340,000	680,000	360,000	136	80
Kyandondo	Islamic	Mixed day and boarding	271	537	730,000	530,000	730,000	530,000		187
Makerere	Christian	Mixed day and boarding	342	464	450,000	250,000	500,000	300,000	85	47
Royal	Christian	Boarding only	461	32	600,000		650,000		110	93
Kinaawa	Islamic	Boarding only	492	430	900,000		900,000			94
Jakayza	Islamic	Mixed day and boarding	525	1047	460,000	230,000	480,000	245,000		25
Mukono	Christian	Mixed day and boarding	527	472	600,000	450,000	600,000	450,000	82	57
Atlas	Christian	Mixed day and boarding	529	170	900,000	450,000	920,000	470,000		40
Gayaza	Islamic	Mixed day and boarding	931	2020	470,000	208,000	500,000	220,000		12
Dynamic	Christian	Mixed day and boarding	1423	2036	550,000	180,000	400,000	180,000	141	
Kulumba	Islamic	Mixed day school	1782	1205		170,000		220,000	21	5
Devine	Islamic	Mixed day and boarding	1799	2007	440,000	210,000	500,000	250,000	53	5

Religion is the religious affiliate reported by the school. Students are taken to pray on religious days and 5 times a day at Islamic schools. Fees are in Ugandan Shillings per year. Class size refers to the size of the class if it participated in the study. Schools either gave the entire class or not at all, never part of a class.

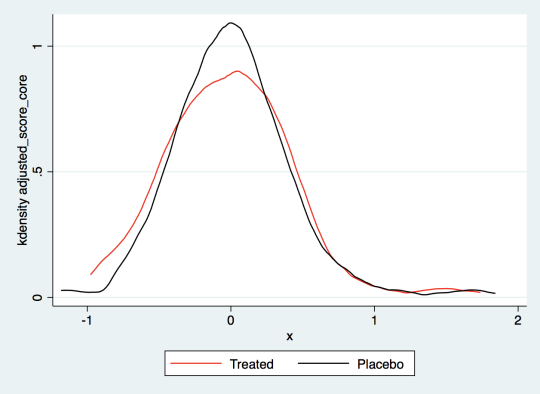
A3 Histograms

Figure A2: Histograms of S4 student results by treatment assignment

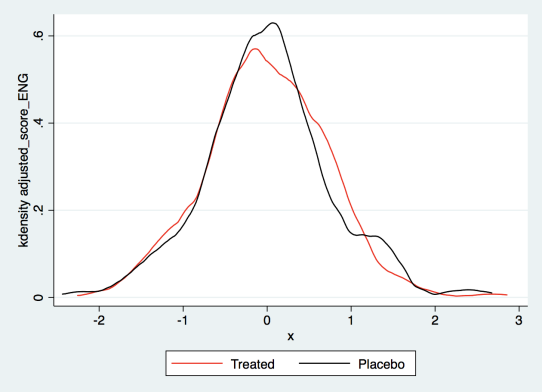
(a) Total score



(b) Core score



(c) English



(d) Maths

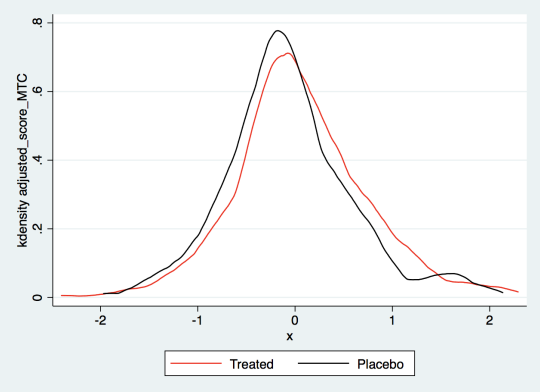
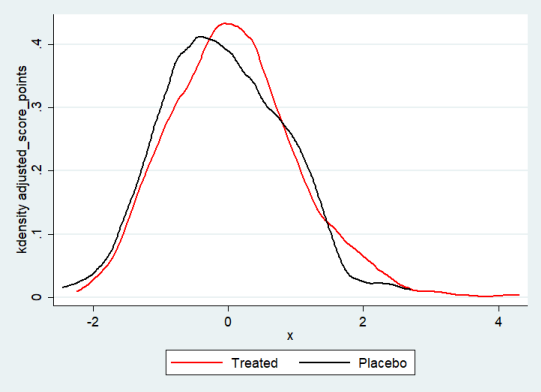
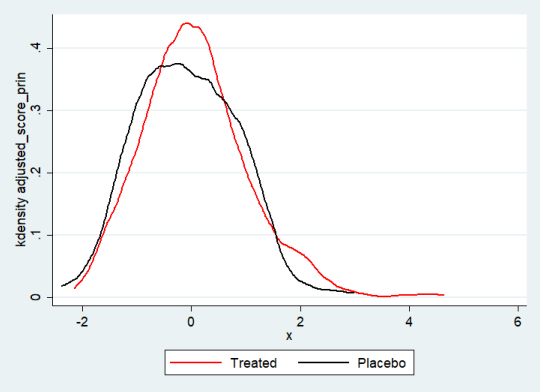


Figure A3: Histograms of S6 student results by treatment assignment

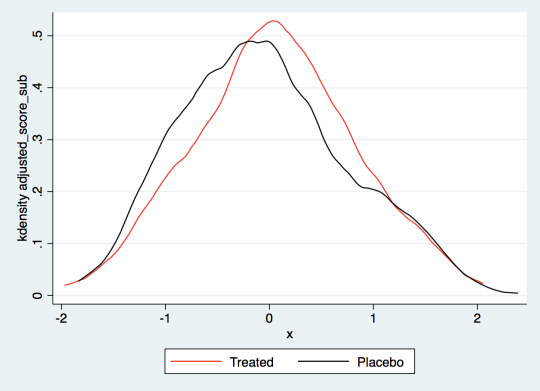
(a) Total score



(b) Principal papers



(c) Maths/computer papers



A4 Logistics regressions

Table A5: Ordered logit regression of the impact of treatment on maths grade at S4

