Increasing students' aspirations: the impact of *Queen of Katwe* on students' educational attainment

Emma Riley *

Department of Economics, Manor Road Building, Oxford OX1 3UQ, UK

(email: emma.riley@economics.ox.ac.uk)

August 18, 2017

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 (classes S4 and S6) were individually randomised to see either an aspirational movie featuring a potential role model, *Queen of Katwe*, or to see a placebo movie. I find that treatment with the aspirational movie leads to a 0.11 standard deviation increase in maths performance for S4 students, with the effect coming from students being 11 percentage points less likely to fail the exam. This effect is being driven by the lowest ability and students at lower ranked schools. For S6 students, their total score on their exams increase by 0.13 standard deviations. This study highlights the power of a role model as a cost-effective way to improve secondary school students' educational attainment, particularly of the worst performing students.

*I would like to thank Richard Sedlmayr for his contribution to the initial development of this study and Rachel Proefke for her assitance with logistics in Uganda, particularly recruiting schools and transportation for the students. 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 Rose Kusingura for her assistance also with recruiting schools and logistics in Uganda. I would like to thank those present at the CSAE workshop, where I presented a draft of this work, and for their insightful comments. 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. An amendment to the pre-analysis plan was uploaded there on the 18th July 2017 before additional data collection was finished and additional analysis begun. All analysis in this paper follows these pre-analysis plans unless clearly stated otherwise.

1 Introduction

Education is a key way to escape from poverty. In Uganda, those finishing secondary education with good grades are set to earn 78% more than those with just primary education, and those with university education earn 170% more (Kavuma, 2014). However, currently only one-third of boys and one-quarter of girls continue from lower to upper secondary education and only half of those completing secondary school continue to university. It is not just the years of education which matter for future earnings, but the quality and amount students actually learn (Hanushek et al., 2014). Finding ways to help students maximise their educational investment by increasing their attainment, along with helping them remain in education for longer, is therefore a key component of poverty alleviation.

A wide range of different types of intervention have been tried to improve learning outcomes in developing countries (Evans and Yuan, 2017), focusing on problems ranging from credit constraints (Baird et al., 2016) to poor teacher incentives (Glewwe et al., 2010). An alternative possible reason for low educational investment that I focus on here is low aspirations. People's aspirations and sense of control over their lives may be beaten down over time by the very difficult nature of living in poverty (The World Bank, 2015), making low aspirations a particular problem for the poor. Low aspirations have been linked to low economic investments, such as in education, to try and bring about a more prosperous future, and hence can lead to a poverty trap (Dalton et al., 2014). Poor people may have no reference of other people making successful investments, which can trap entire communities in a low investment, low aspirations poverty trap (Genicot and Ray, 2017). Raising aspirations and changing mindset have been shown to help overcome the negative impact of poverty on educational attainment (Claro et al., 2016).

One method that's been shown to raise aspirations is through a role model. Role models affect aspirations by demonstrating positive psychological behaviours such as self-belief and hard work (Bernard et al., 2014) and lead to an updating of beliefs about what can be achieved by people from a similar background (Nguyen, 2008). Role models have been shown to have large effects on economic behaviours and change social norms (Beaman et al., 2012; La Ferrara et al., 2012; Chong and Ferrara, 2009).

In this paper, I examine a low cost intervention featuring a role model to address low aspirations and raise educational attainment. Specifically, I examine whether a movie featuring a potential role model can improve exam performance. I do this through the randomised exposure of 1500 secondary school students in Uganda to a treatment movie, *Queen of Katwe*, featuring a potential role model, versus a placebo movie. Students preparing to take their national exams at the end of lower and upper secondary school were individually randomised to see the treatment or placebo movie, allowing me to test the impact of the role model on academic performance.

The treatment movie, *Queen of Katwe*, is base on the true story of a teenage girl from the slums of 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 getting into the top school in Uganda in order to play chess. She may therefore act as a role model about the importance of education and working hard to achieve your dreams, and raise aspirations more broadly through her demonstration of behaviours such as goal setting and fighting to achieve your dreams in the face of obstacles. The use of a placebo movie, here *Miss Peregrine's Home for Peculiar Children*, allowed me to exclude any beneficial effect to aspirations and exam performance from the novelty of going to the cinema and media exposure in general (Bernard et al., 2015).

I find that amongst students taking qualifying 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). I find no effects on a student's aggregate exam score or English score. 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 11 percentage point 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.

When examining these findings for heterogeneity, I find that it's female students who benefit most from treatment in terms of their maths score and probability of failing maths, students below the median age of 17 and students taking fewer subjects than the median. When looking at prior ability as measured by a mock exam taken before the study began, the entire benefit from seeing the treatment movie is experienced by low ability students. Students whose scores in the mock exam were below the mean increase their maths scores by 0.28 standard deviations when exposed to the treatment movie and decrease their probability of failing maths by 0.27 percentage points. This is a from a mean of 54% of students in the control group with below median scores in the mock failing the maths exam, and so a 50% decrease in the probability of failing maths amongst this group. Students who scored above the mean in their mock exam experience no effects on their maths scores or probability of failing maths from seeing the treatment movie.

I also examine heterogeneity by school characteristics by the ranking of the school and by whether the school charges fees above the median for my sample. I find it is lower ranked schools and schools charging lower fees where students taking the lower secondary school exam benefit the most from watching the *Queen of Katwe*. This suggests it is the worst students at poorly performing schools who benefit most from treatment.

Amongst students taking their finishing exams from upper secondary school, I find an overall improvement in their performance of 0.13 standard deviations. This effect is coming from improvements in their chosen subject papers. Students are also 6 percentage points more likely to get a place at public university. Again, it is women who benefit from seeing the treatment movie and see improvement in their overall exam scores of 0.20 standard deviations. At the higher level it is students at the best performing schools charging higher fees who see the most benefit from seeing the treatment movie.

This paper shows that behavioural change is possible after a brief (2 hour) exposure to a role model, and impacts on exam outcomes are seen even as soon as 1 week after exposure. This complements work which has looked at the impact of media exposure to role models and found large behavioural change over time. Bernard et al. (2014), in Ethiopia, invited people to watch documentaries about how people from similar backgrounds to them had improved their socio-economic position. Six months later, the treated group had higher aspirations and a stronger sense that they were able to control their own lives. They also displayed behavioural changes: they saved more, took out more loans, and increased school enrolment of their children. Effects are persistent 2 years later.

There is also other, non-experimental evidence from developing countries that exposure to the lives of alternative role models through TV, who rural individuals might not have encountered in their ordinary experience, can result in major shifts in behaviour. La Ferrara et al. (2012) show that exposure to TV shows with strong female role models and smaller families in Brazil reduces fertility and increases divorce. Jensen and Oster (2009) show that exposure to soap operas depicting urban women reduces fertility and domestic violence and alters beliefs about women's autonomy in rural India. Role model cartoons developed by UNICEF have become household names in South Asia and South America, and have been shown to empower girls through their behaviour (Chesterton, 2004). Role model exposure through media therefore facilitates an updating of beliefs and shifts in norms. My study adds to this by showing that the media exposure to the role model can even be as brief as a 2 hour movie and still lead to behavioural change.

There have also been studies looking at the impact of local successful people and their ability to affect those exposed to them. Beaman et al. (2012) find that, in Indian villages where girls had female role models on the village council (because the village was randomly assigned to reserve a seat on the village council for a woman) the gender gap in occupational aspirations declined among the girls themselves and among their parents. This also altered behaviour: the gender gap in adolescent educational attainment disappeared and girls spent less time on household chores. This shows the power of real life role models over a prolonged period to affect behaviour.

Role models have also been examined in the context of education in developing counties, with a view to changing beliefs about the returns to education and likelihood of someone from a poor background achieving those returns. In Madagascar, Nguyen (2008) used a randomised experiment to compare giving information about schooling returns to exposure to a role model from either a rich or poor background. She finds 0.17 standard deviation 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 students. The effect is even larger for the poorest students, improving test scores by 0.27 standard deviations. This suggests role models can be a powerful tool, particularly for the poorest, by changing beliefs about both the returns to education and the probability of success. My study complements these by showing that the role model does not have to be available in real life to inspire and have a positive effect on students.

The importance of aspirations for education attainment is an area that has traditionally been overlooked in developing countries, particularly compared to that in developed countries. Wydick et al. (2013) is one study looking at the role of aspirations for children and for later life outcomes in a developing country context. They find that children sponsored through Compassion International had improved adult outcomes, with the sponsored children exhibiting significantly higher levels of self-esteem, aspirations and self-expectations, and lower levels of hopelessness. Serneels and Dercon (2014) also show that maternal aspirations are an important factor in determining child educational outcomes, including grade achieved and verbal and maths test scores.

There is extensive evidence for developed countries, especially from psychology, that exposure to role models improves aspirations, particularly among young adults. Stout et al. (2011) find improvements in self-efficacy, career aspirations and effort in science subjects among female calculus students after they are exposed to photographs and videos of female role models in science. Dasgupta and Asgari (2004) show the power of role models in overcoming stereotypes relating to academic achievement, and that exposure to role models can change beliefs about what is possible. Dennehy and Dasgupta (2017) show that female mentors increase female students' feelings of belonging in engineering, their retention and their aspirations for pursuing postgraduate engineering study. Male mentors didn't have these effects. My study therefore adds to this large literature but in a developing country context, with a large sample size and in the field on an important educational outcome.

