Nova School of Business and Economics

Essays on the Economics of Education and Development

by

Ruth Fortmann

A Thesis submitted as part of the requirements

for the degree of

Doctor of Philosophy (Ph.D.) in Economics

at

Nova School of Business and Economics
Universidade Nova de Lisboa

August 2018

Acknowledgements

I am especially thankful to my supervisor Catia Batista, Shan Aman-Rana, Mehregan Ameri, Tim Besley, Gharad Bryan, Daniel Chandler, Svetlana Chekmasova, Alexia Delfino, Dita Eckhardt, Greg Fischer, Simon Franklin, Ria Ivandic, Jay Lee, Johanna Posch, Wolfgang Ridinger, Kilian Russ, Giulia Tagliaferri, Martina Zanella, and Celine Zipfel for their support and time discussing this research.

I would also like to thank Sofia Vala, Jane Dickson, and Rhoda Frith for their help over the years.

I am endlessly grateful to my family and Sohail for their encouragement and belief in me.

Contents

List of Tables

List of Figures

Introduction

Equality of opportunity is the most fundamental characteristic of any fair society. In Economics, equity often comes at the expense of efficiency. Income taxes are a famous example of this trade-off as lowering taxes is thought to incentivise productivity while it limits the resources available for redistribution. However, there are some policies that can increase both equity and efficiency. Investment in education is one of those examples. Equity in education increases both fairness and efficiency by enabling everyone in the economy to achieve their full potential. Nonetheless, in the UK, pupils who are eligible for free-school-meals, compared to more advantaged pupils, lag 9.5 months behind in educational attainment at the end of primary school (Education Endowment Fund (2018)).

Whilst social mobility in the UK is on the rise, with 43% experiencing an increase in living standards compared to their parents, access to careers remains strongly segregated by socioeconomic backgrounds. According to the Social Mobility Report (2017), most traditional professions, such as doctors, lawyers and academics, in the UK are still firmly in the hands of those from high socioeconomic backgrounds and an average pay gap between working class and professional backgrounds of £6,800 p.a. documents that the divide persists also within professions.

In the first chapter of this thesis, I focus on the demand side of this issue by testing whether the inequality in the demand for professional jobs is reflected in the inequality in the demand for education. If the link exists, raising demand for careers that require higher levels of education amongst disadvantaged pupils will contribute to improving equity in education. I analyse the impact on test scores of an intervention that aims at increasing pupils' demand for professional careers by providing information through careers talks to improve the understanding of the relationship between knowledge about high achieving jobs and performance at school.

I develop a theoretical framework to model the impact of careers talks on the optimal level of effort that pupils invest in schooling, which determines performance. I formalise the effect of novel information on professional careers by adding a high utility job outcome to the objective function of the pupil, where the probability of attaining it depends on effort. I also discuss two potential channels that can negate the positive effect, first, if the careers talk speaker induces distraction during the subsequent test, or secondly, if the speaker intimidates the pupil and thereby reduces the perceived probability of attaining the high utility job.

I use data on careers talks in English schools to test whether pupils in cohorts that held careers events increased their performance on standardised tests at the end of primary school. To study the role of the careers talk speaker as the transmitter of information and understand possible role model effects, I estimate treatment effects separately for girls and boys with same and opposite gender speakers. I find that careers talks do not change the performance of pupils on average, but have a negative effect on Maths and Reading scores of girls, and Maths scores of boys, when the speaker is male. Estimating heterogeneous treatment effects by regional views on women's rights in the labour market, shows that the negative effects on girls' and boys' Maths performance are stronger in regions with more traditional gender norms. This implies that region specific characteristics that are correlated with gender norms might be driving these effects.

The main contribution of this paper is to study the effect of careers talks on the educational attainment of primary school pupils, with emphasis on the difference in impacts on girls and boys with same gender and opposite gender speakers. I use novel data on careers talks in English primary schools to estimate the impact of careers talks by male and female speakers from a range of professional backgrounds on primary school pupils, thereby adding to the existing evidence which focuses primarily on the influence of female speakers on older children. The results highlight important heterogeneity in the treatment effects by speaker gender.

In the second chapter, I delve deeper into the specific influence of role models on aspirations as the mechanism behind the effect on performance. I conduct an experiment in which pupils in treatment schools are exposed to a video intervention in which a speaker of varying gender with a career in a Maths related area highlights the importance of the subject in a variety of careers and in daily life. The video intervention allows to rule out any systematic differences in speaker quality across different schools in the treatment group and the variation by gender permits to disentangle treatment effects for pupils who can relate to the speaker via gender and those who cannot. For the entire sample, I collect survey data on aspirations before and after the intervention and data on performance on standardised tests in Maths and Reading after the intervention including information on test scores in earlier tests to measure prior attainment.

I find no significant average treatment effect on aspirations or test scores for pupils who watched the video compared to those who did not. However, I find a positive effect on Maths test scores for boys with male speakers and also on their degree aspirations and the self reported difficulty of their dream job, which I use as a measure of career aspirations. Girls with female speakers are more likely to embrace a Growth Mindset and the average same gender effect is also positive on subject specific aspirations in Maths. The fact that these positive effects are only present in pupils who were exposed to a same gender speaker highlights the importance of same gender role models as a mechanism. Interactions of the same gender effects with peer performance illustrate that high achieving female peers act as a substitute for the speaker in the video for girls whereas these peer effects are not significant for boys.

The positive effects on boys with male speakers stand in contrast with the negative effects in the previous chapter. Whilst, estimations in the first chapter show that the effects are more pronounced in regions with more traditional gender norms, this is not the case in this chapter where the interactions with gender norms are insignificant. Since speaker quality in this chapter is uniform across schools, thanks to the standardised video intervention, while it varies in the first chapter as the availability of speakers is geographically restricted, this contrast suggests that the quality of male speakers in regions with more traditional views explains parts of the negative effects in the first chapter.

These findings contribute to the existing literature by providing experimental evidence for the positive effect of same gender role models on primary school pupils' aspirations for education and Maths in particular, and Maths test scores. This work adds to the results in the previous chapter and the related literature by using survey data to explicitly document the effect of role models on non-cognitive skills such as aspirations as drivers of the impact on test scores. Furthermore, the controlled video intervention allows to, first, isolate the effect of relatability via gender from any speaker specific effects, and secondly the randomised design to clearly identify the causal effect by ruling out bias in the estimations.

In the last chapter, I am concerned with optimal risk sharing behaviour to yield the best informal insurance protection in the absence of formal insurance products. The poor in many rural parts of the developing world do not have access to formal insurance providers and must rely on risk sharing agreements with peers to protect themselves from the consequences of adverse shocks to their income that arise from weather conditions or illness. Since a steady income flow is a prerequisite for incentivising investments in children's education, informal insurance is an important tool to promote social mobility and to prevent the income gap between rich and poor from widening further.

In the analysis, I focus on the optimal choice of risk sharing partners with respect to people's risk profiles. I develop a model that links the optimal choice to the risk variance and covariance of idiosyncratic shocks of two utility maximising agents and show that in equilibrium, risk sharing will take place between agents with similar shock variances and negative covariances of their idiosyncratic shocks.

I test these theoretical implications empirically using data on an insurance network in a rural Tanzanian village and find a strong correlation between the probability of risk sharing and the similarity in people's shock variance as measured by their level of riskiness based on the diversification of their income profiles. I also find evidence for a correlation between insurance links and negative shock covariances as measured by the covariance of income profiles, but this correlation is not significant when I include the similarity of shock variances as a regressor. These results suggest that these risk sharing agreements adhere to the optimal allocation in the model, but that observed behaviour is less likely to deviate from the equilibrium implications regarding similar shock variances than negative shock covariances.

There are two main contributions of this work to the existing literature. First, I derive a model of the endogenous formation of informal insurance links that allows for non-zero covariances between idiosyncratic shocks, thereby extending the framework by Jaramillo et al. (2015). Secondly, this is the first paper to empirically test whether observed risk sharing agreements comply with the theoretical implications on the role of risk profiles of insurance partners. Whilst previous papers have discussed the importance of counter related shocks to ensure insurability, this work is the first to test this directly.

To conclude, the first two chapters of this thesis discuss the importance of knowledge about labour market opportunities and professional role models and aspirations in improving educational attainment of primary school pupils to lay the foundation for greater career choice and professional success in the future. Both chapters highlight strong differences in the magnitude and direction of the effect by gender, and more research is needed to determine the long run effects on career choice and income. The last chapter focuses on informal insurance as a tool to smooth consumption and incentivise investments in the future, such as children's education, and studies the optimal allocation of risk sharing partners with respect to their risk profiles. I find that observed risk sharing largely adheres to the equilibrium implications of the model, which prescribe that utility maximising risk sharing takes place between agents with similar shock variances and negative shock covariances.

Chapter 1

The Effect of Labour Market Information on Educational Achievement: Evidence from Careers Talks in UK Primary Schools

Abstract
In the UK, a child of working class background is 2.3 times more likely to enter a working class job compared to children with higher socioeconomic background. One possible driver of this inequality is an unequal distribution of knowledge about high achieving careers that creates inequality in the demand for careers, which might decrease motivation and performance at school. In this paper, I focus on this channel by studying the impact of providing information about high achieving careers through careers talks on increasing motivation and educational achievement of primary school pupils in end-of-year standardised tests. The results show that the average treatment effect of attending a careers talk is insignificant for both Maths and Reading outcomes, but disentangling the effects by pupil gender each with male and female speakers respectively highlights that both genders when exposed to male speakers have significantly lower test scores in Maths, and girls also in Reading, compared to the control group. Interactions of the treatment effects with measures of regional gender norms suggest that the negative effects on boys and girls on Maths test scores are amplified in regions with more traditional gender views.