Grade	treatment
Fail	-0.052***
	(0.015)
1	-0.016***
	(0.005)
2	0.000
	(0.001)
3	0.013***
	(0.004)
4	0.015***
	(0.004)
5	0.016***
	(0.005)
6	0.012***
	(0.004)
7	0.008***
	(0.003)
8	0.004***
	(0.002)
Observations	729

Papers are graded from fail (0) to highest grade (8). Regressions include school fixed effects and individual control variables (age, gender, number of subjects taken) as well as the standardised mock exam score in maths. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A6: Ordered logit regression of the impact of treatment on core subject grade for S4

	(1) Eng	(2) Chem	(3) Phy	(4) Bio	(5) Hist	(6) Geog
Fail	-0.002	0.018	0.015	0.009	0.011	0.001
	(0.010)	(0.017)	(0.019)	(0.015)	(0.012)	(0.011)
2	-0.001	-0.000	-0.002	0.001	0.001	0.000
	(0.006)	(0.000)	(0.003)	(0.001)	(0.001)	(0.002)
3	-0.001	-0.003	-0.003	-0.001	0.001	0.000
	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)
4	-0.000	-0.003	-0.002	-0.002	0.001	0.000
	(0.003)	(0.003)	(0.003)	(0.004)	(0.001)	(0.001)
5	0.001	-0.004	-0.002	-0.003	0.000	-0.000
	(0.004)	(0.004)	(0.002)	(0.004)	(0.001)	(0.001)
6	0.002	-0.003	-0.002	-0.002	-0.001	-0.000
	(0.008)	(0.003)	(0.002)	(0.003)	(0.001)	(0.004)
7	0.001	-0.003	-0.002	-0.001	-0.003	-0.001
	(0.005)	(0.003)	(0.002)	(0.002)	(0.003)	(0.005)
8	0.000	-0.001	-0.001	-0.001	-0.004	-0.000
	(0.003)	(0.001)	(0.002)	(0.001)	(0.004)	(0.003)
9	0.000	-0.001	-0.000	-0.000	-0.006	-0.000
	(0.002)	(0.001)	(0.000)	(0.000)	(0.006)	(0.002)
Observations	729	728	728	729	727	706

Coefficients are the marginal effects of treatment on the likelihood of achieving each grade. All regressions include school fixed effects and individual characteristics (age, gender, number of subjects taken) and mock score in that subject (for English) or overall. Core subjects are taken by all students at S4 level. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A7: Impact of treatment assignment on S4 standardised tests scores

	(1) Total score excl. maths	(2) Total score excl. maths
Treatment	-0.03	-0.03
	(0.02)	(0.02)
Age		0.00
		(0.01)
Female		-0.03
		(0.02)
No. subjects		-0.05**
		(0.02)
Mock score	0.82***	0.82***
	(0.01)	(0.01)
Constant	-0.15**	0.21
	(0.06)	(0.26)
Observations	735	729
R-squared	0.88	0.88

Total score excl. maths refers to the total score excluding the maths score. Standardized test scores composed of subject standardized scores and renormalised. All regressions include school fixed effects and the standardised mock score.
Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A8: Ordered logit regression of treatment on raw points scored at S6

	(1)	(2)	(3)
Grade	Overall score	Principal score	Maths/computer score
Fail		-0.000	-0.009**
		(0.000)	(0.004)
1	-0.001	-0.002	-0.012**
	(0.001)	(0.001)	(0.005)
2	-0.003	-0.005	-0.016**
	(0.002)	(0.003)	(0.007)
3	-0.004*	-0.006	-0.024**
	(0.003)	(0.004)	(0.009)
4	-0.007*	-0.006	-0.011**
	(0.004)	(0.004)	(0.004)
5	-0.006*	-0.007	-0.003*
	(0.003)	(0.004)	(0.002)
6	-0.008*	-0.005	0.006**
	(0.005)	(0.003)	(0.003)
7	-0.005*	-0.002	0.012**
	(0.003)	(0.002)	(0.005)
8	-0.002*	0.001	0.015**
	(0.001)	(0.000)	(0.006)
9	0.000	0.003	0.014**
	(0.000)	(0.002)	(0.006)
10	0.003*	0.004	0.010**
	(0.002)	(0.003)	(0.004)
11	0.004*	0.005	0.007**
	(0.002)	(0.003)	(0.003)
12	0.005*	0.004	0.007**
	(0.003)	(0.003)	(0.003)
13	0.005*	0.004	0.003*
	(0.003)	(0.002)	(0.001)
14	0.004*	0.004	
	(0.003)	(0.002)	
15	0.003*	0.002	
	(0.002)	(0.001)	
16	0.003*	0.002	
	(0.002)	(0.002)	
17	0.002	0.002	
	(0.001)	(0.002)	
18	0.002	0.001	
	(0.001)	(0.001)	
19	0.003		
	(0.002)		
20	0.001		
	(0.001)		
Observations	708	708	654

Coefficients are the marginal effects of treatment on the likelihood of achieving each grade. Grade is the raw score achieved in that outcome. Principal papers are graded from 0 to 6 in 3 papers, giving a maximum score of 18. Maths/computer papers are graded from 0 to 8 in two papers, giving a maximum score of 16. The overall score is the principal score plus 1 point for each passed maths/computer paper and a general paper (a score of 3 or more), giving a maximum score of 20 (no one scored 0 overall). Regressions include school fixed effects and the overall mock score. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A5 Heterogeneity