This intervention also showed that substantial impacts can be had on exam performance even when the intervention is as short as 1 week before the exam. Over such a short time span, there is limited opportunity for increased study effort to affect exam performance and so effects relating to motivation during the exam are likely to dominate. The size of effect seen in this paper is of a similar magnitude to that seen in experiments which offer to pay students for performance immediately before an exam, thus removing all effects from increased studying, which found 0.12-0.22 standard deviation effects, that are most pronounced for maths (Levitt et al., 2011). Again this highlights that the psychology of how motivated the students feel on the day of the exam can be as crucial as the amount of preparation they do beforehand for their exam performance. This holds especially for maths, which has generally been found to be more elastic than other subjects, where students can improve their scores simply by trying harder and more persistently on a problem (Bettinger, 2010).

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 therefore 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 in Kenya (Glewwe et al., 2010).

The rest of this paper is organised as follows: Section 2 discusses the interventions and study design. Section 3 goes over the data used in this study. Section 4 contains the empirical specification and results. Section 5 discusses the cost effectiveness, results and policy implications of the findings and section 6 concludes.

2 Intervention and Study Design

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 aspirational 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 Treatment movie

Social psychologists have long noted that aspirations, motivation, goal setting and self-efficacy are based on observing others in the immediate environment (Bandura, 1977b,a). Ray (2006) argues that individuals form their aspirations by observing individuals they can identity with and whose behaviour they can observe. The protagonist of *Queen of Katwe* may act as that individual and so impact aspirations.

The protagonist has many characteristics which have been shown in a large psychology literature to matter in a role model. Firstly, Phiona is similar in multiple dimensions to many of the students and hence relevant to them and easy for them to relate to (Lockwood and Kunda, 1997). She is the same age (a teenager) and from the same country and even city as the students in this study (Kampala, Uganda). She is also from a similar or poorer background to most of them, important for creating realistic aspirations (Genicot and Ray, 2017; Ray, 2006). Secondly she is the same gender as half the students, a woman, and research has shown that women require same-sex role models in order to overcome negative stereotypes about their ability, whereas men respond equally well to role models of either gender (Lockwood, 2006). This means that Phiona will act as a potential role model to both male and female students. Additionally, Phiona is a counter-stereotype in that she is a woman doing well at what is traditionally a male dominated game, chess (Dasgupta and Asgari, 2004). The very fact of exposure to a count-stereo type alone has been shown to change attitudes and "inoculate" those exposed against applying stereotypes to themselves (Stout et al., 2011). Phiona therefore displays all the characteristics required in this context of a relevant role model who is likely to appeal to the students and have an effect on their behaviour.

Phiona goes from nothing, living in the slums and selling corn to passing drivers, to getting into the top school in Kampala, succeeding at chess and achieving her dreams. She does this while overcoming numerous difficulties along the way, all through hard work and perseverance. She may therefore act as a constructive role model, encouraging similar character traits in the students and raise aspirations amongst students who wish to emulate her example.

She displays a number of different positive behaviours which students might gain from imitating. These include: perseverance and hard work; over-coming hardship; shaping her own life (Rotter, 1966); a growth mindset (Dweck, 2000); goal setting; achieving long term goals through small incremental steps (Locke and Latham, 2002); and reaching out to others for help. Watching the treatment movie may therefore both encourage these important behaviours needed to succeed in both education and life and raise aspirations by changing beliefs over what its possible to achieve even from a poor background.

2.2 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 aspirational or motivational 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 allows me to remove any effect from simply the activity of attending the cinema and instead ascribe any effects to seeing the treatment movie in particular.

The placebo movie was chosen carefully to be appealing to this age group. It was important the movie was entertaining and suitable for the students, containing characters of a similar age but without a Ugandan background. The content was purely an adventure story focused on overcoming monsters threatening the characters. There was no educational or strongly motivation content.

2.3 Sample

Secondary schools were approached during August and 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 who they had previously worked with and asked if their students sitting national exams in 2016 (the S4 and S6 classes) would like to participate in the study. 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. Schools signed consent forms for the students to be transported to and attend the cinema, and agreed to provide student lists and exam results once they became available.

Schools were recruited until a sample size of approximately 1600 students was reached. The students were equally split between male and female and the S4 and S6 classes taking national exams.

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 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. If there were less than 200 student attending the screening the two small screens would be used, if between 200 and 300 students one screen of 100 and one of 300 would be used and for more than 300 students both screens of 100 and the 300 person screen would be used.

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. 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. The chosen cinema was one of two multi-screen cinemas in Kampala which allowed us space to conduct a randomisation and complete control over the movie schedule and times.

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 for a screening. This was done by an assistant picking a ticket out of a bag without looking and handing 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.

After getting a ticket, students were steered to the designated registration desk for that movie, where their ticket was checked and they registered their name, school, age and gender before proceeding into the theatre. These registration lists were later combined with lists from the schools of student index numbers, which uniquely identify student exam results. Once a ticket had been selected, students with tickets for different movies were kept separate the entire time. I am therefore confident that all students saw the movie they were assigned to. The students also had between 2 and 5 teachers accompanying them depending on the class size. These teachers were split between the theatres randomly.

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. For example, if a school had 250 students then 150 would have to see one movie and 100 the other. This was randomised and balanced over different sessions so that overall we issued 825 treatment movie tickets and 727 placebo movie tickets to students in classes taking national exams. School fixed effects will be used to control for this difference in treatment probability within a school.

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. Looking at Table 1, 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. The standardised mock score was approximately zero in both the treatment and control groups, as would be expected from a standardised score, and not different between them. At S6 level, Table 2 shows that 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. Again, the standardised mock scores were approximately zero and not significantly different between the treatment and control groups.

Table 1: Balance test S4 class

	Placebo		Treatment			
	mean	sd	mean	sd	difference	p-value
Age	17.28	1.25	17.25	1.23	0.03	(0.76)
Female	0.51	0.50	0.51	0.50	0.00	(0.61)
Number of subjects	9.73	0.62	9.68	0.60	0.04	(0.34)
Mock total score	0.01	0.98	-0.01	1.01	0.02	(0.74)
Observations	344		391		735	

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.

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. Missing results meant either that the index number obtained for that student was incorrect or that the student didn't take the exam. All cases of no results for an index number were double checked with the

Table 2: Balance test S6 class

	Placebo		Treatment			
	mean	sd	mean	sd	difference	p-value
Age	19.09	1.24	19.00	1.13	0.09	(0.31)
Female	0.47	0.50	0.50	0.50	-0.03	(0.40)
STEM	0.33	0.47	0.30	0.46	0.02	(0.53)
Mock total score	-0.02	0.97	0.04	1.01	-0.06	(0.45)
Observations	341		370		711	

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.

school, with remaining cases due to students not taking the exam. Attrition was balanced across the treatment and control groups at the 10% level, as shown in Table 3 below. 21 students in the placebo and 33 in the treatment group did not take their national exams, 3.6% of the sample. Attrition varied greatly by school, with some of the schools in particular having very few candidates

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. Attrition means the students didn't take their national exams.

at S6 level taking the exams and many of these students deciding to not actually take the exam. I examined whether student or school characteristics were correlated with attrition in Table 4. Students at Christian schools are more likely to take the exam, as are older students and students in the S4 class.

Table 4: Individual and school characteristics correlated with attrition

	(1)
	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

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

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

3 Data

3.1 Student data

Limited information about the students was collected upon registration at the cinema. This was their name, age, gender, class (S4 or S6) and school, along with which movie they saw. This data was combined with lists of exam enrolment provided by each school. The exam enrolment information listed the student's name, age, index number and subjects entered for exams. The index number was particularly important as it is a unique identifier of a student's exam results. The registration and school exam enrolment data were combined using double data entry with any discrepancies checked. This resulted in a data set of 1500 students who saw a movie and were due to take a national exam.

Mock exam results from a practice for the national exam administered by the schools in the summer before the study began were also obtained for all students in the study. This data is described in the test score data section.

3.2 School data

Information was collected about the schools at the same time as exam results data was collected. This involved asking the schools about their religious affiliation, fees and whether they offered boarding. Publicly available rankings of the schools were also collected. This information is shown in Table 5.

Schools all had a strong religious affiliation as either Christian or Islamic. This determined the kind of prayers they took part in during the school day and was a strong part of the school's character. Nearly all the schools had some boarding element, with only one, Kulumba, having none. This is very common in secondary schools in Uganda, where students might come from very far away and transport is difficult. Note, all the schools were private schools, as two-third of secondary schools in Uganda are (MoES Uganda, 2015).

National rankings of the schools were obtained from the Ministry of Education. At the advanced examination taken by S6 students, UACE, there are 1900 schools ranked of which the schools in my sample varied markedly, from near the top to one of the very worst. Likewise, at the UCE level taken by S4 students, the schools are also spread out in the ranking out of 3300, though not so disburse as for UACE. The fees the schools charge for boarding and day students in the S4 and S6 classes were also collected from schools and display a wide spread, with the most expensive boarding school 900,000 USH a year, or $257 at current exchange rates, while the least expensive is half that at only 440,000 USH or $125 a year.