JEL codes: I24, I26, J16
Keywords: educational inequality, labour market information, role models

1.1 Introduction

Whilst social mobility in the UK is on the rise, career choices are still highly segregated by socioeconomic background, which constitutes one explanation for why 45% of earning inequalities are passed from parents onto their children. According to the Social Mobility Report (2017), a child of working class background is 2.3 times more likely to enter a working class job compared to children with higher socioeconomic background. Similarly, children whose parents work in professional jobs are 2.5 times more likely to pursue a professional career themselves than children from less advantaged backgrounds. These figures raise questions about the drivers of this inequality of opportunity. Possible explanations include discrimination in the hiring process based on social class or disadvantages that arise due to negative social network effects for workers from lower socioeconomic backgrounds. This might lead to an unequal distribution of knowledge about high achieving careers that can create inequality in the demand for these careers, which in turn might decrease motivation and performance at school if lower achieving jobs require less education. In this research, I focus on this channel by studying the impact of the provision of information about high achieving careers through careers talks on increasing motivation and educational achievement of primary school pupils.

Children acquire a significant part of their knowledge about future career options through family and friends, therefore the socioeconomic status of the child can influence the knowledge they can gain. Families of low socioeconomic background are less likely to work in high-achieving jobs, likewise other people with whom the child will be surrounded, due to social segregation. Consequently, careers talks in schools are becoming an increasingly popular instrument to expand pupils' knowledge on career options. In these talks, speakers from the world of work discuss their jobs and professional backgrounds with pupils in an effort to increase motivation and thereby improve performance. In this paper, I use data on careers talks in English primary schools during the school year 2014/15 and 2015/16 to provide evidence on their effectiveness in improving educational outcomes.

The aim of careers talks is to provide pupils with labour market information on career options that they do not yet know about through family or the media, and to discuss entry paths into these jobs. In this paper, I develop a theoretical framework in which job outcomes determine the level of effort pupils invest into schooling. In the model, pupils are introduced to a new high utility job outcome through careers talks, and the probability of attaining it depends on effort. The main implication of the model is that pupils increase effort in response to the novel information, which raises performance on tests. Furthermore, I discuss two side effects of the talks that can negate the positive effect of information on performance either through distracting the pupil and inducing additional stress during the test, or by reducing the probability of attaining the high utility job outcome if the speaker appears unattainable.

In the empirical analysis, I focus on careers talks in English primary schools that take place during year 5 and year 6 of the two cohorts graduating primary school in 2015 and 2016 by estimating the effect of attending a talk on pupils' performance on year 6 standardised tests in Maths and Reading, known as *year 6 SATs*. I estimate heterogeneous treatment effects for pupils who attend a talk by a speaker of the same gender compared to the opposite gender and disentangle the effects for boys and girls. The estimations show that girls and boys perform significantly worse in Maths if they attend a talk by a male speaker, and for girls this effect is also significant for performance on Reading tests. I find that the effect is stronger in geographic regions with more traditional views on women's role in the labour market, which suggests that the effects are to a certain extent driven by gender norms or correlated factors such as speaker quality.

I use data from two different sources that provide information on careers talks, and pupil level data on test scores and some school characteristics. The first source contains information on careers events that stems from administrative data from an educational charity that operates as a platform to facilitate the contact between schools and volunteer speakers across the UK to organise careers talks. The data cover all the expressions of interest from primary schools to host an event between the launch of the charity's programme in the summer of 2014 and the end of the school year in 2016. This information includes the number of speakers that a school invited for a given cohort, the speakers' names from which I can infer the gender, and the number of events that took place. This allows to construct a school-cohort level dataset with information on the number of invites, talks and events for each of the two cohorts per school that graduated in 2015 and 2016.

I populate the school-cohort level dataset with individual level performance data by using data on test scores in year 6 standardised test, and year 2 tests to control for prior attainment, from the National Pupil Database (NPD). In addition to test scores, the NPD data holds information on pupil gender, school averages of free-school-meal eligibility by age and school location. Based on this dataset, I can estimate the effect of the cohort level treatment on individual level test scores.

Identification in this study relies on exploiting variation in the allocation of school-cohorts to treatment and control group, and further limit the scope for endogeneity through different sets of individual and school level controls. Selection bias can arise from three sources in this setting: First, schools actively opt into treatment by inviting speakers for a careers talk. This poses a concern for identification if pupils in schools that select into treatment perform systematically better or worse in year 6 tests than pupils in schools that do not hold career events. To rule out bias through this channel, I construct the entire sample based on schools that actively use the platform to invite speakers and use as the control group those cohorts within the schools that have not attended a talk. This ensures that no bias can arise from a correlation between using the platform and other performance enhancing measures that a school undertakes. In addition to eliminating this bias on the extensive margin, I include the number of invites to speakers as a covariate to control for the intensive margin of the effort that schools put into hosting a talk.

The second source of endogeneity arises from the possibility that the likelihood of a speaker accepting the invite from a school might be correlated with performance. If speakers prefer to accept invites from schools with lower or higher than average performance, then this poses a bias in the estimation as treatment is correlated with the outcome. To account for this, I use the school-cohort structure of the data to further restrict the sample to only those schools that have hosted a careers talk for any age group before. If a school held an event for the cohort that graduated in 2015 but not for the one that graduated in 2016, this school will be represented in the treatment group by the 2015 graduating cohort and in the control group by the 2016 graduating cohort. Additionally, the sample also includes schools that hosted events for cohorts too young to be included in the estimation, namely any cohorts that graduated primary school after 2016. A school that only hosted an event for younger cohorts, the 2015 and 2016 graduating cohorts of that school will both be part of the control group. Likewise, if both graduating cohorts of a school attended a careers talk, the school will only appear in the treatment group.

Finally, selection bias can arise from cohort level selection into treatment. This will be an issue if cohort level characteristics that increase the likelihood of treatment are also correlated with performance, such as quality and engagement of the cohort's teacher. I partly address this issue by controlling for observable confounders with four different sets of controls. The first specification includes no controls, the second a set of ad hoc selected controls, in the third, I use post-double LASSO regularisation to select a set of controls from the full set of potential covariates, and in the fourth I add school fixed effects to the post-double LASSO controls. Even though the selected controls vary, I find that my results are stable across specifications, which adds confidence to the hypothesis that treatment is a significant driver of the estimated effects on test scores. Notwithstanding, I cannot rule out selection bias from unobservable confounders that are uncorrelated with the covariates in the model.

The empirical analysis shows that the average treatment effect of attending a careers talk is insignificant on both Maths and Reading outcomes. Disentangling the average effects for boys and girls each with male and female speakers respectively shows that both genders when exposed to male speakers have significantly lower test scores in Maths, and girls also in Reading, compared to the control group. When controlling for post-double LASSO selected covariates and school fixed effects, girls' performance deteriorates on average by 0.15 standard deviations in Maths and by 0.11 standard deviations in Reading, whereas the effect on Maths scores for boys with male speakers is -0.12 standard deviations. Interactions of the treatment effects with measures of regional gender norms from the World Values Survey (Inglehart et al. (2014)) suggest that the negative effects on boys and girls with male speakers on Maths test scores are amplified in regions with more traditional views on women's role in the labour force. Quantile regressions show that the negative effects on girls' Maths performance are widely spread over the distribution, and on Reading are more concentrated in the top half of the performance distribution, whereas the effects on boys are also spread between the second and sixth decile.

These findings seem to imply that any positive effect of the careers talks that arises due to novel information is negated by negative effects on girls and boys. The negative effect on girls is compatible with the theoretical predictions about male speakers inducing additional stress and distraction during tests, for instance by priming girls on a negative stereotype regarding their aptness for professional success. Under this theory, girls who are subject to the negative stereotype that men are more professionally successful than women might find themselves confronted with the threat of confirming the stereotype in the subsequent test, which represents a distraction and can lead to worse performance (Spencer et al. (1999)). However, the related literature suggests that these effects are temporary as they do not affect long term effort, and will therefore disappear over time.

The effect on boys is less clearly reflected in the predictions of the model, but it might be due to boys discounting the probability of achieving a high job outcome themselves when attending a talk by a speaker whom they consider 'out of reach' given their own ability. According to the theoretical predictions, this effect should be largest on low performing boys as they are most likely to perceive the speakers as unattainable. However, this is not clearly borne out in the quantile regressions, which also show negative effects for boys in the top half of the performance distribution. Furthermore, since not only norms but also speakers vary by region, the heterogeneity in the treatment effect by regional gender norms could also be driven by regional variation in the quality of speakers.

The main contribution of this paper is to study the effect of careers talks on the educational achievement of primary school pupils, with special focus on the differential impacts of same gender and opposite gender speakers. I build a theoretical model in which utility maximising agents choose the level of effort that they invest into education. Careers talks provide information on a new high achieving, high effort job outcome, thereby adding to the set of attainable job options in the utility function. I also introducing two channels that can negate the positive effect from the novel information. In the empirical analysis, I use novel data on careers talks in English primary schools, which allow me to study the impact on the educational achievement of boys and girls with same and opposite gender speakers separately.