Table A9: Heterogeneity in treatment effect for S4 by gender and prior ability

	Impact of treatment assignment on standardized test scores									
	(1) Total Score	(2) Core Score	(3) Maths	(4) English	(5) Fail Maths	(6) Total Score	(7) Core Score	(8) Maths	(9) English	(10) Fail Maths
treatment	0.01	-0.00	0.03	-0.03	-0.08**	-0.02	-0.03	0.01	-0.02	0.05
	(0.04)	(0.05)	(0.07)	(0.08)	(0.04)	(0.04)	(0.05)	(0.07)	(0.07)	(0.03)
	[0.99]	[0.99]	[0.99]	[0.99]	[0.43]	[0.99]	[0.99]	[0.99]	[0.99]	[0.99]
treatment*female	-0.05	-0.04	0.15	0.04	-0.07					
	(0.05)	(0.06)	(0.10)	(0.11)	(0.05)					
	[0.99]	[0.99]	[0.99]	[0.99]	[0.99]					
female	-0.01	-0.02	-0.09	-0.19**	0.07*	-0.04	-0.04	0.03	-0.07	0.02
	(0.04)	(0.05)	(0.07)	(0.08)	(0.04)	(0.03)	(0.03)	(0.05)	(0.05)	(0.03)
treatment*below mock						-0.01	0.02	0.23**	0.03	-0.33***
						(0.05)	(0.07)	(0.10)	(0.10)	(0.05)
						[0.99]	[0.99]	[0.10]*	[0.99]	[0.00]***
below mock						-0.02	-0.02	-0.81***	-0.95***	0.33***
						(0.06)	(0.07)	(0.08)	(0.08)	(0.04)
Age	-0.00	-0.01	-0.07***	-0.07***	0.02**	-0.00	-0.01	-0.06***	-0.05**	0.02**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
No. Subjects	-0.07***	-0.08**	-0.02	0.08	-0.02	-0.07**	-0.08**	-0.05	0.03	-0.01
	(0.03)	(0.03)	(0.05)	(0.06)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.03)
Mock Score	1.01***	0.96***	0.66***	0.52***	-0.12***	1.00***	0.96***	0.51***	0.32***	-0.09***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Constant	0.33	0.33	0.82	0.22	0.42	0.33	0.34	1.27**	0.78	0.16
	(0.31)	(0.37)	(0.59)	(0.65)	(0.30)	(0.31)	(0.38)	(0.55)	(0.57)	(0.28)
Observations	729	729	729	729	729	729	729	729	729	729
R-squared	0.88	0.83	0.54	0.47	0.30	0.88	0.83	0.61	0.59	0.36

All regressions include school fixed effects. Age is age in years. No. subjects is a dummy variable if the student is taking less subjects than the median for the UCE exams. Below median mock is a dummy variable for if the student performed below the median in a mock exam. Mock score is the standardised score achieved in the maths mock exam taken prior to treatment. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sharped q-values in square brackets

Table A10: Heterogeneity in treatment effects for S4 by mock decile

	(1) maths	(2) fail maths
treatment	0.22*	-0.25***
	(0.13)	(0.07)
treatment * decile 2	-0.12	0.01
	(0.21)	(0.11)
treatment * decile 3	0.05	0.03
	(0.20)	(0.10)
treatment * decile 4	0.22	-0.12
	(0.20)	(0.10)
treatment * decile 5	0.02	0.13
	(0.19)	(0.10)
treatment * decile 6	-0.13	0.30***
	(0.20)	(0.10)
treatment * decile 7	-0.09	0.27***
	(0.19)	(0.10)
treatment * decile 8	-0.27	0.27**
	(0.20)	(0.10)
treatment * decile 9	-0.27	0.21**
	(0.19)	(0.10)
treatment * decile 10	-0.29	0.28***
	(0.20)	(0.10)
Overall treatment effect		
decile 1	0.22*	-0.25***
	(0.13)	(0.07)
decile 2	0.10	-0.24***
	(0.16)	(0.08)
decile 3	0.27*	-0.23***
	(0.15)	(0.08)
decile 4	0.44***	-0.38***
	(0.15)	(0.08)
decile 5	0.24*	-0.13*
	(0.13)	(0.07)
decile 6	0.10	0.05
	(0.15)	(0.08)
decile 7	0.14	0.01
	(0.14)	(0.07)
decile 8	-0.05	0.01
	(0.16)	(0.08)
decile 9	-0.05	-0.05
	(0.14)	(0.07)
decile 10	-0.06	0.02
	(0.15)	(0.08)
Constant	0.74	0.72***
	(0.55)	(0.28)
Observations	729	729
R-squared	0.63	0.43

Decile refers to decile of mock exam score. Regressions include school fixed effects and individual controls of age, gender, number of subjects taken and mock decile. The second panel shows the overall effect of treatment for each decile. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A11: Heterogeneity in treatment effect for S6 by stem subject and mock performance