The schools also differed in which classes they provided to attend the cinema. We offered to take both the S4 and S6 classes if they wanted. For mainly timetabling reasons and staff constraints, some schools only offered one class. The schools are also dramatically different sizes, with the largest having nearly 200 S6 students and the smallest only five.

3.3 Test score data

The primary outcomes in this paper are standardised exam scores on national exams. Secondary school students sit their national exams in October and November and the results are released in January and February of the following year. Ordinary exams are taken after 4 years of secondary school by the S4 class, the Uganda Certificate of Education (UCE), and began on the 19th October, 1 week after the last movie screening. Advanced level exams are taken after a further 2 years of study by the S6 class, the Uganda Advanced Certificate of Education (UACE), and began on the 14th November, 1 month after the last screening. 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.

Data on national exam results was collected in February 2017 once the exam data sat by both the S4 and S6 classes had been released. The data was collected in two ways. Initially, 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. The Ugandan National Exam Board allows you to text in an index number to obtain results for that student. Results obtained in this way provide an equal amount of information as those provided to the schools. Results were collected via the text interface for all remaining students that results were missing for and entered into Excel. A random sample of results obtained via text-messaging were audited to ensure they had been entered correctly.

Mock exam results data was collected for all the students in the study. This data was provided by the schools. All students sit a mock exam during the summer before their national exam. This corresponded to August 2016 for the study sample, two months before treatment took place and one month before schools were approached about taking part in the study. This mock exam is administered by the schools and based on previous exams. Students in the S4 class sit mock exams in English and Maths only. Students in the S6 class sit the mock exam in the principal and subsidiary subjects they are registered for in the national exam. Schools were requested to provide

*All the schools signed consent forms agreeing to provide exam results as part of being in the research 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.

Table 5: School Characteristics

			Ranking		Fees				Class size	
School	Religion	Boarding	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.

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.

All these outcomes have been pre-specified in the pre-analysis plan unless explicitly stated as not in the pre-analysis plan.

3.3.1 Standard 4 Exam

After 4 years of secondary education candidates take the UCE exam. 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. 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 90.

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.

For students taking UCE exams the following outcomes are examined:

1. **Exam score aggregate**: aggregate score composed of exam scores across all eight-ten subjects taken by a student

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

3. **Individual subject grade**: Standardised score achieved in Maths and English subjects

Effects of treatment are more likely to be expected on subjects related to chess, of which maths has the clearest link. I therefore examine the particular effect of the treatment on the maths exam outcome, and also look at English individually since it is a common outcome examined in the education literature.

3.3.2 Standard 6 Exam

In their final year of secondary school, students sit the UACE exam. The UACE is taken in five subjects, three of which are from a list of principle subjects, one in a subsidiary subject out of mathematics or computer and one in a general 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. Any student achieving a 6 or below on a subsidiary paper or the general paper gets one point. The principal papers are marked on a A, B, C scale, with an A earning 6 points, a B 5 points etc. The maximum of 2 points earned on the subsidiary and general paper are added to the points earned on the principal papers. This means the highest total score a subject could earn is three As and passes on the subsidiary and general paper, giving 20 points.

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.

For students taking UACE exams, the following outcomes were examined:

1. **Total exam score**: aggregate exam score composed of exam scores across all principal and subsidiary subjects taken by a student, with subsidiary subjects scoring a maximum of 1 point.

2. **Principal score**: aggregate score in the principal papers only.

3. **General paper and subsidiary paper score**: standardised score on the general paper and subsidiary paper in maths or computer taken by all students. This will be an inverted scale of the 1-9 score on these papers.

An additional outcome examined is a dummy variable for whether a student achieves the grades to get into public university. Public University in Uganda is the best type of tertiary education and the grades required are set nationally. The requirement is passing grades in two principal subjects, where a pass is any score greater than 0. I therefore construct a dummy variable equal to one if a student got at least two passes in their principal subjects and zero otherwise. 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.

3.4 University place

In an amendment to the original pre-analysis plan two further outcomes were specified before data was collected on them:

1. An indicator for whether the student obtained a government scholarship

2. An indicator for whether the student gained entry to Public University

These outcomes were obtained from the Ugandan National Council of Higher Education who hold records on all public University entry and determine scholarship awards. These records are publicly available and include identifiable information for the students, such as name, id number and school, which I used to match this data to my study sample.

4 Empirical Strategy and Results

4.1 Empirical strategy

To examine the effect of the treatment on exam outcomes, I run 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. If available, the mock result in the specific subject outcome will be controlled for. 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.

Specification 1 is the basic specification used here, as set out in the pre-analysis plan. Any departures from the contents of that plan will be clearly stated.

The parameter of interest is β_1, the average treatment effects of the *Queen of Katwe* movie on an exam outcome. The school fixed effects, θ_s, are included to account for differential treatment probability depending on which movie was played on the larger cinema screen. They also control for substantial school heterogeneity (and so improve precision), as seen in Table 5 in the large dispersion of rankings of the schools. Robust standard errors are calculated to allow for heteroskedasticity.

Individual characteristics, $\boldsymbol{x_i'}$, are included to improve precision. These are:

1. dummy for whether the student is female
2. the age of the student in years
3. the number of subjects taken (for S4 students)
4. whether the student choose to take any subjects in maths or science (STEM subjects) at S6 level

4.2 Main Results

4.2.1 S4 Class

Table 6 shows the impact of assignment to see the treatment movie on the S4 exam outcomes defined in section 3.3. 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 when controls are included, and 0.14 standard deviations still significant at the 5% level without any controls. This is a large positive effect on the maths exam outcome, and is examined in more detail below.

Table 6: Impact of treatment assignment on S4 standardized test scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Total score	Total score	Core score	Core score	Maths	Maths	English	English
Treatment	0.01	-0.02	0.00	-0.03	0.14**	0.11**	-0.04	-0.06
	(0.07)	(0.03)	(0.07)	(0.03)	(0.07)	(0.05)	(0.07)	(0.05)
Age		0.00		-0.00		-0.03		-0.03*
		(0.01)		(0.01)		(0.02)		(0.02)
Female		-0.01		-0.01		-0.03		0.10**
		(0.03)		(0.03)		(0.05)		(0.05)
No. subjects		-0.06**		-0.08**		-0.16***		-0.05
		(0.03)		(0.03)		(0.05)		(0.05)
Mock score		0.99***		0.95***		0.80***		0.77***
		(0.02)		(0.02)		(0.03)		(0.03)
Constant	-0.67***	0.32	-0.66***	0.61	-0.35***	1.70***	-0.53***	0.60
	(0.09)	(0.36)	(0.12)	(0.43)	(0.13)	(0.65)	(0.12)	(0.60)
Observations	735	729	735	729	735	729	735	729
R-squared	0.31	0.88	0.29	0.82	0.17	0.57	0.25	0.62

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.

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Mock exam performance is a strong predictor of national exam score, with each additional standard deviation scored in the mock associated with a 0.99 standard deviation increase in total score. I will examine heterogeneity by mock exam performance later.

To examine the effect of treatment on the maths exam performance in more detail, I break down the maths exam into dummies by grade achieved. As mentioned, the exam is graded from 0, fail, to 8, the maximum result. I look at the impact of treatment on a dummy for obtaining each grade in Table 7. In column (1) it can clearly be seen that seeing the treatment movie reduces the probability that a student obtains the bottom, failing, grade in maths by 11 percentage points from a control group mean of 27%. This is a 40% decrease in the probability of failing maths. Seeing the treatment movie increases the probability by 5 percentage points a student scores 2 or 3 on the maths test, suggesting that seeing the treatment movie might be pushing students to the next couple of grades above what they would have achieved, though this is only significant for grade 3 at the 10% level. No impact is seen for higher scores, and in fact less than 1% of students achieve the top grade at all in this sample. I find no effect of treatment on the probability of failing any other core subject (see Robustness section, Table 20).

Table 7: Impact of treatment on probability of getting each maths grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Fail	1	2	3	4	5	6	7	Top
Treatment	-0.11***	0.05	0.05*	-0.01	0.01	0.01	0.00	-0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Age	0.01	0.01	-0.02**	0.00	-0.00	0.00	-0.01	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Female	0.03	-0.05*	0.02	0.02	0.02	-0.04*	-0.00	0.01	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
No. Subjects	0.01	0.07**	0.01	-0.02	-0.04	-0.01	-0.02	-0.02	0.01*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)
Mock score	-0.17***	-0.11***	-0.04***	0.02	0.05***	0.09***	0.09***	0.06***	0.03***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Constant	-0.04	-0.67*	0.50	0.33	0.46*	0.05	0.29	0.16	-0.07
	(0.35)	(0.38)	(0.33)	(0.33)	(0.27)	(0.27)	(0.21)	(0.15)	(0.05)
Mean in control	0.27	0.17	0.13	0.17	0.10	0.08	0.05	0.03	0.01
Observations	729	729	729	729	729	729	729	729	729
R-squared	0.31	0.09	0.02	0.06	0.06	0.10	0.10	0.08	0.05

Maths papers are graded from fail (0) to highest grade (8). All regressions include school fixed effects. Mean in control is the mean proportion of the control group obtaining that grade
Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

I also look at the effect of treatment on the probability of getting each maths grade using an ordered logit regression to improve power. 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 are shown in Table 8. 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 3 scores, particularly the probability of getting the failing score declines by 5% percentage points. The treatment movie also increases the probability of obtaining higher grades, with the effect significant at at least the 10% level in all cases, though the magnitudes are small. The largest effect is seen on grades 6 and 7 where seeing the treatment movie increases the probability of obtaining that grade by 1.4 percentage points. I also do the same ordered logit for the other subjects in the core exams. These are shown in the Appendix in Table 21. I find no effect of seeing the treatment movie on the probability of getting any particular grade.