This paper ties in with a long standing literature on human capital investment and the relationship between education and inequality. The early literature on the topic (Becker (1962) and the subsequent book Becker (1964)) discusses human capital as an input to production, which grows through investments, instead of innate ability which is constant over time. In this work, Becker highlights education and training as the most important investments into human capital as measured by later returns in earnings. Around the same time a series of papers established the link between education and the observed skewness in the distribution of income. Mincer (1958) builds a model that relates the unequal distribution of earnings to education, and Becker and Chiswick (1966) measures the contribution of schooling to income inequality in the US. Furthermore, Chiswick (1969) develops a model and conducts a cross country analysis of the effect of equal opportunity schooling legislation such as minimum schooling laws on inequality in the UK and the Netherlands. He shows that minimum schooling laws tend to increase the level and skewness, and decrease the inequality in income and schooling.

More recent work by Autor (2014) shows that the rising wage premium for high-skilled labour explains a substantial share of the increase in inequality in many industrialised countries. In a cross country comparison, Hanushek and Woessmann (2011) document that a one standard deviation increase in numeracy skills is associated with 12-15% higher wages in Nordic countries or 28% higher wages in the US, and that differences in cognitive skills account for a large share of growth differences between OECD countries, more so than a variety of other economic institutions (Hanushek et al. (2014)).

By linking family income and credit constraints to child development and schooling, Becker and Tomes (1979, 1986) show that differences in the demand for education by income also yield unequal investment into children's schooling. A number of studies have since focused on the impact of labour market information on educational attainment, and school and degree choice. Jensen (2010) in the Dominican Republic and Nguyen (2008) in Madagascar show that increasing the perceived returns to education through provision of information on returns in the labour market increases years of schooling and performance on tests. Attanasio and Kaufmann (2014) find that expected returns to education are important determinants of schooling decisions in Mexico.

Furthermore, Hastings et al. (2015) use data from Chile to show that students have inaccurate beliefs about major specific earnings, Baker et al. (2017) find that low-income, low-achieving students who apply to low-earning degrees in California (USA) overestimate salaries by more than 100%, and Reuben et al. (2017) document that girls have lower earnings expectations than boys. All studies find that labour market information affects major choice. A recent study by Lergetporer et al. (2018) analyses the relationship between information on the cost and returns to attending university and educational aspirations of parents in Germany. They find that the information increases educational aspirations of parents with and without a university degree, however, since the effect is larger on parents with university degrees, they conclude that the inequality in information does not explain the socioeconomic gap in educational aspirations.

The careers talks in this paper can be thought of as a similar type of intervention to raise the perceived returns to education by increasing knowledge about high achieving jobs. I contribute to the existing literature by showing how the effect of labour market information on primary school pupils' educational achievement crucially depends on their relationship to the speaker who acts as the transmitter of information.

The work most similar to the one in this chapter is a recent study in France by Breda et al. (2018), which exploits a randomised experiment to study the effect of female STEM role models on course choice and college enrolment of high school pupils. They find a large effect on enrolment in science programmes for high-achieving girls, and show that the role model effect is stronger if the pupils can relate to the individual via gender and also ability. In this paper, I add to these findings by analysing the effect of careers talks on the educational achievement of primary school pupils, and by disentangling the impact of same gender versus opposite gender speakers for both boys and girls. I show that gender matching matters significantly and can have adverse effects on both genders.

In the remainder of the paper, I first provide a simple theoretical framework to guide the interpretation in section 1.2, and lay out the institutional background in section 1.3. In section 1.4, I illustrate the identification strategy, and in section 1.5, I present the data that I use in the empirical analysis, as well as summary statistics and balancing tests. In section 1.6, I discuss the estimation and methods, and in section 1.7 the results. I discuss the compatibility of theoretical and empirical results, and the implications of the limitations of the identification strategy in section 1.8. Finally, section 1.9 concludes with a summary and suggestions for future research.

1.2 Theoretical Framework

The following framework is intended to model how careers talks might influence the educational achievement of pupils. The main hypothesis in this work is that careers talks improve educational achievement by providing novel information on high achieving, high effort career options being attainable, thereby increasing the perceived returns to education.

1.2.1 Base Case

The basic setup of the model features an agent who maximises utility from potential job outcomes through effort. I formalise the effect of novel information from careers talks by adding a new high achieving job opportunity θ^h to the utility function. The probability of attaining the new job option depends on the level of effort e that the agent invests into schooling, which in turn determines performance. Let this probability be $p(e)$, and the probability of attaining the low achieving job option θ_L be $1 - p(e)$. Therefore, agents maximise the utility function

$$\max_e (1 - p(e))\theta_L + p(e)\theta_H - \frac{1}{2}e^2 \tag{1.1}$$

where $\frac{1}{2}e^2$ is the standard quadratic cost function.

Since effort is costly, agents invest effort only if the new high achieving job outcome provides enough utility to offset the cost. Thus, by providing information on high-achieving, high utility job options, careers talks will raise effort, and thereby performance.

1.2.2 Possible Deviations from the Base Case

I discuss two possible mechanisms that might lead the reduced form effect of careers talks to deviate from this prediction. First, if the reference point of achievement that the speaker in the careers talk defines, appears unattainable to the pupil, i.e. the distance between the reference point and the pupil's subjective belief about their potential achievements appears too far, the talk might discourage the pupil instead of motivating them.[1] I formalise this scenario as a subjective reduction of the probability of achieving the high outcome $p(e)$ for all effort levels by a, with $a < 1$ if the speaker induced reference point is too distant.

The second channel that might outweigh the positive effect from providing novel information on high achieving career outcomes is the possibility that speakers cause additional stress and distraction in pupils during tests, which increases the cost of

[1] See Genicot and Ray (2017) for a formal discussion.

effort. Defining this additional cost of effort as $s > 1$ yields the new utility function

$$\max_e (1 - p(e))\theta_L + p(e)a\theta_H - \frac{1}{2}se^2 \tag{1.2}$$

The stereotype threat literature in social psychology suggests that such additional distraction might be caused by careers talk speakers who remind pupils of a negative stereotype, such as the negative connotation attached to women's Maths ability (Spencer et al. (1999)). The salience of the stereotype might induce the threat of confirming it, in subsequent tests in pupils, which adds an additional layer of burden to the test situation and thereby increases the likelihood of self-fulfilling the stereotype. According to the theory, priming people on the stereotype will negatively affect the performance of people who are subject to the stereotype. Experimental evidence of the theory in Steele and Aronson (1995) confirms that this prediction will hold for performance in tasks that are diagnostic of stereotyped feature and also non-diagnostic tasks, which implies that the negative impact could be reflected also in Reading scores.

The stereotype threat literature focuses only on the effect on tests that take place immediately after the priming, which is not the case in this setup as the pupils take the tests up to two years after the careers talk. However, if the speakers leave a lasting impression of pupils are reminded of the talks when taking the test and the negative priming becomes salient again, the negative effect could still be at play.

1.2.3 Predictions

The framework implies that careers events will improve performance on standardised tests if pupils learn about new, high utility career opportunities, as these increase the optimal level of effort they invest into schooling. In addition, the model gives rise to two potential channels that might negate the positive effect and reduce performance.

First, pupils might discount the probability of achieving the high utility job outcome at all levels of effort if the speaker appears out of reach compared to pupils' own perceived ability. A lower probability of attaining the high utility job outcome leads to a lower optimal level of effort, which reduces performance. This channel is particularly relevant for low achieving pupils as they are more likely to consider the speakers unattainable.

The second mechanism that might outweigh the positive effect from providing information on a high utility job option in this model is captured by negative associations that the pupil might connect with the speaker, which distract them during the subsequent test and thereby lower performance. These negative associations might be related to a negative stereotype that a group of pupils feel reminded of when attending a talk by a speaker of another group. This might include male speakers reinforcing the negative stereotype about girls' Maths ability or aptness for professional success. Furthermore, pupils of an ethnic minority may face this stereotype threat when attending a talk by a white speaker. Unfortunately, I am unable to detect the presence of this particular type of stereotype threat as the data does not contain information on the ethnicity of pupils or speakers.

1.3 Institutional Background

In this project, I collaborate with a charity that serves as a platform for schools to connect with speakers from the world of work. The platform facilitates contact with professionals who volunteer to give careers talks in schools. To host an event, schools initiate the process by inviting speakers for a talk, while invites to speakers are based on information on their age, gender, preferences regarding geographic location of events, industry, area of specialisation and a short bio. The platform does not provide indications of volunteers' quality. Depending on the format of the event, schools send out invites to one or multiple volunteers (which they can either decline or accept).

The mission of the charity is to foster employer engagement in schools. Conversations with both the charity and schools suggest that most schools host events to introduce their pupils to speakers with different professional backgrounds in order to expand their knowledge of careers beyond those of their parents and to raise aspirations. Schools may use the platform to reach out to volunteers for a variety of events, which include, but not limited to, careers weeks, subject specific events (e.g. Science week), and guest speakers in the morning assembly.

An event takes place if the invited volunteer accepts the school's invitation. The primary reasons for volunteers to decline an invite are that they are unavailable on the day of the event, they face difficulties reaching the school with public transport, or their professional profile is not suited for the type of event that a school organises. There is no evidence for strategic acceptance decisions based on school quality, though this may be a concern in exceptional cases.

1.4 Identification Strategy

Since treatment is not randomly assigned, the setup in this study is susceptible to selection bias. Participating schools opt into hosting an event and are thus inherently different from schools that do not select in. However, I exploit the unique nature of the data to reduce the scope for endogeneity substantially. In what follows, I will first characterise the selection bias that threatens causal identification in more detail and then propose a remedy.