	Overall score (1)	principal score (2)	maths/ computer score (3)	Overall score (4)	principal score (5)	maths/ computer score (6)
Treatment	0.15**	0.17***	0.15**	0.13*	0.12*	-0.00
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)
	[0.03]**	[0.03]**	[0.03]**	[0.54]	[0.29]	[0.99]
Treatment * STEM	-0.04	-0.11	-0.07			
	(0.10)	(0.11)	(0.12)			
	[0.99]	[0.99]	[0.99]			
STEM	0.58***	0.28***	0.49***	0.57***	0.27***	0.40***
	(0.07)	(0.08)	(0.09)	(0.06)	(0.06)	(0.06)
Treatment * Below median mock				0.00	0.01	0.21*
				(0.09)	(0.10)	(0.11)
				[0.94]	[0.99]	[0.37]
Below median mock				-0.29***	-0.36***	0.05
				(0.09)	(0.10)	(0.09)
Age	-0.03	-0.02	-0.04	-0.03	-0.02	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Female	0.05	0.12**	0.03	0.06	0.13**	0.03
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Mock overall score	0.76***	0.71***	0.39***	0.64***	0.58***	0.48***
	(0.03)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)
Constant	-0.16	-0.08	0.29	-0.03	0.05	-0.82
	(0.43)	(0.45)	(0.51)	(0.43)	(0.45)	(0.57)
Overall treatment effect						
STEM	0.11	0.06	0.08			
	(0.09)	(0.09)	(0.10)			
Below median mock				0.13*	0.13*	0.20**
				(0.07)	(0.07)	(0.08)
Observations	708	708	708	708	708	646
R-squared	0.62	0.59	0.49	0.63	0.60	0.52

Overall score is the aggregate score achieved in the principal papers plus a point for each passed maths/computer and general paper, principal score is the score in 3 chosen principal papers, maths/computer score is the score on the two compulsory maths or computer and general papers. All scores are standardized. All regressions control for school fixed effects.Age refers to age in years. Mock score is the standardised mock score for that outcome. Below mock is a dummy variable if the student scored in the bottom half of scores for that outcome variable in a mock exam. STEM refers to choosing maths, biology, chemistry or physics as a principal subject. The bottom panel shows the overall effect for each subgroup. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sharpened q-values in square brackets.

Table A12: Heterogeneity in treatment effects for S6 by mock decile

	(1)	
	total score	maths/ computer score
treatment	0.13	0.11
	(0.12)	(0.36)
treatment * decile mock 1	-0.12	0.13
	(0.20)	(0.38)
treatment * decile mock 2	0.19	0.09
	(0.19)	(0.41)
treatment * decile mock 3	-0.19	0.21
	(0.24)	(0.39)
treatment * decile mock 4	-0.18	-0.04
	(0.25)	(0.39)
treatment * decile mock 5	-0.02	-0.00
	(0.28)	(0.40)
treatment * decile mock 6	-0.40	-0.16
	(0.25)	(0.41)
treatment * decile mock 7	-0.21	-0.07
	(0.27)	(0.41)
treatment * decile mock 8	-0.18	-0.47
	(0.27)	(0.41)
Constant	2.00***	-0.41
	(0.44)	(0.63)
Overall treatment effect		
decile 1	0.12	0.24**
	(0.11)	(0.12)
decile 2	0.42***	0.20
	(0.14)	(0.19)
decile 3	0.04	0.32**
	(0.14)	(0.16)
decile 4	0.06	0.07
	(0.15)	(0.15)
decile 5	0.21	0.11
	(0.19)	(0.18)
decile 6	-0.16	-0.05
	(0.16)	(0.21)
decile 7	0.03	0.05
	(0.19)	(0.21)
decile 8	0.06	-0.35
	(0.18)	(0.21)
Observations	708	708
R-squared	0.57	0.38

Decile refers to decile of mock performance for that exam score outcome. Regressions include school fixed effects and individual controls of age, gender, if taking a stem (maths or science) paper and the mock decile. The second panel shows the overall effect of treatment for each decile. There are 9 deciles since two deciles had equal scores associated with them, decile 9 is the excluded category. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A13: Heterogeneous effects of treatment assignment by background characteristics on standardized test score at S4

	(1) Maths	(2) Fail Maths	(3) Maths	(4) Fail Maths	(5) Maths	(6) Fail Maths	(7) Maths	(8) Fail Maths	(9) Maths	(10) Fail Maths
Treatment	0.04	-0.04	0.12**	-0.10***	0.09	-0.10***	0.13**	-0.12***	0.06	-0.08**
	(0.08)	(0.04)	(0.06)	(0.03)	(0.06)	(0.03)	(0.06)	(0.03)	(0.06)	(0.03)
	[0.91]	[0.83]	[0.24]	[0.01]***	[0.10]*	[0.00]***	[0.06]*	[0.00]***	[0.52]	[0.05]**
Treatment * Below PLE	0.12	-0.12**								
	(0.12)	(0.06)								
	[0.45]	[0.12]								
Below median PLE	-0.21**	0.08*								
	(0.08)	(0.05)								
Treatment * Kampala			-0.14	0.03						
			(0.15)	(0.07)						
			[0.99]	[0.99]						
Kampala			0.21*	-0.10*						
			(0.11)	(0.05)						
Treatment * Mother					0.04	-0.01				
					(0.18)	(0.08)				
					[0.97]	[0.97]				
Mother					0.04	-0.08				
					(0.13)	(0.06)				
Treatment * professionals							-0.22*	0.09		
							(0.14)	(0.08)		
							[0.33]	[0.33]		
Professionals							0.12	0.07		
							(0.10)	(0.06)		
Treatment * Bursary									0.19	-0.14*
									(0.15)	(0.08)
									[0.36]	[0.20]
Bursary									-0.20*	0.06
									(0.11)	(0.06)
Mock	0.65***	-0.11***	0.66***	-0.11***	0.66***	-0.11***	0.66***	-0.11***	0.66***	-0.10***
	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)
Constant Constant	1.95**	0.39	1.79**	0.45	1.74**	0.48	1.73**	0.53	1.72**	0.54
	(0.82)	(0.45)	(0.81)	(0.43)	(0.82)	(0.44)	(0.81)	(0.43)	(0.81)	(0.44)
Observations	562	562	562	562	562	562	562	562	562	562
R-squared	0.54	0.30	0.54	0.30	0.54	0.29	0.54	0.30	0.54	0.30

All scores are standardized. Below median PLE school is a dummy variable capturing whether the student performed below the median in the PLR exam taken at the end of primary school. Kampala is a dummy variable equal to one if the student comes from Kampala. Mother is a dummy variable equal to one if the student was raised by their mother alone (father is dead). Professional is a dummy variable equal to one if one of the student's parents has a professional occupation. Bursary is a dummy variable equal to one if the student received a bursary from their school. All regressions include school fixed effects and control for age, whether the student is female the number of subjects taken at UCE. Mock score is the score in the mock exam prior to the intervention in maths. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sharpened q-values in square brackets.