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 Figure 1. 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 0.25, 0.25 and 0.28 respectively. Hence I cannot reject equality of the distributions. However, in 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. Now the p-value for the Kolmogorav test is 0.008, so I can reject equality of the distributions at the 1% significance level and confirm that treated students achieve higher maths scores.

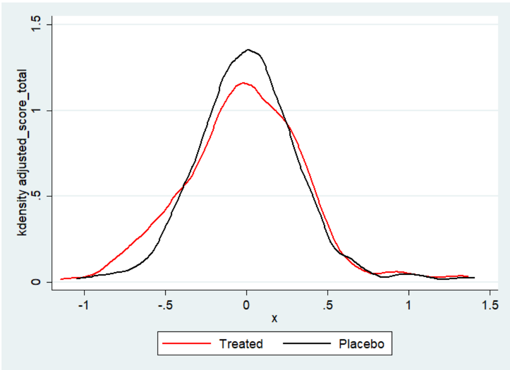
Table 8: Ordered logit regression of the impact of treatment on maths grade at S4

Grade	treatment
Fail	-0.048***
	(0.015)
1	-0.013***
	(0.004)
2	-0.000
	(0.001)
3	0.009***
	(0.003)
4	0.010***
	(0.003)
5	0.014***
	(0.005)
6	0.014***
	(0.005)
7	0.009***
	(0.003)
8	0.004***
	(0.002)
Observations	729

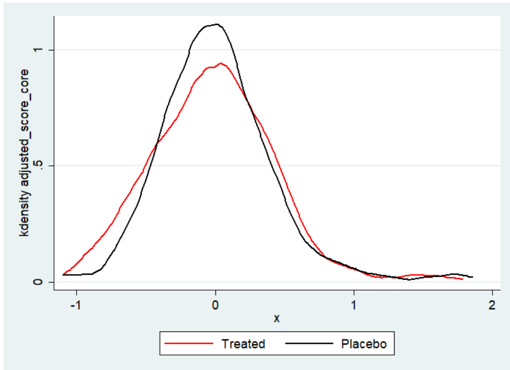
Maths papers are graded from fail (0) to highest grade (8). Regressions include school fixed effects and individual control variables (age, gender, number of subjects taken and standardised mock score).
Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 1: Histograms of S4 student results by treatment assignment

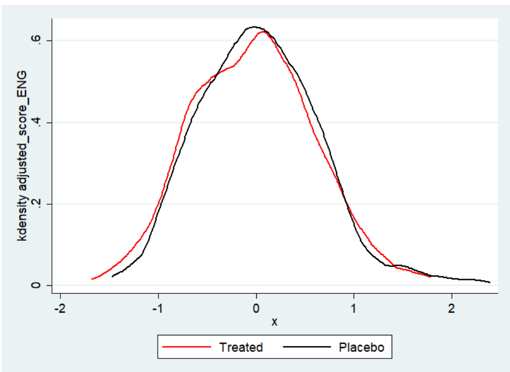
(a) Total score



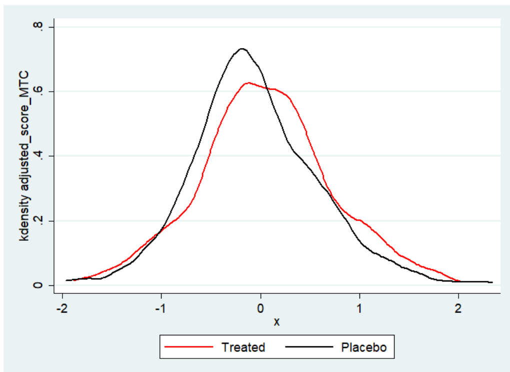
(b) Core score



(c) English



(d) Maths



4.2.2 S6 class

Table 9 shows the impact of assignment to see the treatment movie on the S6 exam outcomes defined in section 3.3. Results are shown both without and with individual control variables, but I will discuss only the results with control variables for brevity. Seeing the treatment movie results in the overall exam score being 0.13 standard deviations higher, a large effect on an education outcome. This is significant at the 1% level. The score on the principal exam papers increases by 0.13 standard deviations, significant at the 5% level. There is no effect on the score achieved on the general and subsidiary papers.

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. Students taking STEM subjects do significantly better, possibly because only the best students take STEM subjects. A 1 standard deviation better performance on the mock is associated with a 0.76 standard deviation better performance on the overall and principal papers, but only a 0.28 standard deviation improvement on the subsidiary subjects.

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 on a dummy for each decile of overall score and by examining histograms of adjusted scores by treatment assignment. Firstly, the impact on decline of overall score, shown in Table 10. I include control variables but the results do not change without them. Treatment has no effect on the probability that a student's final score is in a particular decile except for the very top decile shown in column (9). Treatment results in an increase of 3 percentage points in the probability the student scores in the top decile, though this is only significant at the 10% level.

Again, I 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 Figure 2. The histograms all show a shift to the right for students assigned to treatment around the middle of the distribution. There is also an effect at the top of the distribution for principal subjects, with the top tail of high scores extending further for treated students. This supports the results found in the decile analysis that its the top students who benefit most from seeing the treatment movie and perhaps indicates there are also some positive effects around the mean of the distribution that I am not powered to detect.

I again look at the Kolmogorov test for equality of the distributions. For the overall score, the p-value is 0.046 so I reject equality. The treatment shifts the distribution to the right. For principal papers, the Kolmogorov p-value is 0.073, so I can just reject equality at the 10% level. For the subsidiary paper I get a p-value of only 0.041, so I can also reject equality of the distributions here. This suggests that while I cannot detect differences in the mean subsidiary outcome by treatment

Table 9: Impact of treatment on S6 standardized test scores

	(1) Overall score	(2) Overall score	(3) Principal subjects	(4) Principal subjects	(5) Subsidiary subjects	(6) Subsidiary subjects
Treatment	0.17**	0.13***	0.17**	0.13**	0.08	0.07
	(0.07)	(0.05)	(0.08)	(0.05)	(0.06)	(0.06)
Age		-0.03		-0.02		-0.04
		(0.02)		(0.02)		(0.03)
Female		0.05		0.10*		-0.09
		(0.05)		(0.05)		(0.06)
STEM		0.55***		0.40***		0.79***
		(0.06)		(0.06)		(0.07)
Mock score		0.76***		0.77***		0.28***
		(0.03)		(0.03)		(0.03)
Constant	-1.50***	-0.13	-1.48***	-0.25	-0.36	0.69
	(0.42)	(0.43)	(0.41)	(0.43)	(0.35)	(0.57)
Observations	711	708	711	708	711	708
R-squared	0.20	0.62	0.15	0.59	0.29	0.44

Overall score refers to the aggregate score in the principal and subsidiary papers. Principal subjects refers to the standardised score on the 3 chosen subject papers. Subsidiary subjects refers to the standardised score on the two mandatory subsidiary papers. Standardized test scores composed of subject standardized scores and renormalised. All regressions include school fixed effects.

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

status there may be differences elsewhere in the distribution, suggesting heterogeneity is important here.

In Table 11 I report whether the student achieved the required 2 passes in principal papers to get into public university and whether they obtained a place at public university or not [†]. Students who saw the treatment movie were 4 percentage points more likely to get the necessary grades and 6 percentage points more likely to get a place at University. However both these results are only just significant at the 10% level and so should only be taken as indicative. This is from a mean of 79% getting the required grades to get into university and 31% being offered a place at University

[†] Whether a student obtained the grades to get into University was not included in the original pre-analysis plan. Whether a student obtained a place at University was included in the pre-analysis plan amendment

Table 10: Deciles of overall score at S6

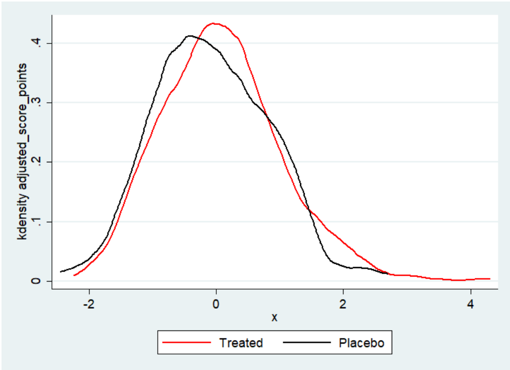
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	decile 1	decile 2	decile 3	decile 4	decile 6	decile 7	decile 8	decile 9	decile 10
Treatment	-0.02	-0.01	-0.02	0.03	-0.00	-0.00	0.01	-0.01	0.03*
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Age	0.01	-0.01	0.00	0.01	0.01	-0.01	0.01	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.01	0.00	-0.04	0.04	-0.01	-0.00	-0.02	0.03	0.01
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
STEM	0.11***	-0.01	-0.00	-0.03	-0.04	0.05	-0.05**	-0.03	-0.01
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Mock score	-0.13***	-0.08***	-0.05***	-0.05***	0.01	0.08***	0.04***	0.07***	0.11***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Constant	0.35	0.11	-0.07	-0.16	0.13	0.34	-0.17	0.30	0.16
	(0.23)	(0.22)	(0.22)	(0.31)	(0.24)	(0.28)	(0.19)	(0.21)	(0.18)
Control mean	0.15	0.10	0.10	0.18	0.10	0.16	0.07	0.09	0.06
Observations	708	708	708	708	708	708	708	708	708
R-squared	0.25	0.07	0.04	0.04	0.02	0.06	0.05	0.13	0.24

Decile 1 is worst and decile 10 best. Overall score refers to the aggregate score in the principal and subsidiary papers. STEM refers to taking a principal paper in maths or science. Robust standard errors in parentheses. Regressions include school fixed effects.