There are three sources of selection bias. First, schools opt into treatment by signing up to the platform and inviting speakers. This is the only way to initiate an event as the charity running the platform merely facilitates the contact but does not set up events themselves. Therefore, schools that set up an event are inherently different from those that do not. To address this source of selection bias, I limit the sample to only those schools that are signed up to the platform and are actively using it to invite speakers. This procedure eliminates selection bias from opting into treatment since all schools in the remaining sample have at least attempted to do so.

The second source of selection bias stems from potential differences in volunteers' propensity to accept an invite, for instance based on school location or quality. Volunteers may be more likely to accept an invite if a school is well connected to public transport or is of higher quality, or the reverse. Either way, the regression would suffer from omitted variable bias if location and quality were correlated with test scores. To combat this issue, I exploit the school-cohort panel nature of the data, which allows to construct a treatment indicator for each cohort per school and to exclude schools that never held an event (despite sending invites to volunteers) from the sample. This way, I can eliminate any time-invariant volunteer selection bias as all schools in the remaining sample hosted at least one careers talk for one of their year groups during the period of observation. This may include events for year groups who graduate from primary school after 2016 in which case both cohorts of interest, those graduating in 2015 and 2016, are included in the control group. The sample also includes schools where both the 2015 and 2016 graduating cohorts attended events, which implies that the school is exclusively represented in the treatment group and not in the control group.

Overall, of the 101 schools in the sample, both cohorts are in the control group in 39 cases, the two cohorts are split between treatment and control in 37 cases, and in 25 schools, both cohorts are in the treatment group.

Furthermore, to control for any remaining time-invariant bias at the school level, I present my main specifications with and without school fixed effects.

Finally, this leaves scope for bias due to selection at the cohort level within a school, which is a concern for identification if the probability of hosting an event is correlated with performance of the cohort. This would require that schools are more likely to invite speakers if a cohort performs better or worse than the school average, and / or speakers' propensity to accept the invitation depends on cohort quality relative to the school average. While I cannot entirely rule this possibility out, this should not be a major reason for concern. The data shows that most invites are targeted jointly at multiple year groups, which suggests that schools rarely target specific year groups based on their performance. Furthermore, as discussed above, volunteers' propensity to accept invites largely depends on fixed school characteristics such as location, which I control for with school fixed effects, or volunteers' schedules that can be considered as good as random in this exercise.

To control for any persisting endogeneity, I present estimations for four different sets of cohort and individual level controls. In the first specification, I estimate the effect without any controls. Secondly, I choose a set of ad hoc selected covariates, in the third specification I use a post-double LASSO regularisation to select optimal controls from a set of potential covariates, and squares and interactions thereof, and in the fourth one I add school fixed effects to the post-double LASSO controls. For the LASSO, I follow Belloni et al. (2014) and use a double selection procedure that first selects from the set of available controls those that best predict treatment and secondly the outcome. Comparing estimates across specifications will offer some useful insights into the internal validity of the results, and the finding that the estimates do not differ significantly across specifications lends some confidence to the identification strategy. Nonetheless, since data on some important individual level covariates such as free-school-meal eligibility and measures of speaker and teacher quality are missing due to data limitations, I cannot rule out that these are causing bias to the estimations.

1.5 Data Description & Sample Characteristics

1.5.1 Event Data

The first source of data in this analysis is internal data from the educational charity, covering all invites and communications between schools and volunteers from the roll out of the program in summer 2014 to summer 2016. In this invite level dataset, each observation contains an event date and the age group at which the event is targeted. Based on this information, I construct an event level dataset by grouping invites by event date and target age. For each event that was initiated, including those that did not take place, I have information on the number of speakers and their names. Based on the name, I can infer the gender of speakers using the R package *gender* that predicts the gender using historical data. From this information, I can construct the gender share of speakers at an event. I generate a speaker gender dummy equal to one if more than half of the speakers at an event were female and zero otherwise. Therefore, the dummy assuming a value of one implies that a child who participated in this event was more likely to encounter a female speaker than a male speaker, or in case of a child attending multiple talks per event, encountered more female than male speakers.

Subsequently, I transform this event level dataset into a panel dataset by school-cohort. For each school-cohort that graduates from primary school between 2015 and 2016, I observe the events that took place during during year 5 and 6. I do not observe events from earlier years since the program was rolled out in summer 2014. Thus, pupils who graduated in 2015 were only affected by events during year 6 and those who graduated in 2016 in year 5 and 6. For each school-cohort, I construct an aggregate speaker gender dummy based on the most recent event that the cohort attended.

1.5.2 Test Score Data

Secondly, I use data from the NPD on individual level test scores and school level administrative data. These data contain scores in tests in Reading, Maths and Writing in year 2 (aged 6 - 7) and year 6 (aged 10 - 11). The latter serve as the main outcome variables of interest.

In England, primary school pupils take tests in Reading, Writing, and Maths at the end of year 2, in which achievement is measured on a scale from 0 to 30 where 15 refers to the age related expected level of achievement. The test papers are standardised, but assessment is not. The results are used primarily to track overall school performance and do not have a direct effect on pupils educational career. In year 6, pupils take fully standardised tests (papers and assessment) in Reading, Maths (and Science) and non-standardised tests in Writing. Here, achievement is measured on a scale from 1 - 110 in Maths, 0 - 49 in Reading, and Writing on a scale from 0 - 6. Since Writing tests are not standardised, I exclude this subject from the analysis. In the regressions, I standardise test scores to a standard deviation of one and mean zero in the control group.

Similarly to year 2 tests, these tests do not affect pupils directly and they do not serve as entry exams for secondary schools. However, performance of primary schools in England is closely monitored and schools and pupils are under significant pressure to adhere to centralised performance standards. Therefore, these tests are taken seriously and test scores serve as a reliable measure of pupils' academic performance.

1.5.3 Comparison of Experimental Sample to All English Primary Schools

Table 1.1: All schools vs. sample

	All Schools		Sample			
	Obs	Mean	Obs	Mean	Diff(C-T)	OLS P-val
Panel A: 2015 Cohort						
Test Scores						
Maths Year 2	15559	15.33	100	15.20	0.1381	0.59
Reading Year 2	15559	15.37	100	15.11	0.2609	0.33
Writing Year 2	15559	14.03	100	13.89	0.1416	0.57
School Characteristics						
Number of pupils	15605	36.33	100	45.00	-8.6706	0.00
Girls	15605	0.48	100	0.49	-0.0131	0.32
% FSM Eligible	15605	4.12	100	6.83	-2.7040	0.00
Population Density (per hectare)	15605	21.16	100	37.58	-16.4191	0.00
Panel B: 2014 Cohort						
Test Scores						
Maths Year 6	14364	70.87	94	70.78	0.0898	0.92
Reading Year 6	14377	31.18	94	30.32	0.8624	0.03
Maths Year 2	14832	15.29	94	15.34	-0.0516	0.85
Reading Year 2	14832	15.30	94	15.26	0.0358	0.90
Writing Year 2	14832	13.96	94	14.03	-0.0664	0.80
School Characteristics						
Number of pupils	14876	35.14	94	45.06	-9.9189	0.00
Girls	14876	1.52	94	1.52	-0.0014	0.92
% FSM Eligible	14875	4.08	94	6.81	-2.7278	0.00
Population Density (per hectare)	14876	21.31	94	39.43	-18.1257	0.00

Note: Balancing tests based on school level data, with unclustered OLS P-values of the differences in means reported in the last column. *% FSM Eligible* refers to an average across shares of free-school-meal eligible pupils at age 9, 10 and 11 per school. *Population Density (per hectare)* is the average number of people per hectare resident in the postcode district of the school. In panel A, I present school averages for the cohort graduating primary school in 2015. Since parts of this cohort are included in the sample, I do compare in terms of year 6 test scores. Panel B shows school averages for the cohort graduating primary school in 2014, and since no pupils from this cohort are included in the sample, I also compare year 6 test performance.

To get a better sense of the schools in the sample and how they fit into the universe of English primary schools, I compare the schools in the sample to all primary schools in terms of test scores and school characteristics. In the top panel of table 1.1, I compare schools based on the school cohorts that graduated primary school in 2014, and in the bottom panel, the comparison is based on the cohorts that graduated in 2015. In the 2014 comparison, seven sample schools are missing from the statistics because of missing historic data as they converted to a different school type, which makes the school identifier codes incompatible. To show how the comparison varies when all but one sample school is included, I also show the comparison for the 2015 graduating cohorts. Since the year 6 test scores of the 2015 school-cohorts in the sample are endogenous to treatment, I do not compare year 6 test scores in panel B.

Panel A shows that the sample is representative of all primary schools in the country in terms of achievement in year 6 and year 2 tests, except in year 6 Reading performance where the sample performed significantly worse than the rest of English primary schools. As balance tests, I conduct a simple T-test and report P-values of the two-sided test. P-values on the differences in achievement are far larger than conventional significance levels, except for year 6 Reading scores. However, schools in the sample are significantly larger than the average English primary school in terms of number of year 6 pupils. Furthermore, a larger share of pupils are eligible for free school meals, which is a measure of deprivation, and sample schools are located in significantly more densely populated areas.

Panel B shows that the comparison based on the 2015 cohort yields a very similar pattern. The sample is representative in terms of year 2 performance but schools are larger, and in more densely populated and deprived locations. These differences suggest that large schools in deprived, inner-city districts are more likely to be interested in raising pupils' aspirations, and to sign up with the careers talks platform. Whilst these differences do not invalidate the results in this paper, they are important in assessing the external validity of the findings. These may not translate to the average school with fewer pupils, in less densely populated and less deprived locations if the program became mandatory for all schools including those that would not otherwise have joined.