Table A14: Heterogeneous effects of treatment assignment by background characteristics on standardized test score at S6

	(1) Overall	(2) Prin	(3) MTC	(4) Overall	(5) Prin	(6) MTC	(7) Overall	(8) Prin	(9) MTC	(10) Overall	(11) Prin	(12) MTC	(13) Overall	(14) Prin	(15) MTC
Treatment	0.12	0.12	0.16*	0.13**	0.11*	0.13**	0.12*	0.15**	0.06	0.19***	0.18***	0.17***	0.12*	0.12*	0.06
	(0.09)	(0.10)	(0.08)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
	[0.42]	[0.42]	[0.28]	[0.13]	[0.18]	[0.13]	[0.14]	[0.14]	[0.71]	[0.02]**	[0.02]**	[0.02]**	[0.24]	[0.24]	[0.67]
Treat * below UCE	0.17	0.17	-0.00												
	(0.11)	(0.12)	(0.12)												
	[0.44]	[0.44]	[0.99]												
Below median UCE	-0.31***	-0.37***	-0.12												
	(0.08)	(0.08)	(0.08)												
Treat * Kampala				0.13	0.20	-0.04									
				(0.13)	(0.13)	(0.15)									
				[0.87]	[0.81]	[0.99]									
Kampala				-0.11	-0.11	0.01									
				(0.09)	(0.09)	(0.10)									
Treat * Mother							0.21	0.03	0.50***						
							(0.15)	(0.16)	(0.15)						
							[0.39]	[0.99]	[0.01]***						
Mother							-0.09	0.07	-0.38***						
							(0.10)	(0.12)	(0.11)						
Treat * professional										-0.18	-0.16	-0.29**			
										(0.14)	(0.15)	(0.14)			
										[0.53]	[0.53]	[0.21]			
Professional										0.06	0.07	0.14			
										(0.11)	(0.12)	(0.10)			
Treat * bursary													0.14	0.11	0.25*
													(0.12)	(0.13)	(0.13)
													[0.68]	[0.72]	[0.31]
Bursary													-0.01	-0.02	-0.00
													(0.08)	(0.09)	(0.09)
Mock	0.72***	0.63***	0.45***	0.75***	0.68***	0.44***	0.74***	0.67***	0.45***	0.74***	0.67***	0.44***	0.75***	0.67***	0.45***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Constant	0.99**	0.04	1.16**	0.81*	-0.11	1.12**	0.76	-0.20	1.18**	0.74	-0.21	1.12**	0.74	-0.20	1.13**
	(0.47)	(0.52)	(0.52)	(0.48)	(0.52)	(0.52)	(0.47)	(0.52)	(0.51)	(0.47)	(0.52)	(0.52)	(0.47)	(0.52)	(0.52)
Obs	535	535	535	559	559	559	559	559	559	559	559	559	559	559	559
R-squared	0.62	0.58	0.55	0.60	0.56	0.54	0.60	0.56	0.55	0.60	0.56	0.54	0.60	0.56	0.54

All scores are standardized. Below median UCE school is a dummy variable capturing whether the student performed below the median in the UCE exam taken prior to entering upper secondary school. Kampala is a dummy variable equal to one if the student comes from Kampala. Mother is a dummy variable equal to one if the child was raised by their mother (father died). Professional is a dummy variable equal to one if one of the student's parents has a professional occupation. Bursary is a dummy variable equal to one if the student received a bursary from their school. All regressions include school fixed effects and control for age, whether the student is female and whether the student was taking STEM subjects (maths or science paper). Mock score is the score in the mock exam prior to the intervention for that outcome. Robust standard errors in parentheses. Prin is the principal paper score and MTC the maths and computer paper score. *** p<0.01, ** p<0.05, * p<0.1. Sharpened qvalues in square brackets

A6 Long term outcomes

Table A15: Heterogeneous impacts of treatment assignment on continuation in upper secondary school

	(1)	(2)	(3)	(4)
Outcome variable: continuing to S5				
treatment	0.06*	0.06*	0.02	0.04
	(0.04)	(0.04)	(0.05)	(0.05)
treatment*female			0.10	
			(0.07)	
female		-0.03	-0.08	-0.03
		(0.04)	(0.05)	(0.04)
treatment*below median mock				0.05
				(0.07)
below median mock				0.11
				(0.07)
age		-0.03**	-0.03**	-0.03**
		(0.01)	(0.01)	(0.01)
no. subjects		-0.07**	-0.07*	-0.08**
		(0.04)	(0.04)	(0.04)
Mock score		0.05**	0.05**	0.11***
		(0.02)	(0.02)	(0.04)
Constant	0.39***	1.50***	1.50***	1.49***
	(0.10)	(0.40)	(0.40)	(0.41)
Mean control	0.29	0.29	0.25	0.25
Observations	729	729	729	729
R-squared	0.04	0.06	0.06	0.06
Overall treatment effect				
female			0.11**	
			(0.05)	
below median mock				0.08*
				(0.05)