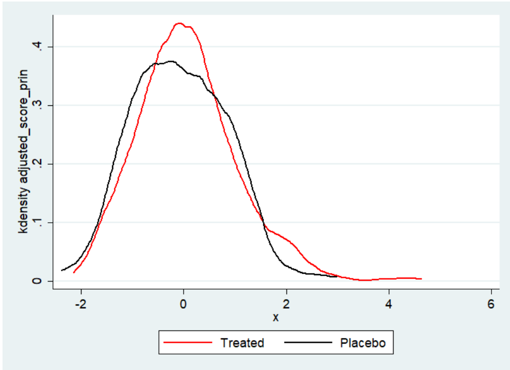
*** p<0.01, ** p<0.05, * p<0.1

Figure 2: Histograms of S6 student results by treatment assignment

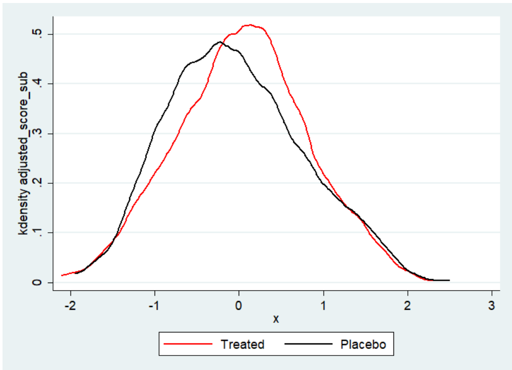
(a) Total score



(b) Principal papers



(c) Subsidiary/generate papers



in the control group. Seeing the treatment movie therefore increases by 20% the odds that a student will get a place at University. This result shows that not only has seeing the treatment movie improved exam outcomes but that there will be long term effects from students seeing the treatment movie being more likely to get a place at university.

I also looked at whether seeing the treatment movie increase the odds that a student obtained a government scholarship. However I find no effects here on the tiny sample of 16 students who obtained scholarships from my study sample, and so I do not include the results.

Table 11: Impact of treatment on probability obtain scores to get into public university

	(1) University passing grade	(2) University passing grade	(3) Place at University	(4) Place at University
Treatment	0.06*	0.04*	0.06*	0.06*
	(0.03)	(0.02)	(0.03)	(0.03)
Age		-0.01		0.01
		(0.01)		(0.02)
Female		-0.01		-0.01
		(0.03)		(0.04)
STEM		-0.24***		-0.08*
		(0.03)		(0.04)
Mock score		0.15***		0.10***
		(0.01)		(0.02)
Constant	0.37	0.82***	0.16	0.18
	(0.22)	(0.29)	(0.20)	(0.34)
Mean in control	0.79		0.31	
Observations	711	708	711	708
R-squared	0.05	0.32	0.10	0.15

Regressions include school fixed effects. University passing grade refers to the minimum two principal passes to get into public University. Place at University refers to obtaining a space at a public University.

Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

4.3 Heterogeneity

Heterogeneous treatment effects across variables collected at treatment assignment are tested by augmenting equation 1 to include the variable and the interaction between treatment and that variable. This gives the following specification:

$$\begin{aligned} y_{is1} = &\beta_0 + \beta_1 \text{QofK}_i + \boldsymbol{x_i'} \cdot \gamma + y_{is0} \\ &(\boldsymbol{x_i'} \cdot \text{QofK}) \cdot \lambda + \theta_s + \epsilon_{is}, \end{aligned} \tag{2}$$

where i indexes student in school s, y_{is1} denotes the exam outcome of interest, y_{is0} is the standardised mock value of the same outcome variable[‡] 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 student characteristics, θ_s is a vector of school fixed effects and ϵ_{is} is a random error. The parameter of interest here is λ, the heterogeneous treatment effect of seeing *Queen of Katwe*.

4.3.1 Individual characteristics

The individual student variables I examine are:

1. An indicator equal to one if the respondent is female
2. An indicator variable equal to one if respondent's age is above the sample median for that grade.
3. An indicator variable equal to one if the respondent is taking fewer subjects than the median for that grade (at UCE)
4. An indicator variable equal to one if the student is taking at least on principal science subject (maths, physics, biology and chemistry) at UACE
5. An indicator equal to one if the student was below the median exam performance in their mock exams.
6. Indicators of decile of exam performance in the mock exam

The heterogeneity by gender would reveal whether the treatment movie, featuring a female protagonist, appealed more to women or equally to each gender, as research from psychology suggests it might (Lockwood, 2006). The heterogeneity by age would pick up whether older students, who might have been held back years or had to postpone their studies for a while, perform less well as a result but benefit more from seeing the treatment movie featuring a girl who had stopped school but returns to it. Reports from those familiar with the Ugandan education system suggests

[‡] if provided by the school, if not available the standardised mock total score will be used instead

students who are struggling often take fewer subjects at UCE in order to trade off quality versus quantity. Looking at heterogeneity by students taking fewer subjects than the median would reveal if the weakest students benefited most from the movie. At UACE students are encouraged to commit to a sciences or humanities subject track. Since the treatment movie was most concerned with what is regarded as a scientific game, looking at heterogeneity would reveal if the treatment movie had greater effects on students taking more scientific subjects who might more closely relate to the protagonist. Lastly heterogeneity by mock exam performance will reveal whether students at the bottom or top of the ability distribution benefited more from treatment.

Heterogeneous treatment effects on maths exam performance for the S4 class are shown in Table 12. 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. At S4 level, 50% of students are female, 35% are above the median age of 17, 28% are taking fewer than the median number of 10 subjects and 49% score below the median score on the mock exam.

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 for failing maths negative, neither are significant. However the overall effect of treatment for women is a positive and significant 0.17 standard deviations for maths score and a 14 percentage point decrease in the probability of failing maths. The simple treatment effect for maths score is no longer significant, and for failing maths is only a significant 9 percentage point decrease. This suggests that girls benefit more from seeing the treatment movie than boys.

Amongst older students (columns (3) and (4)), the point estimate on the interaction term for maths score is actually negative and for failing maths positive, implying treatment could worsen maths performance for older student. However again these are not significant. Overall, students older than the median have no benefit from seeing the treatment movie on their maths score but still experience a 9 percentage points decrease in the probability of failing maths.

There is a large point estimate for the interaction between taking fewer subjects than the median and treatment on the maths score, but it is not significant. The overall effect for those taking fewer subjects is a 0.25 standard deviation improvement in maths score, though this is only significant at the 10% level. The point estimate on the interaction for failing maths in column (6) is significant and negative, resulting in those students who have chosen to take fewer subjects being 19 percentage points less likely to fail maths after treatment compared to 9 percentage points less likely for those taking more subjects than the median. It could be the case that those taking less subjects find it easier to shift effort from one subject to another, resulting in treatment having larger effects.

Looking at students who scored below the median in their mock exam, there is a large and

Table 12: Heterogeneity in treatment effect for S4 by gender, age and number of subjects

Impact of treatment assignment on S4 standardized test scores								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	maths	fail maths	maths	fail maths	maths	fail maths	maths	fail maths
Treatment	0.12	-0.09**	0.17**	-0.13***	0.11	-0.09***	0.02	0.04
	(0.08)	(0.04)	(0.07)	(0.03)	(0.07)	(0.03)	(0.08)	(0.04)
Female	-0.10	0.06	-0.07	0.04	-0.07	0.04	-0.07	0.03
	(0.08)	(0.04)	(0.06)	(0.03)	(0.06)	(0.03)	(0.06)	(0.03)
Above median age	-0.09	0.02	-0.06	0.00	-0.09	0.02	-0.10	0.02
	(0.07)	(0.03)	(0.09)	(0.04)	(0.07)	(0.03)	(0.07)	(0.03)
Below median subject	0.10	0.02	0.10	0.02	0.02	0.07	0.11	0.01
	(0.07)	(0.03)	(0.07)	(0.03)	(0.10)	(0.05)	(0.07)	(0.03)
Below median mock	-1.15***	0.25***	-1.15***	0.25***	-1.14***	0.25***	-1.28***	0.41***
	(0.07)	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.09)	(0.04)
Treatment * female	0.05	-0.05						
	(0.11)	(0.05)						
Treatment * above median age			-0.07	0.03				
			(0.12)	(0.06)				
Treatment * below median subject					0.14	-0.10*		
					(0.13)	(0.06)		
Treatment * below median mock							0.26**	-0.31***
							(0.12)	(0.05)
Constant	0.66***	0.06	0.63***	0.08	0.66***	0.06	0.69***	0.02
	(0.14)	(0.06)	(0.13)	(0.06)	(0.13)	(0.06)	(0.13)	(0.06)
Overall treatment effect								
female	0.17**	-0.14***						
	(0.08)	(0.04)						
above median age			0.11	-0.09**				
			(0.10)	(0.05)				
below median subject					0.25*	-0.19***		
					(0.11)	(0.05)		
below median mock							0.28***	-0.27***
							(0.08)	(0.04)
Mean in control		0.32		0.29		0.42		0.54
Observations	730	730	730	730	730	730	730	730
R-squared	0.42	0.28	0.42	0.28	0.42	0.28	0.42	0.31