1.5.4 Sample Characteristics and Balance Tests

Table 1.2: Balance tests

	Control		Treatment			
	Obs	Mean	Obs	Mean	Diff(C-T)	OLS P-val
Panel A: Test Scores						
Year 2 Maths	4752	15.32	3723	15.54	-0.2226	0.25
Year 2 Reading	4753	15.31	3723	15.47	-0.1546	0.47
Year 2 Writing	4754	14.05	3723	14.20	-0.1487	0.49
Panel B: Event Data						
Total number events	5069	0.00	3963	1.65	-1.6488	0.00***
Number events year 4	5069	0.00	3963	0.01	-0.0073	0.32
Number events year 5	5069	0.00	3963	0.76	-0.7623	0.00***
Number events year 6	5069	0.00	3963	0.88	-0.8791	0.00***
Total number invites	5069	12.50	3963	50.47	-37.9675	0.00***
Number invites year 4	5069	0.00	3963	0.01	-0.0073	0.32
Number invites year 5	5069	0.00	3963	18.76	-18.7613	0.00***
Number invites year 6	5069	12.50	3963	31.70	-19.1989	0.03**
Panel C: Cohort Characteristics						
Number of pupils	115	44.05	87	45.53	-1.4766	0.64
Girl	112	0.49	85	0.49	0.0006	0.96
% FSM eligible	113	8.23	87	8.84	-0.6052	0.50
Population density (per hectare)	115	32.25	87	43.89	-11.6427	0.04

Note: Balancing tests based on pupil level data, with OLS P-values of the differences in means clustered at the school-cohort level reported in the last column. *% FSM Eligible* refers to an average across shares of free-school-meal eligible pupils at age 9, 10 and 11 per school. *Population Density (per hectare)* is the average number of people per hectare resident in the postcode district of the school.

In table 1.2, I present summary statistics and balance checks for covariates in the pooled sample of the two cohorts graduating primary school in 2015 and 2016 that are included in the analysis. As balance tests, I show P-values from regressions of the form

$$y_i = \alpha + \beta T_i + \varepsilon_i$$

where y_i are the outcomes and T_i is an indicator that turns one when pupil i is in a treatment cohort, standard errors are clustered at the cohort level.

Panel A shows average test scores per school-cohort in year 2 tests in Maths, Reading and Writing. In all three subjects, cohorts in the treatment group have higher average test scores than those in the control group. Whilst these differences exist, the last column shows that they are not significantly different from zero, which implies that they will not bias the estimates in the analysis. However, they might be an indication of other unobservable confounders that I cannot control for and that might threaten the internal validity of the results.

In panel B, I present summary statistics of the event data. The figures show that pupils in treatment cohorts attended on average 1.65 events, 1% of those took place during year 4[2], 46% during year 5, and 53% during year 6. By construction, school-cohorts in the control group did not hold any events. Notwithstanding, control cohorts actively used the charity platform to invite on average 13 speakers. Cohorts in the treatment group sent out 50 invites on average, 1% for events in year 4, 37% in year 5, and 62% for events in year 6. These averages are significantly higher than in control schools as suggested by the P-values of the differences in the last column. In the estimations, I account for these differences by adding the number of invites as a possible control for the post-double LASSO to select from, or controlling for them directly in the specification with ad hoc selected covariates.

Cohort level characteristics are presented in panel C. Treatment and control co-

[2]Since these events took place before the official roll out of the programme and they took place more than two years prior to the test, I exclude them from the analysis.

horts consist of 46 and 44 pupils respectively, and the average gender split is almost identical across the two groups. In the control group, on average 8.23% of pupils per school-cohort are eligible for free school meals compared to 8.84% of pupils in treatment cohorts. The last column suggests that all three differences are indistinguishable from zero, whereas the population density in post code districts of treatment cohorts is significantly higher than for control cohorts. I partially address this issue by adding population density to the list of possible controls in the LASSO or control for it directly in the specification with ad hoc selected covariates. Albeit, this remains a reason for concern if this imbalance suggests that treatment is correlated with other unobservable confounders that cannot be controlled for with the existing covariates.

1.6 Estimation & Methods

1.6.1 Estimation of Main Effects

I estimate four sets of specifications, one without controls, one with a set of ad hoc selected covariates, one in which I control for the post-double selection covariates from the LASSO, and in the last one I add school fixed effects to the post-double LASSO controls. In all regressions, I control for prior attainment to get an estimate of the treatment effect in terms of progress between year 2 and year 6 tests. For each set of controls, I estimate the following set of equations for individual i in cohort c in school s and year t.

$$
\begin{aligned}
y_{icst} &= \beta_0 + \beta_1 T_{cst} + \phi \tilde{y}_{icst} + \mu X_{sct} + \gamma_s + \delta_t + \varepsilon_{icst} && (1.3) \\
y_{icst} &= \beta_0 + \beta_1 T_{cst} + \beta_2 Same\ Gender_{icst} + \phi \tilde{y}_{icst} + \mu X_{isct} + \gamma_s + \delta_t + \varepsilon_{icst} && (1.4) \\
y_{icst} &= \beta_0 + \beta_1 T_{cst} + \beta_2 Girl * Female_{icst} + \beta_3 Boy * Male_{icst} + \\
&\quad + \phi \tilde{y}_{icst} + \mu X_{isct} + \gamma_s + \delta_t + \varepsilon_{icst} && (1.5) \\
y_{icst} &= \beta_0 + \beta_1 Girl * Female_{icst} + \beta_2 Boy * Male_{icst} + \beta_3 Girl * Male_{icst} + \\
&\quad + \beta_4 Boy * Female_{icst} + \phi \tilde{y}_{icst} + \mu X_{isct} + \gamma_s + \delta_t + \varepsilon_{icst} && (1.6)
\end{aligned}
$$

The treatment indicator T_{cst} is defined at the cohort level and is equal to one if the school-cohort that pupil i is part of attended a careers talk. I denote test scores in year 2 exams in the outcome subject as $\tilde{y}_{icst}$, δ_t are year fixed effects and γ_s are school fixed effects. The vector of controls X_{isct} varies depending on whether I estimate the effect with no controls, ad hoc selected controls, post-double LASSO selected controls, or post-double LASSO selected controls plus school fixed effects.

In equation 1.3, I estimate the average treatment effect of a careers talk on all pupils. In equation 1.4, I estimate the treatment effect for pupils whose gender matches that of the majority of speakers in an event compared to pupils where the gender does not match as the coefficient on $Same\ Gender_{icst}$, whereas the coefficient on T_{icst} compares the outcome when the gender does not match compared to the control group, and in equation 1.5 I further disentangle this effect by gender. Lastly, in equation 1.6, I run the fully satiated model to compare in-gender and out-gender matches for boys and girls to the control group.

In equation 1.4 - 1.6, the vector of covariates X_{isct} includes pupil gender. Across all equations, the set of ad hoc selected covariates further includes cohort averages of performance on year 2 tests in Maths and Reading and standard deviations thereof.

Furthermore, cohort shares of pupils aged 10 - 11 who are eligible for free-school-meals, the gender distribution of the cohort consisting of the mean and standard deviation, and the population density at the school's postcode. Lastly, the number of volunteers a school invited for events targeted at year 5 and 6.

1.6.2 Post-double LASSO Regularisation

Due to the large number of potential covariates in my data, I use LASSO regularisation to check the validity of my estimations without any and with ad hoc selected controls. I use the post-double LASSO following Belloni et al. (2014), which selects covariates in two steps distinguishing between "noise reducing" and "identifying" controls. To reduce noise in the estimation, the post-double LASSO first selects the combination of covariates that best predicts the outcome variable. Secondly, it selects the regression model that best predicts the treatment variable. Under the assumption that treatment is exogenous conditional on covariates, this step selects those covariates that achieve identification of the causal effect.

Using the post-double LASSO, I select different sets of controls for Maths and Reading test scores, and for the specification with and without school fixed effects. The regularisation selects covariates from the set of ad hoc controls plus individual level data on test scores in year 2 in Maths and Reading, and pupil gender. Furthermore, the squares of all these variables and pairwise interactions, which leads to a total of 135 possible covariates to select from. Of these, the post-double LASSO selects up to 19 controls for year 6 Maths scores when I control for school fixed effects. A full list of controls is provided in table A.1 of the appendix. The sets of LASSO selected controls are different from the ad hoc selected ones, but, the noise reducing covariates for the two outcome variables, year 6 scores in Maths and Reading, are very similar, while the set of identifying controls is identical by design.

1.7 Results

1.7.1 Main Effects

Table 1.3: Treatment effect Year 6 test scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Maths Year 6								
Treatment	-0.04	-0.08	-0.07	-0.05	-0.06	-0.09*	-0.09	-0.07
	(0.04)	(0.05)	(0.05)	(0.06)	(0.04)	(0.06)	(0.06)	(0.06)
Same Gender Speaker					0.04	0.04	0.04	0.04
					(0.03)	(0.03)	(0.03)	(0.03)
Observations	8295	8295	8299	8323	8295	8295	8299	8323
Panel B: Reading Year 6								
Treatment	-0.02	-0.02	-0.02	-0.04	-0.04	-0.05	-0.05	-0.07
	(0.04)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)
Same Gender Speaker					0.05**	0.06**	0.06**	0.06**
					(0.03)	(0.03)	(0.03)	(0.03)
Observations	8286	8286	8291	8314	8286	8286	8291	8314
Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Girl Dummy					Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes				Yes			
Post-Double LASSO	Yes	Yes			Yes	Yes		
Ad hoc			Yes				Yes	
Without				Yes				Yes

Note: Standard errors are clustered at the school*cohort level and reported in brackets below. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring performance on year 2 tests, which are included in the regression to express treatment effects in terms of progress between year 2 and year 6 tests. Test scores are standardised to a standard deviation of one and mean zero in the control group by gender.