Continuation to S5 is a dummy variable equal to one if that student continued to the first grade of upper secondary school, S5. No. subject is the number of subjects taken at UCE. Age is the age in years. Female is a dummy variable for female student. Below median mock is a dummy variable if the student scored below the median mark on the mock exam. Mock score is the standardised aggregate mock score. The control mean is the mean for the control group, and for a particular sub-group if heterogeneity is examined. The bottom panel shows the overall treatment effect for each sub-group. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A16: Impact of treatment assignment on standardized test scores at S6 for original S4 students

	(1) total	(2) principal	(3) maths/ computer	(4) total	(5) principal	(6) maths/ computer
treatment	0.10	0.07	0.05	0.07	0.04	0.03
	(0.11)	(0.09)	(0.11)	(0.15)	(0.13)	(0.16)
treatment*female				0.07	0.06	0.04
				(0.21)	(0.19)	(0.22)
female	-0.02	0.03	0.08	-0.06	-0.00	0.05
	(0.10)	(0.09)	(0.11)	(0.16)	(0.12)	(0.17)
age	0.03	-0.04	0.00	0.03	-0.04	0.00
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
no. subjects	-0.10	0.00	-0.13	-0.10	0.00	-0.13
	(0.09)	(0.09)	(0.13)	(0.09)	(0.09)	(0.13)
Constant	0.01	0.01	0.73	0.01	-0.01	0.72
	(1.02)	(1.01)	(1.22)	(1.02)	(1.01)	(1.22)
Observations	221	188	188	221	188	188
R-squared	0.44	0.64	0.50	0.45	0.64	0.50
Overall treatment effect						
Female				0.14	0.10	0.07
				(0.15)	(0.13)	(0.16)

All scores are standardized. No. subject is the number of subjects taken at UCE. Age is the age in years. Female is a dummy variable for female student. Mock score is the standardised aggregate mock score. The bottom panel shows the overall effect of treatment from adding the treatment and interaction effects together. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A17: Impact of treatment assignment on choosing maths as principal subject at S6, original S4 sample

	(1) Maths	(2) Maths
treatment	0.012	-0.066
	(0.056)	(0.074)
treatment*female		0.173
		(0.109)
female	0.038	-0.066
	(0.055)	(0.085)
Constant	-1.314*	-1.290*
	(0.692)	(0.690)
Observations	222	222
R-squared	0.111	0.122
mean control	0.187	0.162
Overall effect		
Female		0.11
		0.08

Maths is a dummy variable capturing taking that subject as a principal subject at S6. Female is a dummy capturing student gender. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A18: Impact of treatment assignment on applying for university for original s4 class

	(1)	(2)	(3)
Dependent variable: applied for university			
Treatment	-0.01	0.04	-0.02
	(0.07)	(0.10)	(0.10)
Treatment * Female		-0.04	
		(0.13)	
Female	-0.08	-0.07	-0.10
	(0.07)	(0.10)	(0.07)
Treatment * Below median mock			0.10
			(0.14)
Below median mock			0.01
			(0.15)
Age	0.04	0.03	0.03
	(0.03)	(0.03)	(0.03)
subjects	-0.03	-0.03	-0.04
	(0.06)	(0.06)	(0.06)
Constant	-0.44	-0.45	-0.35
	(0.66)	(0.66)	(0.67)
Observations	185	185	185
R-squared	0.11	0.12	0.12
Control mean	0.29	0.27	0.18
Overall effect			
Female		-0.00	
		(0.10)	
Below median mock			0.08
			(0.10)

Admitted to University refers to obtaining a space at a public University. All regressions include school fixed effects and control for the score in a mock exam prior to the intervention. Age is age in years. Female is a dummy for being a female student. Control mean is the mean in the control group for heterogeneity group Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A19: Impact of treatment assignment on minimum grades at S4 to do a vocational certificate

	(1)	(2)	(3)
Dependent variable: Vocational passing grade			
treatment	0.10***	0.09***	0.07**
	(0.03)	(0.03)	(0.03)
age		-0.03**	-0.03**
		(0.01)	(0.01)
female		-0.03	-0.05
		(0.03)	(0.04)
treatment*female			0.04
			(0.05)
no. subjects		-0.00	-0.00
		(0.03)	(0.03)
Mock score		0.15***	0.15***
		(0.03)	(0.03)
Constant	0.30**	0.85***	0.84***
	(0.12)	(0.30)	(0.30)
control mean	0.70	0.70	
control mean female			0.67
Observations	735	729	729
R-squared	0.26	0.40	0.40

Vocational passing grade refers to achieving 3 pass marks on UCE papers, including maths and English. Age is age in years, female is a dummy for a female gendered student. No. of subjects is the number of papers taken at UCE. Control mean is the mean achieving 3 passes in the control group. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.

1

Table A20: Impact of treatment on University admittance

	(1) University passing grade	(2) University passing grade	(3) Admitted to University	(4) Admitted to University
treatment	0.04*	0.04	0.06*	0.00
	(0.02)	(0.04)	(0.03)	(0.05)
treatment* female		0.01		0.13*
		(0.05)		(0.07)
Mock score	0.15***	0.15***	0.08***	0.08***
	(0.01)	(0.01)	(0.02)	(0.02)
Age	-0.01	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.02)
Female	-0.01	-0.01	-0.07*	-0.13***
	(0.03)	(0.04)	(0.04)	(0.05)
stem	-0.24***	-0.24***	-0.12***	-0.12***
	(0.03)	(0.03)	(0.04)	(0.04)
Constant	0.82***	0.82***	0.29	0.30
	(0.29)	(0.29)	(0.33)	(0.34)
Observations	708	708	708	708
R-squared	0.32	0.32	0.14	0.14
Control mean	0.79		0.30	
Control mean female		0.82		0.25
Overall effect				
Female		0.05		0.13***
		(0.03)		(0.05)