Maths is a standardized maths score. Fail maths is a dummy for whether a student got a fail in the maths exam. Above median age refers to a dummy if the student is above the median age for students in S4. Less median subject is a dummy variable if the student is taking less subjects than the median for the UCE exams. The middle panel shows the overall treatment effect for each group. The mean in control shows the control mean for that sub-group. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

significant heterogeneous effect of treatment. Maths scores increase by 0.26 standard deviations amongst this group from treatment, and the probability of failing maths falls an additional 31 percentage points. The coefficient on the simple treatment effect on the maths score result becomes only 0.02 standard deviations and is no longer significant, likewise for failing maths. The entire improvement in maths from watching the treatment movie is seen from the group who were in the bottom half of performance on the mock exam. Overall, students who performed below the median in the mock exam are 27 percentage points less likely to fail maths. Compared to a mean of 54% of students with below median mock scores failing maths, this means treatment reduced the probability a previously poorly performing student failed maths by 50%. These are very large effects, and suggest that those students struggling are more able to improve their maths scores from treatment.

I breakdown the treatment effect by mock exam performance further by interacting the treatment with each decile of mock score. These are shown in Table 15 , again only for the maths score and dummy variable for failing maths since there are no significant effects for total score, core score or English outcomes. The bottom decile, 1, is the excluded group. In column (1), the simple treatment coefficient is positive but insignificant. None of the interaction terms are significant, but the combined linear effect of the treatment and the treatment interacted with being in that mock decile (shown in the second panel) is significant and positive for deciles 3, 4 and 5. This suggests it is those in the bottom of the distribution who are benefiting from the treatment and not those at the top.

In column (2), failing maths, this time the simple treatment effect is large, negative and significant. This means the bottom decile by mock score is 24 percentage points less likely to fail maths if treated. This large negative effect is also true for deciles 2-4, suggesting all the lower deciles see a reduction in the probability they fail maths by seeing the treatment movie. Deciles 5-10 though show a large positive interaction effect with seeing the treatment, resulting in overall no effect of seeing the treatment movie on their likelihood of failing maths. This seems intuitive since they were very unlikely to fail maths to start with.

Moving onto the S6 class, heterogeneous treatment effects on exam performance are shown in Table 14 for the total score outcome variable only. Results for the principal subjects score and subsidiary paper score are similar. At S6 level, 49% of the students are women, 30% are above the median age of 19 years, 31% are taking a STEM subject and 43% scored below the median mock score. Looking first at gender, the total effect of treatment is significant for women $(0.12 + 0.08)$ at the 5% level and the coefficient on the simple treatment effect is no longer significant. This suggests that the beneficial effects from treatment are going to female students. For students above the median age, taking stem subjects and below the median in the mock, the point estimates on the

Table 13: Heterogeneity in treatment effects for S4 by mock decile

	(1) maths	(2) fail maths
treatment	0.14	-0.24***
	(0.14)	(0.07)
treatment * decile 2	-0.05	-0.00
	(0.23)	(0.11)
treatment * decile 3	0.14	0.02
	(0.22)	(0.10)
treatment * decile 4	0.26	-0.13
	(0.22)	(0.10)
treatment * decile 5	0.24	0.14
	(0.21)	(0.10)
treatment * decile 6	-0.16	0.37***
	(0.22)	(0.11)
treatment * decile 7	0.03	0.28***
	(0.21)	(0.10)
treatment * decile 8	-0.30	0.31***
	(0.24)	(0.11)
treatment * decile 9	-0.22	0.19*
	(0.21)	(0.10)
treatment * decile 10	-0.21	0.27**
	(0.22)	(0.11)
Overall treatment effect		
decile 1	0.14	-0.24***
	(0.14)	(0.07)
decile 2	0.10	-0.25*
	(0.18)	(0.08)
decile 3	0.28*	-0.22***
	(0.16)	(0.09)
decile 4	0.41**	-0.37***
	(0.17)	(0.08)
decile 5	0.39***	-0.10
	(0.15)	(0.07)
decile 6	-0.02	0.12
	(0.17)	(0.08)
decile 7	0.17	0.03
	(0.15)	(0.07)
decile 8	-0.16	0.07
	(0.19)	(0.09)
decile 9	-0.08	-0.05
	(0.15)	(0.07)
decile 10	-0.07	0.03
	(0.17)	(0.08)
Observations	730	730
R-squared	0.56	0.40

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. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

interaction with treatment are negative, but the standard errors are very large. There is no overall effect of treatment for these groups, suggesting it may be younger students, those not taking stem subjects and those who did better in the mock exam than the median who benefit from treatment.

Once again, I breakdown the treatment effect by mock exam performance further by interacting the treatment with each decile of mock score. These are shown in Table 13 for the total score only (there are no differences for principal score or subsidiary score). I find no differential effects of treatment by decile.

Table 14: Heterogeneity in treatment effect for S6 by gender, age and taking stem subjects

Impact of treatment assignment on S6 standardized test scores

	(1)	(2)	(3)	(4)
	Dependent variable: total score			
Treatment	0.08	0.17**	0.17**	0.16**
	(0.08)	(0.07)	(0.07)	(0.08)
Above median age	-0.15**	-0.09	-0.15**	-0.15**
	(0.07)	(0.09)	(0.07)	(0.07)
Female	-0.03	0.03	0.03	0.03
	(0.08)	(0.06)	(0.06)	(0.06)
STEM	0.45***	0.45***	0.50***	0.45***
	(0.07)	(0.07)	(0.09)	(0.07)
Below median mock	-1.20***	-1.20***	-1.20***	-1.16***
	(0.06)	(0.06)	(0.06)	(0.08)
Treatment * female	0.12			
	(0.11)			
treatment * above median age		-0.11		
		(0.12)		
treatment * STEM			-0.10	
			(0.12)	
treatment * below median mock				-0.07
				(0.11)
Constant	-0.51	-0.56	-0.56	-0.56
	(0.37)	(0.37)	(0.37)	(0.36)
Overall treatment effect				
Female	0.20**			
	(0.08)			
Above median age		0.06		
		(0.10)		
STEM			0.07	
			(0.10)	
Below median mock				0.10
				(0.08)
Observations	711	711	711	711
R-squared	0.50	0.50	0.50	0.50

Total is the aggregate score achieved, prin the score in 3 principal papers, sub the score on the subsidiary papers. All scores are standardized. Median age refers to being above the median age for students in S6. STEM refers to choosing maths, biology, chemistry or physics as a principal subject. The bottom panel shows the overall effect for each group.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 15: Heterogeneity in treatment effects for S6 by mock decile

	(1)
	total score
treatment	0.13
	(0.12)
treatment * decile mock 2	0.12
	(0.20)
treatment * decile mock 3	-0.19
	(0.19)
treatment * decile mock 4	0.01
	(0.20)
treatment * decile mock 5	0.03
	(0.21)
treatment * decile mock 6	-0.18
	(0.21)
treatment * decile mock 7	0.07
	(0.23)
treatment * decile mock 8	0.07
	(0.21)
treatment * decile mock 9	0.07
	(0.21)
Overall treatment effect	
decile 2	0.26
	0.16)
decile 3	-0.06
	(0.14)
decile 4	0.14
	(0.16)
decile 5	0.16
	(0.17)
decile 6	-0.05
	(0.16)
decile 7	0.20
	(0.19)
decile 8	0.20
	(0.17)
decile 9	0.20
	(0.17)
Observations	708
R-squared	0.58

Decile refers to decile of mock exam score. 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 only 9 deciles since two deciles had the same scores associated with them. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.3.2 School Characteristics

Since the Pre-Analysis Plan was lodged and exam results data analysed, additional data on the schools was collected. This data covered:

- The national ranking of the school
- The fees of the school

and is summarised in Table 5.

Heterogeneous treatment effects are also analysed by these school characteristics since information about the schools could provide important information about which types of students benefit most from seeing the treatment movie. For example, both the national ranking of schools and the fees charged by the school give an indication of how good the school is. If students at worse schools benefit more from seeing the treatment movie than those at better schools, then treatment might be able to partially mitigate factors related to poor schooling, such as teacher quality.