In table 1.3, I present the results from estimating equation 1.3 and 1.4. I estimate each equation with post-double LASSO with and without school fixed effects, ad hoc selected and without controls. In the top panel of the table, I show estimates for regressions of the progress between year 2 and year 6 Maths tests on treatment and in the bottom panel, I show the results for progress in Reading as the outcome variable. The estimates in the first four columns indicate that the average treatment effect of the intervention is negligible across all specifications on both outcome variables. For Maths, the effect sizes range between -0.08 for LASSO without fixed effects and -0.04 for LASSO with fixed effects, whereas for Reading the effect sizes range from -0.02 with LASSO and ad hoc selected covariates, to -0.04 standard deviations without controls. Standard errors in parentheses below the estimates suggest that the effect sizes are not significantly different across specifications.

In the second half of the table's columns, I estimate heterogeneous treatment effects on pupils who attend a talk by a speaker of the same gender compared to those with a speaker of the opposite gender, and to those in the control group. The coefficients on the treatment indicator suggest that the progress of pupils in Maths with an opposite gender speaker is 0.06 to 0.09 standard deviations smaller than in the control group, while the estimates for pupils with a same gender speaker show a positive and insignificant change in progress by 0.04 standard deviations compared to those with an opposite gender speaker. In Reading, pupils with a same gender speaker improve progress compared to those with an opposite gender speaker by 0.05 to 0.06 standard deviations, whereas pupils the treatment effect for pupils with opposite gender speakers relative to the control group is imprecisely estimated at -0.04 to -0.07 standard deviations. As above, the coefficients across specifications in columns 5 - 8 on either outcome are not significantly different from each other.

In table 1.4, I show the results for the same gender treatment effect split by gender in equation 1.5 and the fully satiated model by gender in equation 1.6. Estimations of the first model indicate that girls with female speakers improve their progress in Maths by 0.07 to 0.11 standard deviations compared to girls with male speakers, and in Reading by 0.08 to 0.11 standard deviations. However, estimates of the second model highlight that these effects are largely driven by negative effects of male speakers on girls compared to girls in the control group. In Maths, the negative effects for girls with male speakers range from -0.13 without to -0.15 with controls and in Reading from -0.09 to -0.11 standard deviations, while the performance of girls with female speakers in either subject does not change significantly compared to the control group. Boys with male speakers lose out to the control group by 0.11 to 0.12 standard deviations in Maths progress whereas the decrease in Reading test scores is not significant.

Whilst the effect of treatment is not perfectly causally identified as discussed in section 1.4, the fact that the estimates do not vary significantly between specifications implies that they are robust to including different sets of covariates. This adds confidence to the assumption that the estimates identify a causal effect unless they are biased by confounders that are orthogonal to the covariates in the model. Since all estimates in this chapter bear this risk, any conclusions derived from the results are subject to this concern.

The results indicate an interesting pattern of heterogeneity in the impact of the intervention on progress between year 2 and year 6 tests. The average effect of a careers event is negligible and indistinguishable from zero in both Maths and Reading tests. These results stand in contrast to findings by Jensen (2010) who documents positive effects on the number of years of schooling from correcting beliefs about returns to education through statistical information, and Nguyen (2008) who finds a positive impact on performance, attendance and future enrolment. However, the latter, who tests different ways of transmitting the novel information, namely statistical information, a role model or both, finds too that the role model intervention is less effective than providing statistical information. The fact that the speakers in this intervention did not provide hard statistical facts on actual returns to schooling but instead provided more general information on different careers, may be one explanation for why my findings differ from their main conclusion.

My estimates of heterogeneous effects by same and opposite gender match between pupils and speakers show that an opposite gender match decreases girls' progress in Maths and Reading compared to the control group, whereas boys with male speakers

Table 1.4: Treatment effect Year 6 test scores by gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Maths Year 6								
Treatment	-0.06	-0.09	-0.09	-0.07				
	(0.04)	(0.06)	(0.06)	(0.06)				
Boy*Male Speaker	-0.01	-0.02	-0.03	-0.04	-0.12**	-0.10	-0.11*	-0.10
	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Girl*Female Speaker	0.07*	0.09	0.09	0.11*	0.01	-0.01	-0.00	0.02
	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.07)	(0.08)	(0.07)
Boy*Female Speaker					-0.03	-0.05	-0.04	-0.03
					(0.05)	(0.07)	(0.07)	(0.07)
Girl*Male Speaker					-0.15***	-0.14**	-0.15**	-0.13
					(0.05)	(0.07)	(0.07)	(0.08)
Observations	8295	8295	8299	8323	8295	8295	8299	8323
Panel B: Reading Year 6								
Treatment	-0.05	-0.05	-0.05	-0.07				
	(0.04)	(0.05)	(0.05)	(0.05)				
Boy*Male Speaker	0.02	0.01	-0.01	0.01	-0.06	-0.03	-0.05	-0.05
	(0.04)	(0.06)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	(0.08)
Girl*Female Speaker	0.08**	0.11**	0.11***	0.10**	0.04	0.05	0.05	0.01
	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.07)	(0.06)
Boy*Female Speaker					-0.02	-0.01	-0.01	-0.05
					(0.05)	(0.07)	(0.07)	(0.07)
Girl*Male Speaker					-0.11**	-0.09	-0.11*	-0.11*
					(0.05)	(0.06)	(0.06)	(0.06)
Observations	8286	8286	8291	8314	8286	8286	8291	8314
Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Girl Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes				Yes			
Post-Double LASSO	Yes	Yes			Yes	Yes		
Ad hoc			Yes				Yes	
Without				Yes				Yes

Note: Standard errors are clustered at the school*cohort level and are reported in brackets below. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring performance on year 2 tests, which are included in the regression to express treatment effects in terms of progress between year 2 and year 6 tests. Test scores are standardised to a standard deviation of one and mean zero in the control group by gender.

only decrease progress in Maths but not in Reading. The fact that the effects on the same pupils vary by subject suggests that they are driven by an interaction effect between subject and speakers that not only prevents students from reaping the benefits of the careers talks but lead to deteriorating performance.

1.7.2 Interaction with Gender Norms

To study how the treatment effect varies by regional gender norms, I estimate regressions of the form

$$\begin{aligned} y_{icst} &= \beta_0 + \beta_1 Girl * Female_{icst} + \beta_2 Boy * Male_{icst} + \\ &+ \beta_3 Girl * Male_{icst} + \beta_4 Boy * Female_{icst} + \\ &+ \beta_4 Girl * Male * Women\ Rights_{icst} + \beta_5 Women\ Rights_s + \\ &+ \phi \tilde{y}_{icst} + \mu X_{isct} + \delta_t + \varepsilon_{icst} \end{aligned} \tag{1.7}$$

where $Girl * Male * Women\ Rights_{icst}$ is an interaction between the opposite gender effect on girls with regional data on gender norms, and $Women\ Rights_s$ is the corresponding main effect. A positive sign on the interaction term implies that the negative opposite gender effect on girls is stronger in regions with more traditional gender norms and a negative sign implies that the effect is stronger in regions with more progressive norms.

I use data on gender norms from the 2005 - 2009 wave of the World Values Survey (Inglehart et al. (2014)) for the UK, which is the most recent survey wave with data on the UK. I use two questions from the survey to measure gender norms, one captures the attitude towards women in the labour market, and the second one women's rights more generally. The first question reads *"When jobs are scarce, men should have more right to a job than women"* and the corresponding variable $Work\ Rights$ assumes a value of one if the respondent disagrees and is therefore positively correlated with progressive gender norms.[3] The question behind the second variable, $General\ Rights$, asks *"How essential of a characteristic of democracy*

[3] The variable used in the regression is an average over all responses per region and therefore non-binary.

is it that women have the same rights as men" and is answered on a scale from one to ten, where one refers to "not an essential characteristics of democracy" and ten to "an essential characteristic of democracy".

The smallest geographical unit in the data is the region level, which implies that all schools from one region are assigned the same gender norms.[4] Table 1.5 shows the regional distribution of pupils and schools across the nine regions, and figure 1.1 illustrates the distribution of norms across regions with the number of pupils that each bar represents indicated on top.