Admitted to University refers to obtaining a space at a public University. All regressions include school fixed effects. Age is age in years. Female is a dummy for being a female student. Stem is a dummy capturing students in the science track. Mock score is the score in the mock exam prior to the intervention for that outcome. Control mean is the mean in the control group for heterogeneity group Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

A7 Persistence of effects

To examine if effects on exam performance tail off with time I use the natural variation in exam date depending on the subjects chosen at S6 to see if the treatment effects differ for students taking subjects closer or further away from the intervention[29]. Exams for different subjects at S6 level took place between 14th November and 29th November. Depending on which combinations of subjects student's chose, some students had their exams closer to the intervention than others, presenting natural variation I can exploit to see if treatment effect vary over time. Note the subject choice and exam timing were all determined before the movies were shown to students. I use this natural variation to see if treatment effects decline overtime. Results for the impact of treatment by whether exams were earlier or later are shown in Appendix Table A21. I do not find any significant difference in treatment effects for those with later versus earlier exams.

[29]This exercise is not possible for the S4 class since the screenings took place over only 5 days and all students had the same first exam

Table A21: Impact of treatment assignment on standardized test scores at S6 by students taking early exams

	(1) Over-all score	(2) Prin. score	(3) maths/ computer score	(4) Uni pass grade	(5) Over-all score	(6) Prin. score	(7) maths/ computer score	(8) Uni pass grade
Treatment	0.12*	0.09	0.07	0.08**	0.11	0.09	0.06	0.07**
	(0.07)	(0.07)	(0.08)	(0.03)	(0.07)	(0.07)	(0.08)	(0.03)
Treat*early first exam	0.04	0.09	-0.01	-0.06				
	(0.10)	(0.11)	(0.12)	(0.05)				
Early first exam	-0.01	-0.11	0.15	-0.02				
	(0.11)	(0.11)	(0.12)	(0.05)				
Treat*early average exams					0.05	0.08	0.04	-0.06
					(0.10)	(0.11)	(0.12)	(0.05)
Early average exams					0.03	0.04	0.07	0.02
					(0.08)	(0.08)	(0.09)	(0.04)
Constant	-0.12	-0.22	0.69	0.81***	-0.13	-0.24	0.68	0.80***
	(0.50)	(0.53)	(0.59)	(0.25)	(0.51)	(0.53)	(0.59)	(0.25)
Overall treatment effect								
Early first exam	0.15**	0.18**	0.07	0.01				
	(0.07)	(0.07)	(0.08)	(0.04)				
Early average exam					0.16**	0.17**	0.09	0.01
					(0.07)	(0.08)	(0.09)	(0.04)
Observations	708	708	708	708	708	708	708	708
R-squared	0.62	0.59	0.44	0.33	0.62	0.59	0.44	0.32

Overall score is the aggregate score achieved, prin. score the score in 3 principal papers, maths/computer the score on the compulsory maths or computer and general papers, uni pass grade that they got 2 passes on the principal papers. All scores are standardized. Early first exam means the first exam that student took was before the median first exam for all students. Early average exam means the average days since the intervention of all that student's exams was below the median for all students. I also carryout the same exercise using the number of days from treatment to first exam and mean days between treatment and all exams, but find no differences. Regressions include school fixed effects and individual controls (age, gender, mock score whether they were taking STEM subjects). Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

A8 School heterogeneity

Table A22: Heterogeneity in treatment effect for S4 by school characteristics

	(1) maths	(2) fail maths	(3) maths	(4) fail maths
Treatment	0.14**	-0.16***	0.12*	-0.16***
	(0.07)	(0.04)	(0.06)	(0.03)
	[0.07]*	[0.00]**	[0.09]*	[0.00]***
Treatment * top 500	-0.07	0.11**		
	(0.11)	(0.05)		
	[0.72]	[0.09]*		
Top 500	0.28***	-0.37***		
	(0.11)	(0.05)		
Treatment * high fees			-0.04	0.16***
			(0.11)	(0.06)
			[0.99]	[0.01]***
High fees			0.26**	-0.40***
			(0.11)	(0.06)
Constant	0.82	0.42	0.82	0.43
	(0.59)	(0.29)	(0.59)	(0.29)
Overall treatment effect				
Top 500	0.07	-0.05		
	(0.08)	(0.04)		
High Fees			0.08	0.00
			(0.09)	(0.04)
Mean in control		0.13		0.05
Observations	729	729	729	729
R-squared	0.54	0.30	0.54	0.31

Maths is a standardized maths score. Fail maths is a dummy for whether a student got a fail in the maths exam. Top 500 refers to if the school is within the top 500 out of 3300 nationally ranked schools. High fees refers to if a school charges above the median of school fees in the sample. All regressions include school fixed effects and student individual characteristics (age, gender, mock score and number of subjects taken). Mean in control refers to the control mean of that sub-group. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sharpened q-values in square brackets.