Heterogeneous treatment effects on maths exam performance for the S4 class are shown in Table 16. I only show here heterogeneity by maths score and failing maths as there are no effects for the total score, core score or English score. First looking at whether the school is a top 500; schools in Uganda are nationally ranked and this is freely available from the Ministry of Education. Out of the 3300 schools included in the ranking I consider the top 500 as the leading schools and look at treatment heterogeneity by this variable. The top 500 defines in the ranking what are considered good schools. 46% of the schools in my sample are ranked in the top 500. The treatment interactions with being at a top 500 school for the maths score is small and insignificant. The interaction coefficient on failing maths is positive 10 percentage points and significant at the 10% level. Overall, treatment results in students at school in the top 500 scoring 0.16 standard deviations higher on their maths exam, though this is only significant at the 10% level. There is no benefit from treatment to students at top 500 school in terms of failing maths, perhaps because students at top 500 schools already rarely failed maths (only 13% fail it in the control group). This result indicates that it's schools not in the top 500, so schools performing less well nationally, which benefit most from seeing the treatment movie in terms of the probability of failing maths, with treatment resulting in students at lower ranked schools being 16 percentage points less likely to fail maths. Seeing the treatment movie may therefore help to mitigate some aspects of being at a poorly performing school.

I also look at whether a school charges above the median fees in my sample, of which 31% do. For schools charging the highest fees, the interaction with the treatment has large and negative but not significant effect on the maths score. This would balance out against the pure effect from treatment of positive 0.20 standard deviations, resulting in students at high fees schools getting no

Table 16: Heterogeneity in treatment effect for S4 by school characteristics

Impact of treatment assignment on standardized test scores

	(1) maths	(2) fail maths	(3) maths	(4) fail maths
Treatmen	0.14*	-0.16***	0.20***	-0.17***
	(0.08)	(0.04)	(0.07)	(0.03)
Treatment * top 500	0.03	0.10*		
	(0.12)	(0.05)		
Top 500	0.54***	-0.45***		
	(0.20)	(0.09)		
Treatment * high fees			-0.17	0.18***
			(0.13)	(0.06)
High fees			-0.09	-0.09
			(0.16)	(0.07)
Constant	-0.09	0.51***	0.61***	0.11*
	(0.19)	(0.08)	(0.14)	(0.06)
Overall treatment effect				
Top 500	0.16*	-0.06		
	(0.09)	(0.04)		
High Fees			0.03	0.01
			(0.10)	(0.05)
Mean in control		0.13		0.05
Observations	730	730	730	730
R-squared	0.42	0.28	0.42	0.29

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

overall benefit to their maths score from seeing the treatment movie. Likewise, the coefficient is large, positive and significant for the interaction term for the likelihood of failing maths outcome which balances out against the large negative simple treatment effect giving no overall impact for students at high fees schools from seeing the treatment movie on failing maths. This indicates it is students at lower fee charging schools which see improvements in their maths scores and reduction in the probability of failing maths from watching the treatment movie. Again, this is likely because students at high fee schools already do much better, with only 5% of students at high fee schools failing maths.

This could be interpreted as because high charging schools are already doing so much better than low charging schools, that incremental improvements are much harder. If a student is at a poor school and almost failing maths, then small increases in effort or motivation could push that student over the threshold to no longer fail. If a student is at a high fee paying school, where presumably more and better quality resources are already being invested in their education, then if they are one of the few students failing maths, seeing the movie is not enough to improve their performance compared to everything already being done. Improvements in overall maths score are also much harder for students already doing very well at schools invested in their education.

Overall these heterogeneous effects highlight that it is students at worse performing schools, whether by ranking or fees who benefit most from seeing the treatment movie.

Turning to the S6 class, heterogeneous effects by school characteristics are shown in Table 17. Here I show the effect on the total score, principal paper score and subsidiary paper score. I examine the impact of a school being in the top 200. I use the top 200 to make it comparable to the top 500 out of 3300 schools at the S4 level, since at S6 only 1800 schools provide teaching at this level. I find some large but insignificant effects for a school being in the top 200 interacted with treatment on all the outcomes. These result in overall positive and significant effects from being at a top 200 school on total and principal paper scores. There is no significant effect on the subsidiary paper score. However the simple treatment effect is actually negative for the subsidiary score, suggesting if there is a positive effect of treatment on this score it is all coming through top 200 schools.

Secondly I examine whether there are heterogeneous effects by whether the school fees charged are higher than the median. Higher fee charging schools have students which perform better on the exams, especially for the subsidiary paper. The coefficients on the interaction of treatment with being at a high fee school are significant for both the overall score and subsidiary paper score, resulting in overall positive effects from treatment on all the outcomes for high fee schools, ranging from 0.16 standard deviations to 0.23 standard deviations. This suggests that the only group experiencing positive effects on the subsidiary paper is the high fees schools.

Note, there is no heterogeneity by school fixed effects, and so results are not shown here for brevity.

Table 17: Heterogeneity in treatment effect for S6 by school characteristics

Impact of treatment assignment on standardized test scores						
	(1)	(2)	(3)	(4)	(5)	(6)
	total	prin	sub	total	prin	sub
Treatment	-0.01	-0.01	-0.11	0.12*	0.11*	0.00
	(0.11)	(0.12)	(0.14)	(0.07)	(0.07)	(0.08)
Treatment * top 200	0.18	0.18	0.22			
	(0.13)	(0.13)	(0.16)			
Top 200	0.14	0.17*	0.11			
	(0.10)	(0.10)	(0.12)			
Treatment high fees				0.06	0.05	0.23*
				(0.10)	(0.11)	(0.13)
High fees				0.19**	0.10	0.33***
				(0.07)	(0.08)	(0.09)
Constant	0.15	0.07	-0.08	0.48	0.40	0.37
	(0.44)	(0.46)	(0.55)	(0.44)	(0.45)	(0.53)
Overall treatment effect						
Top 200	0.17***	0.17***	0.12			
High fees				0.18**	0.16**	0.23**
Observations	708	708	708	708	708	708
R-squared	0.58	0.56	0.30	0.58	0.55	0.34

Total is the aggregate score achieved, prin the score in 3 principal papers, sub the score on the subsidiary 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

4.4 Robustness

To check the robustness of my results I calculate permutation p-values and also check the robustness of the effect of treatment on the probability of failing maths for the S4 class to multiple hypothesis testing.

4.4.1 Randomisation test

I use permutation tests to compute exact test statistics which do not depend on asymptotic theorems. To do this I use Stata's permute function to randomly assign students to the treatment and control group and calculate the probability of observing the treatment effect I did under the null hypothesis that there is no treatment effect. I do this for the S4 and S6 outcomes defined in section 3.3 using 10000 permutations and without individual control variables, only school fixed effects. These are reported in Tables 18 and 19 underneath the robust p-values. At the S4 level,

Table 18: S4 main results robustness tests

Impact of treatment assignment on standardized test scores

	(1) Total score	(2) Core score	(3) Maths	(4) English
Treatment	0.006	-0.003	0.140	-0.044
Robust p-value	(0.930)	(0.966)	(0.045)**	(0.517)
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. Permutation p-value calculated using 10000 permutations.

*** p<0.01, ** p<0.05, * p<0.1

treatment still only has a positive impact on the maths exam and is still significant at the 5% level. At the S6 level, for both the overall score and principal subjects score, though the permutation p-values are higher than the robust p-values, treatment still has a positive effect, significant at the 5% level.

Table 19: S6 main results robustness tests

Impact of treatment assignment on standardized test scores

	(1) Overall score	(2) Principal subjects	(3) Subsidiary subjects
Treatment	0.169	0.165	0.079
Robust p-value	(0.024)**	(0.032)**	(0.222)
Permutation p-value	(0.041)**	(0.046)**	(0.409)
Observations	710	710	710
R-squared	0.196	0.150	0.288

Overall score refers to the aggregate score in the principal and subsidiary papers. Principal subjects refers to the standardised score on the 3 chosen subject papers. Subsidiary subjects refers to the standardised score on the two mandatory subsidiary papers. 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

4.4.2 Multiple hypothesis testing

The outcomes examined in the main results section were pre-specified in the pre-analysis plan as well as conforming to the standard when examining educational outcomes of focusing on overall scores and Maths and English. As a robustness check, I illustrate the treatment effect on failing any core subject to highlight the stability of the maths result to multiple testing. To do this, I compare the result of treatment for failing maths to the impact of treatment on the probability of failing each of the core exams listed in Section 3.3. I perform this only for the core subjects since all the students took these subjects.

To correct for multiple hypotheses, I compute sharpened q-values. Q-values adjust p-values to control for the false discovery rate. The false discovery rate is an approach which 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. This is a less stringent approach than those controlling for the probability of any type I error, such as the Bonferroni correction, and it therefore allows more power with a trade off of a higher rate of type I error. The method used here is Benjamini, Krieger and Yekutieli (2006) sharpened q-values as described in Anderson (2008) and using the code provided by Anderson online. This is one of the least conservative methods to control for false discovery rates. However the findings are unchanged even using conservative methods such as Bonferroni to calculate the q-values.

In Table 20 I show the impact of treatment assignment on the probability of failing each of the core exam subjects. Instead of displaying standard errors, I display both robust p-values and sharpened q-values below each coefficient.

It can be seen that the only subject for which the significant result is robust to multiple hypothesis testing is the maths result, where treatment results in an 11 percentage point reduced probability of failure and is significant at the 1% level even using sharpened q-values. Of the other core subjects, none is significant using either conventional p-values or sharpened q-values.