Table 1.5: Regional distribution by pupils and schools

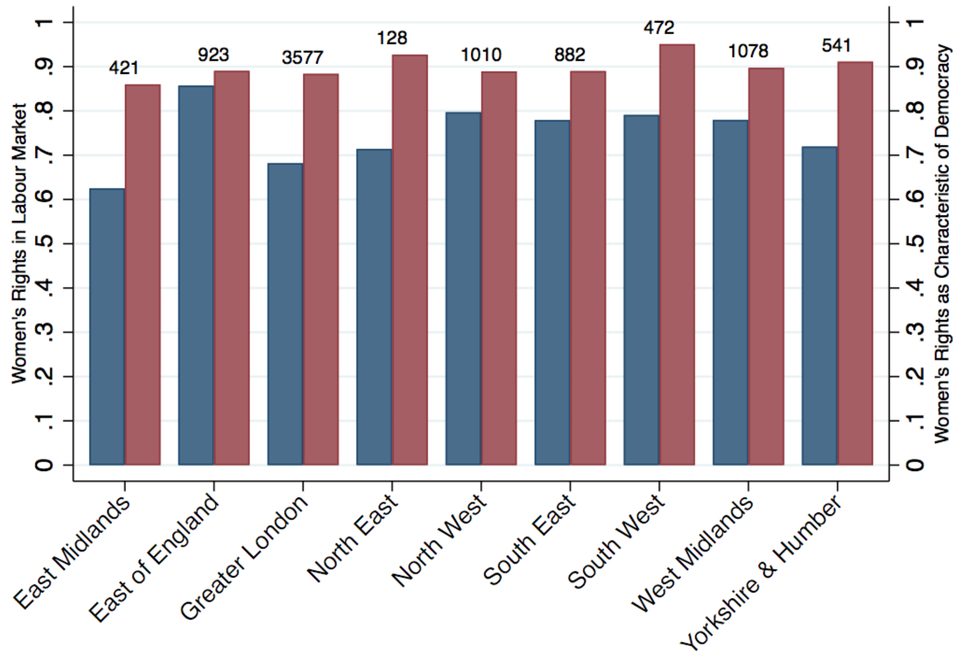
	Pupils			Schools		
	# Pupils	Share	Cum	# Schools	Share	Cum
East Midlands	421	4.66	4.66	6	5.94	5.94
East of England	923	10.22	14.88	12	11.88	17.82
Greater London	3577	39.60	54.48	31	30.69	48.51
North East	128	1.42	55.90	3	2.97	51.49
North West	1010	11.18	67.08	14	13.86	65.35
South East	882	9.77	76.85	10	9.90	75.25
South West	472	5.23	82.07	7	6.93	82.18
West Midlands	1078	11.94	94.01	11	10.89	93.07
Yorks & Humberside	541	5.99	100.00	7	6.93	100.00
Total	9032	100.00		101	100.00	

Note: Regional split by number of pupils and schools based on school postcode.

In table 1.6, I present estimates for the interaction effects of the coefficients of the satiated model with the two measures of gender norms. The estimates for Maths scores imply that the negative effect of male speakers on both girls and boys is amplified in regions with more traditional gender norms, as indicated by the

[4]Since the gender norm data is at the school level, this regression does not include school fixed effects.

Figure 1.1: Gernder norms by region



Note: Measures of gender norms are based on data from the 2005 survey wave of the World Values Survey (Inglehart et al. (2014)). Both y-axes are increasing in more progressive norms. The blue bars represent region averages of question V44 (*Do you agree, disagree or neither agree nor disagree with the following statements? - When jobs are scarce, men should have more right to a job than women.*), where I obtain a binary measure by treating *neither agree nor disagree* as missing values. The red bars represent region averages of question V161 (*Many things may be desirable, but not all of them are essential characteristics of democracy. Please tell me for each of the following things how essential you think it is as a characteristic of democracy. Use this scale where 1 means "not at all an essential characteristic of democracy" and 10 means it definitely is "an essential characteristic of democracy" - Women have the same rights as men.*), where I divide the averages by 10 to obtain a scale from 0.1 to 1. On top of each bar, I indicate the relative size of the region as measured by the number of students.

positive coefficient on the interaction with $Work\ Rights$. These results suggest that the negative effect of male speakers on Maths progress of girls and boys is related to the prevailing local gender norms, or other correlated regional factors like the aptness of male speakers to motivate pupils for Maths related careers, potentially driven by local labour market conditions. For Reading, the estimates point in the same direction but are insignificant, as are the coefficients on the interactions of $Work\ Rights$ with boys and girls with female speakers. The interactions with the other measure of gender norms, $General\ Rights$, are all insignificant.

Table 1.6: Treatment effect on girls with male speakers interacted with gender norms

	Maths Year 6						Reading Year 6					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Boy*Male Speaker	-1.70**	-1.62**	-1.83**	-0.78	-0.65	0.21	-0.88	-0.86	-1.06	1.82	2.42	2.83
	(0.69)	(0.68)	(0.73)	(2.37)	(2.35)	(2.53)	(0.82)	(0.87)	(0.87)	(2.30)	(2.29)	(2.51)
Girl*Female Speaker	-0.21	0.01	0.16	-1.53	-0.73	0.00	-0.11	-0.32	-0.18	-0.17	0.28	0.97
	(0.82)	(0.84)	(0.88)	(3.75)	(3.78)	(4.10)	(0.66)	(0.67)	(0.63)	(2.38)	(2.48)	(2.47)
Boy*Female Speaker	0.01	0.23	0.34	-0.55	0.30	1.12	-0.12	0.03	0.13	0.05	1.08	1.85
	(0.78)	(0.80)	(0.84)	(3.20)	(3.34)	(3.77)	(0.69)	(0.75)	(0.70)	(2.99)	(3.23)	(3.23)
Girl*Male Speaker	-1.83**	-1.80**	-2.06**	-0.48	-0.14	0.65	-0.80	-0.77	-0.96	0.80	0.95	1.09
	(0.76)	(0.75)	(0.87)	(3.06)	(3.01)	(3.47)	(0.67)	(0.70)	(0.74)	(2.47)	(2.49)	(2.73)
Boy*Male Speaker*Work Rights	2.12**	2.01**	2.31**				1.13	1.08	1.34			
	(0.93)	(0.92)	(0.97)				(1.10)	(1.19)	(1.19)			
Girl*Female Speaker*Work Rights	0.23	-0.07	-0.29				0.19	0.49	0.22			
	(1.12)	(1.14)	(1.18)				(0.90)	(0.92)	(0.85)			
Boy*Female Speaker*Work Rights	-0.13	-0.42	-0.60				0.12	-0.08	-0.30			
	(1.09)	(1.10)	(1.15)				(0.96)	(1.04)	(0.96)			
Girl*Male Speaker*Work Rights	2.24**	2.19**	2.56**				0.94	0.88	1.14			
	(0.98)	(0.99)	(1.11)				(0.87)	(0.92)	(0.96)			
Boy*Male Speaker*General Rights				0.08	0.06	-0.03				-0.21	-0.28	-0.32
				(0.26)	(0.26)	(0.28)				(0.26)	(0.26)	(0.28)
Girl*Female Speaker*General Rights				0.17	0.08	0.00				0.02	-0.03	-0.11
				(0.42)	(0.42)	(0.46)				(0.27)	(0.28)	(0.28)
Boy*Female Speaker*General Rights				0.06	-0.04	-0.13				-0.01	-0.12	-0.21
				(0.36)	(0.38)	(0.42)				(0.34)	(0.36)	(0.36)
Girl*Male Speaker*General Rights				0.04	-0.00	-0.09				-0.10	-0.12	-0.13
				(0.34)	(0.34)	(0.39)				(0.28)	(0.28)	(0.31)
Observations	8295	8299	8323	8295	8299	8323	8286	8291	8314	8286	8291	8314
Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Girl Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
General Rights				Yes	Yes	Yes				Yes	Yes	Yes
Work Rights	Yes	Yes	Yes				Yes	Yes	Yes			
Post-Double LASSO	Yes			Yes			Yes			Yes		
Ad hoc		Yes			Yes			Yes			Yes	
Without			Yes			Yes			Yes			Yes

Note: Standard errors are clustered at the school*cohort level and are reported in brackets below. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring performance on year 2 tests, which are included in the regression to express treatment effects in terms of progress between year 2 and year 6 tests. Test scores are standardised to a standard deviation of one and mean zero in the control group by gender. *Work Rights* refers to a binary variable based on question V44 from the 2005 survey wave of the World Values Survey (Inglehart et al. (2014)) (*Do you agree, disagree or neither agree nor disagree with the following statements? - When jobs are scarce, men should have more right to a job than women.*), where I obtain a binary measure by treating *neither agree nor disagree* as missing values. The variable *General Rights* is based on question V161 of the same survey (*Many things may be desirable, but not all of them are essential characteristics of democracy. Please tell me for each of the following things how essential you think it is as a characteristic of democracy. Use this scale where 1 means "not at all an essential characteristic of democracy" and 10 means it definitely is "an essential characteristic of democracy" - Women have the same rights as men.*).

1.7.3 Quantile Regressions

Table 1.7: Quantile regressions of fully satiated model

	Mean	10th	20th	30th	40th	50th	60th	70th	80th	90th
Panel A: Maths Year 6										
Boy*Male Speaker	-0.12**	-0.02	-0.19***	-0.16**	-0.17***	-0.14*	-0.17**	-0.10	-0.08	-0.11*
	(0.06)	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)
Girl*Female Speaker	0.01	0.04	-0.04	0.01	0.07	0.05	-0.02	0.02	0.02	-0.01
	(0.05)	(0.08)	(0.08)	(0.07)	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)
Boy*Female Speaker	-0.03	0.01	-0.08	-0.04	-0.02	-0.01	-0.04	-0.03	-0.03	-0.00
	(0.05)	(0.09)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.05)	(0.05)
Girl*Male Speaker	-0.15***	-0.07	-0.25***	-0.19**	-0.18***	-0.15**	-0.16**	-0.15**	-0.08	-0.12**
	(0.05)	(0.09)	(0.08)	(0.08)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.06)
Observations	8295	8295	8295	8295	8295	8295	8295	8295	8295	8295
Panel B: Reading Year 6										
Boy*Male Speaker	-0.06	-0.00	-0.06	-0.02	-0.10	-0.12	-0.17**	-0.14*	-0.09	-0.06
	(0.06)	(0.09)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)	(0.10)	(0.11)
Girl*Female Speaker	0.04	0.03	0.03	0.04	0.03	0.02	-0.02	0.01	0.08	0.07
	(0.05)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)
Boy*Female Speaker	-0.02	0.01	-0.06	-0.05	-0.02	-0.03	-0.08	-0.01	0.02	0.02
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
Girl*Male Speaker	-0.11**	0.02	0.02	-0.05	-0.19***	-0.20***	-0.24***	-0.20**	-0.16*	-0.09
	(0.05)	(0.09)	(0.09)	(0.07)	(0.07)	(0.06)	(0.07)	(0.08)	(0.08)	(0.10)
Observations	8286	8286	8286	8286	8286	8286	8286	8286	8286	8286
BL / Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Girl Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Post-Double LASSO	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors are clustered at the school*cohort level and reported in brackets below. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring performance on year 2 tests, which are included in the regression to express treatment effects in terms of progress between year 2 and year 6 tests. Test scores are standardised to a standard deviation of one and mean zero in the control group by gender.