Table A23: Heterogeneity in treatment effect for S6 by school characteristics

	(1) total	(2) prin	(3) MTC	(4) total	(5) prin	(6) MTC
Treatment	-0.03	0.03	0.02	0.12*	0.12	0.10
	(0.12)	(0.12)	(0.13)	(0.07)	(0.07)	(0.08)
	[0.99]	[0.99]	[0.99]	[0.30]	[0.30]	[0.30]
Treatment * top 200	0.21	0.13	0.13			
	(0.13)	(0.13)	(0.14)			
	[0.57]	[0.66]	[0.66]			
Top 200	0.91***	0.58***	1.33***			
	(0.15)	(0.16)	(0.16)			
Treatment high fees				0.05	0.03	0.06
				(0.10)	(0.11)	(0.11)
				[0.99]	[0.99]	[0.99]
High fees				0.43***	0.44***	1.36***
				(0.12)	(0.13)	(0.15)
Constant	-0.26	-0.26	0.17	0.28	-0.28	0.17
	(0.45)	(0.46)	(0.49)	(0.40)	(0.47)	(0.50)
Overall treatment effect						
Top 200	0.18***	0.16***	0.16**			
	(0.06)	(0.06)	(0.06)			
High fees				0.16**	0.15**	0.17**
				(0.07)	(0.07)	(0.07)
Observations	708	708	708	708	708	708
R-squared	0.62	0.59	0.42	0.62	0.59	0.42

Total is the aggregate score achieved, prin the score in 3 principal papers, MTC the score on the maths/computer and general papers. All scores are standardized. Top 200 refers to if the school is within the top 200 out of 1800 nationally ranked schools. High fees refers to if a school charges above the median of school fees in the sample. Regressions include school fixed effects and individual controls (age, gender, mock score and number of subjects taken). Mean in control refers to the control mean of that sub-group. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sharped q-values in square brackets.

A9 Robustness

Table A24: S4 main results robustness tests

	(1) Total score	(2) Core score	(3) Maths	(4) English
Treatment	-0.016	-0.019	0.116**	-0.006
Robust p-value	(0.563)	(0.570)	(0.022)**	(0.924)
Permutation p-value	(0.917)	(0.962)	(0.038)**	(0.493)
Observations	735	735	735	735
R-squared	0.310	0.293	0.172	0.249

Total score refers to standardised aggregate score across all subjects taken in the exam. Core score refers to standardised aggregate score in the 6 mandatory subjects at S4 level. Standardized test scores composed of subject standardized scores and renormalised. Regressions include school fixed effects and control for the mock score in that exam. Permutation p-value calculated using 10000 permutations.
*** p<0.01, ** p<0.05, * p<0.1

Table A25: S6 main results robustness tests

	(1) Overall score	(2) Principal score	(3) Maths/computer score
Treatment	0.169	0.160	0.132
Robust p-value	(0.024)**	(0.037)**	(0.016)**
Permutation p-value	(0.041)**	(0.046)**	(0.013)**
Observations	710	710	710
R-squared	0.196	0.150	0.288

Overall score refers to the aggregate score in the principal and maths/computer papers. Principal score refers to the standardised score on the 3 chosen subject papers. Maths/computer score refers to the standardised score on the two mandatory papers in maths or computer and a general paper. Standardized test scores composed of subject standardized scores and renormalised. Regressions include school fixed effects. Permutation p-value calculated using 10000 permutations. *** p<0.01, ** p<0.05, * p<0.1

Table A26: Multiple hypothesis test for failing core subjects at S4

	(1) Maths	(2) Eng	(3) Chem-istry	(4) Biol-ogy	(5) Physics	(6) His-tory	(7) Geogra-phy
Treatment	-0.12	-0.02	0.04	-0.00	0.01	0.01	0.01
p-value	(0.00)***	(0.38)	(0.11)	(0.94)	(0.59)	(0.79)	(0.69)
q-value	(0.001)***	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
control mean	0.27	0.11	0.38	0.35	0.49	0.16	0.12
Observa-tions	729	729	728	729	728	727	706
R-squared	0.30	0.26	0.51	0.57	0.51	0.36	0.35

Core subjects are taken by all students at S4 level. All regressions include school fixed effects and individual controls (age, gender, number of subjects taken and mock exam score). Q-values calculated using the 2 step procedure of Benjamini et al. (2006). Robust p-value in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A9.1 Negative effect on the control group

It is possible that rather than *Queen of Katwe* improving the scores of students, that the placebo movie worsened scores. I consider this unlikely for a number of reasons. Firstly, many of the results show treatment closing gender gaps between men and women by improving the scores of girls to the same level as boys. For the placebo movie having a negative effect to be driving my results, only placebo girls would have to be experiencing a negative impact, when there is no reason to think the placebo movie would have any differential effect by gender. Additionally, I find heterogeneity by those who were below the median mock at in the S4 class, and not in the sciences track in the S6 class which fits with my prior hypothesis of which groups the role model would impact the most. There is no reason for the placebo movie to have a negative effect only for these subgroups.

I also check whether on average the placebo group perform as well as other year groups at the same school. Detailed breakdown of school performance by the number of students achieving certain divisions is available online. Divisions are five categories of overall score, with for example division one requiring at least 5 subjects to score a credit (6 or above) and a passing grade in at least one other subject and a score in the 6 best subjects of at least 31 points. The division help easy comparison of the score distribution of different schools.

In Table A27 I compare the 2016 placebo group's UCE performance to the UCE performance of the same schools two years later. Looking at two years later helps reduce the risk of students who received the intervention dropping out of the 2016 class taking the exam the following year[30]. I see that the division breakdown is very similar between the 2016 placebo and the 2018 class, with in fact the placebo group being slightly more likely to score in divisions 2 and 3 over division 4. This provides evidence that the placebo movie did not negatively affect student exam performance.

Table A27: Comparison of 2016 placebo and 2018 UCE exam performance

Class	Div1	Div2	Div3	Div4	Div5	Total students
2016 placebo	85	100	71	77	11	344
	25%	29%	21%	22%	3%	100%
2018 class	237	261	172	225	32	927
	26%	28%	19%	24%	3%	100%

[30]Attrition was only 3.5% and at least some of these students would not ever take the exam.