Table 20: Multiple hypothesis test for failing core subjects

Impact of treatment on S4 core subject fail

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Maths	English	Chemistry	Biology	Physics	History	Geography
Treatment	-0.11	-0.01	0.04	-0.00	0.01	0.00	0.01
p-value	(0.000)***	(0.703)	(0.123)	(0.959)	(0.654)	(0.867)	(0.76)
q-value	(0.001)***	(1.00)	(0.740)	(1.00)	(1.00)	(1.00)	(1.00)
control mean	0.27	0.11	0.38	0.35	0.49	0.16	0.12
Observations	729	729	728	729	728	727	706
R-squared	0.31	0.28	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

5 Cost effectiveness and discussion

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 sd in maths at S4 and 0.13 sd overall at S6. To compare this to some 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 sd) for $100 of spending. This is comparable to a remedial education programme in India which generated a 3 sd test score gain per $100 spent (Banerjee et al., 2007) or to teacher incentives in Kenya (Glewwe et al., 2010). Baird et al. (2016) find similar effects of 0.15 sd on maths score from giving a $5 conditional cash transfer a month to girls if they stay in school. Equally importantly is that my study had an effect over a period of 1-4 weeks whereas many studies are finding similar impacts after years of treatment. This intervention is therefore as effective and a similar cost as others aimed at impacting education attainment.

Additionally, this study showed the *Queen of Katwe* movie at a cinema for three main reasons: one, so that students could see the movie immediately upon its release before sitting their exams, two, because it was logistically simpler and faster then arranging screenings at schools, and three, to allow an individual randomisation at the cinema. If the study was scaled up though, the movie could be shown in schools to entire classes, perhaps through a projector or a specially arranged screening for many schools, and this might lower the cost further.

5.2 Interpretation of effects

The results for S4 and S6 students appear to come from different parts of the ability distribution. For the S4 class, it is lowest ability students as measured on a mock exam who benefit most from the treatment. This suggests the treatment is helping to compensate for being a poor student. Likewise effects are concentrated amongst students at lower ranked schools charging lower fees. This suggests effects are greatest at lower performing schools.

At the S6 level I find it is students at the highest ranked and highest fee schools who gain the most from treatment, and if anything, the decile analysis suggests the effects are most pronounced at the top of the distribution.

There are a number of possible reasons for these differences in effect. Firstly, the class profiles are very different. All schools have minimum requirements to go from S4 to S6, and on average nationally only one-third of students continue to S6 (MoES Uganda, 2015). The students for which

Queen of Katwe had an effect on at S4 are therefore unlikely to continue onto S6.

Secondly, the profile of subjects also differs greatly between S4 and S6. At S4, most subjects are compulsory and students take many subjects, whereas at S6, the principal papers, for which I see an improvement from seeing the treatment movie, are all optional and students take only three subjects.

Thirdly, there were very different time lags between seeing the treatment and the exam for S4 (one week) and S6 (one month). Having only one week between seeing the treatment and the exam means that there is only time to make a limited amount of improvement in the S4 class. If we assume it is easier to improve an exam score from fail to one above fail than from a B to an A, then 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, whereas at the top grades more effort is needed to reach a higher grade. Hence the time constraint might have meant I could only detect effects at the bottom of the distribution[§]. The fact that effects were only seen for the maths exam is a common result for this type of short term intervention and maths exams are considered more elastic then English or reading exams (Bettinger, 2010). At S6 they had at least a full month between seeing the treatment movie and the exams which is enough time for extra effort to pay off across the ability distribution and across their chosen subjects.

Lastly, schools are primarily judged in Ugandan at S4 level by how many students get the highest score, rather than how many get low scores, and so generally invest less in low performing S4 students and focus instead on a few best students. This could leave a large cohort of low performing students for which a small investment can have large payoffs in terms of exam performance. At S6, this effect is less pronounced and the focus is more on helping students achieve the grades to get into public university (2 passes in principal papers), rather than just those at the top. This could explain the more across the board effects seen in S6.

5.3 Policy implications

This paper stresses the importance of role models for raising students' aspirations about what they can achieve and calibrating the potential returns to education. Through this, a role model can have significant effects on students' educational attainment. An implication 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.

[§]Since the schools were randomly allocated across the 5 days of screening, I examined whether treatment effects for the S4 class varied by treatment day. However I do not find any difference between those students who were treated on the first days to compared to those students treated on the last days.

It is also important that schools don't just focus on the best performing students and leave the weakest behind. The fact that the *Queen of Katwe* movie had such a big effect on S4 students failing maths, especially at the worst schools, suggests that small changes at those schools could also have a big effect. One way to do this, as demonstrated in this study, is to place more emphasis on motivation and inspiration through example, to give more meaning to the students of how education can help them to achieve their life goals.

6 Conclusion

I find that exposing secondary school students to a movie featuring a potential role model who could raise aspirations improves national exam performance. Amongst S4 students completing lower secondary school, seeing the treatment movie increases maths scores by 0.14 sd, with the effect coming from lower ability students at worse schools being less likely to fail maths. 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 0.16 sd and increases the probability by 5 percentage points that they get the necessary grades for university.

This paper argues that the link between being exposed to a role model and improvements in exam performance is through imitation of the positive academic behaviours Phiona demonstrates to get into a top school and play chess, and also through increased aspirations by raising their beliefs over what goals are possible for them to achieve. Further work would hope to disentangle these potential mechanisms to find out whether seeing the treatment movie only increased studying effort or whether it also raised aspirations and caused them to aim their goals higher.

Appendix

Table 21: Ordered logit regression of the impact of treatment on core subject grade for S4

	(1)	(2)	(3)	(4)	(5)	(6)
	Eng	Chem	Phy	Bio	Hist	Geog
Fail	0.011	0.021	0.016	0.014	0.002	-0.005
	(0.010)	(0.017)	(0.019)	(0.015)	(0.012)	(0.011)
2	0.005	-0.000	-0.003	0.001	0.000	-0.001
	(0.005)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)
3	0.003	-0.004	-0.004	-0.001	0.000	-0.001
	(0.002)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)
4	0.002	-0.003	-0.003	-0.004	0.000	-0.000
	(0.002)	(0.003)	(0.003)	(0.004)	(0.001)	(0.001)
5	-0.005	-0.004	-0.002	-0.004	0.000	0.000
	(0.005)	(0.004)	(0.002)	(0.004)	(0.000)	(0.001)
6	-0.008	-0.004	-0.002	-0.003	-0.000	0.002
	(0.007)	(0.003)	(0.002)	(0.003)	(0.001)	(0.004)
7	-0.004	-0.004	-0.002	-0.002	-0.001	0.002
	(0.004)	(0.003)	(0.002)	(0.002)	(0.003)	(0.005)
8	-0.002	-0.001	-0.002	-0.001	-0.001	0.002
	(0.002)	(0.001)	(0.002)	(0.001)	(0.004)	(0.003)
9	-0.002	-0.001	-0.000	-0.000	-0.001	0.001
	(0.002)	(0.001)	(0.000)	(0.000)	(0.007)	(0.002)
Observations	729	728	728	729	727	706

Core subjects are taken by all students at S4 level. All regressions include school fixed effects and individual characteristics (age, gender, number of subjects taken and mock score). Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

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.*

Baird, S., Mcintosh, C., and Özler, B. (2016). When the Money Runs Out Do Cash Transfers Have Sustained Effects on Human Capital Accumulation? *World Bank Policy Research Working Paper.*

Bandura, A. (1977a). Self-efficacy : the exercise of control. *Psychological Review*, 84(2):191–215.

Bandura, A. (1977b). *Social learning theory.* Englewood Cliffs, NJ: General Learning Press.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics.*

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls. *Science*, 335(6068):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.*

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.

Bettinger, E. P. (2010). Paying to learn: the effect of financial incentives on elementary school test scores. *NBER Working Paper Series.*

Chesterton, P. (2004). Evaluation of the Meena Communication Initiative. *UNICEF Impact Evaluation.*

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.

Claro, S., Paunesku, D., and Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences*, 113(31):201608207.

Dalton, P. S., Ghosal, S., and Mani, A. (2014). Poverty and Aspirations Failure. *The Economic Journal*, 126:165–188.

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*.

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

Evans, D. K. and Yuan, F. (2017). The Economic Returns to Interventions that Increase Learning.

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

Glewwe, P. W., Ilias, N., and Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, pages 205–227.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., and Woessmann, L. (2014). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73:103–130.

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.

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

Kremer, M., Brannen, C., and Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science (New York, N.Y.)*, 340(6130):297–300.

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.

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2011). The Impact of Short-term Incentives on Student Performance.

Locke, E. A. and Latham, G. P. (2002). Building a Practically Useful Theory of Goal Setting and Task Motivation A 35-Year Odyssey. *American Psycologist*.

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.

MoES Uganda (2015). Education Statistical Abstract.

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

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

Rotter, J. B. (1966). . Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: general and applied*, 80(609).

Serneels, P. and Dercon, S. (2014). Aspirations, Poverty and Education: Evidence from India. *Young lives working paper*.

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.

The World Bank (2015). *Mind Society and Behaviour World Development Report*.

Wydick, B., Glewwe, P. W., and Rutledge, L. (2013). Does International Child Sponsorship Work? A Six-Country Study of Impacts on Adult Life Outcomes. *Journal of Political Economy*, 121(2):393–436.