Finally, in table 1.7, I present quantile regressions for the fully satiated model in equation 1.6.

The estimates show that the effect of male speakers on boys in Maths is located in the middle of the distribution, decreasing progress of boys in the second to sixth decile by 0.14 to 0.19 standard deviations. For girls, the effects are even more spread out and affect nearly all parts of the distribution except the first and eighth decile. The effect sizes range from -0.12 standard deviation in the ninth decile to -0.25 in the second, where this appears to represent a trend that the girls at the bottom of the distribution are most affected by the treatment. In Reading, the effect on boys is concentrated around the sixth and seventh decile, and is highest in the sixth decile with -0.17 standard deviations. For girls, the negative effects on Reading scores are most prevalent in the top half of the distribution, ranging from -0.16 in the ninth to -0.24 standard deviations in the sixth decile.

1.8 Discussion

1.8.1 Compatibility of Model and Empirical Results

The main hypothesis of the model suggests that careers talks increase pupils' effort by providing information on higher utility job outcomes than those they are already familiar with. The new information makes increased effort worthwhile as it increases the probability of achieving a higher utility job. The empirical analysis in this paper does not provide any evidence in favour of this channel, at least not as the dominating mechanism. Compared to the control group, pupils with careers talks perform the same on average or worse in the case of girls with male speakers on both subjects and boys with male speakers in Maths.

The model illustrates two channels that may interfere with the positive effect of new information. First, careers talk speakers might trigger additional stress in pupils that leads to a depletion of resources and an increase in the cost of effort during the subsequent test, which reduces performance. According to the stereotype threat literature in social psychology, stress can be induced through priming on a negative stereotype (Steele and Aronson (1995), Spencer et al. (1999)). Insofar as careers talks by male speakers constitute a negative stereotype, this might be an explanation for the negative effect on girls that is visible in the data.

The estimations in section 1.7.2, where I interact the negative effect of male speakers on test scores with regional data on gender norms, appear to support the hypothesis that stereotype threat is at least partially responsible for explaining the negative effect on girls. The estimates show that the negative effect on girls and boys is stronger in regions with more traditional views on women's role in the labour force. However, since the data on gender norms is from 2005 - 2009, it may not accurately capture current gender attitudes and the results should be interpreted with caution. Furthermore, the analysis can also not rule out other correlated mechanisms, such as regional variation in the quality of male speakers.

Since the effect on boys is also stronger in regions with more traditional norms, stereotype threat triggered by male speakers is unlikely to be the only driver of the effect. Lastly, since the stereotype threat literature documents the negative effect from priming on a stereotype only for performance on tests that take place immediately afterwards, the stereotype threat theory can explain the results only if the girls were reminded in some way reminded of the negative associations with the male speaker shortly before or during the test, which might arise if speakers mentioned year 6 SATs during the talk.

Since the stereotype literature suggests that the effect targets test performance but not long term effort, a way of testing for this channel would be to measure effort of pupils after the talk, or to follow pupils over time and test whether the negative effect wears out as would be expected if long term effort is not affected. I cannot draw a final conclusion on this channel, as I do not observe effort or performance on later tests in the data.

The model also suggests that the positive effect of novel information could be reversed if the pupils consider the speaker unattainable, which reduces the subjective probability of achieving the higher utility outcome, and thereby the optimal level of effort exerted by pupils. This mechanism might be partly responsible for the negative effect of male speakers. Whilst I do not observe measures of speakers' professional success to test for the interaction between the negative effect and distance in achievement explicitly, the quantile regressions in table 1.7 offer some insights. If lack of perceived attainability were the channel at play, the negative effect of male speakers would be largest for pupils at the bottom of the performance distribution, as their achievement is furthest away from that of high achieving speakers. However, the estimates do not perfectly match that prediction, instead they imply that the effect is driven by boys in the second to sixth decile, and girls across almost the entire distribution. Therefore, further research is necessary to examine this relationship more closely.

1.8.2 Imbalance between Treatment and Control Group

Since the identification strategy in this study does not perfectly rule out endogeneity in the treatment allocation, a remaining concern is that the estimates are partly influenced by selection or other omitted variable bias. As I discuss above, there is scope for endogeneity in the treatment variable due to selection at the school-cohort level. Any unobservable cohort level characteristics that make the cohort more or less likely to host a careers event and that also affect performance of pupils will bias the estimates. For instance, teacher quality that is different from the school average might affect test scores and also increase the likelihood of being in the treatment group if the teacher's good practices include organising careers talks. Since I do not observe these potential differences between treatment control group, I cannot control for their influence in the estimations.

Moreover, table 1.2 shows that even though all schools in the sample have hosted careers talks for some age group, schools in the control group have invited significantly fewer speakers to give talks for year 5 and 6 during the sample period. I can control for these observable confounders in the estimations, but the existing imbalances might be indicative of differences in unobservable characteristics that I cannot control for. Lastly, treated school-cohorts are more often in more densely populated postcode areas than untreated cohorts. As before, I can control for population density directly but not for any related unobservable factors such as infrastructure and labour market opportunities.

Even though the setup in this paper does not lend itself to dealing with these remaining endogeneity concerns directly, the fact that the estimates are stable across specifications with different sets of controls adds confidence to the assumption that careers talks play a major role in driving the results and shows that the estimated impact is robust to including these controls. Unless none of the sets of controls included any omitted variables, and all the controls are uncorrelated with unobservable confounders, the stability of estimates suggests that there is no significant scope for omitted variable bias in the estimations. However, due to the large number of potential unobservables at the speaker, cohort and individual level that arises from the data limitations, I cannot conclude from this that I identify the causal effect of careers talks on pupils.

Furthermore, as far as any conclusions that are based on comparisons of the differential effects on Maths and Reading are concerned, selection bias is less likely to be an issue unless it is correlated with subject specific performance of the cohort rather than average performance across subjects. Yet, this might arise if the class teacher who initiates the event is also better at teaching one subject than the other, and therefore yields higher average test scores for their pupils. Alternatively, children from more advantaged backgrounds for whom the careers talk represents less novel information may do systematically better in one subject than the other.

Future research could address some of these issues with richer data on student and cohort level characteristics and speaker quality. In the next chapter, I conduct an experiment in which treatment of a similar intervention is randomly allocated to schools, and speaker quality is uniform, which rules out the endogeneity issues in this discussion.

1.9 Conclusion

In this paper, I study the impact of careers talks in primary schools on pupils' performance on year 6 standardised tests. The careers talks are aimed at providing pupils with information about a variety of careers to expand their knowledge about career options beyond those that they observe in their social network. This information is intended to raise the effort that pupils invest into education and thereby foster social mobility through increased career choice.

I present a theoretical framework in which I model the effect of careers talks as a higher utility job outcome in pupils' utility function whose probability of materialising increases with effort. Therefore, the new information makes it worthwhile for pupils to invest more effort into education as this increases the probability of achieving the higher utility job. I then extend this standard choice model by adding two channels that can lead to changes in the optimal behaviour of pupils.

First, I allow for careers talks to induce additional stress and distraction that act as an increase in the cost of effort during tests if the speakers evoke negative emotions in pupils. Such distractions can be triggered through priming on a negative stereotype regarding pupils' ability, as suggested by the literature on stereotype threat, which predicts that when negative stereotypes are made salient, pupils who are subject to the stereotype will be distracted by the pressure not to fulfil the stereotype.

A second channel that can undo the positive effect of careers talks in the model captures the negative effect on pupils when the high utility job outcome introduced in the careers talk is perceived as too far from their own set of attainable options. If the distance between the new reference point and pupils' perceived ability is too big, pupils discount their subjective probability of achieving it, which will lead to lower optimal investment of effort into education. Overall, the model predicts that pupils' performance in year 6 tests will improve in response to careers talks, unless one or both of the two negative channels dominate the main effect.

To study this relationship, I use data on careers talks in English primary schools during the school years 2014/15 and 2015/16, and individual level data on performance in year 6 tests following the event. Since treatment is at the cohort level, I can exploit within and between school variation in treatment assignment to estimate the effect of careers talks on pupils. I use different sets of individual and school level covariates to limit the scope for endogeneity in the estimations further. While I find that the results are consistent across specifications, I cannot entirely rule out threats to internal validity that arise from selection and omitted variables bias.

The empirical results show no significant average treatment effect on progress between year 2 and year 6 tests in either Maths or Reading. Yet, I find a significant, negative effect on Maths and Reading performance for girls with a male speaker and on Maths scores of boys with male speakers. The estimations further suggest that negative effect in Maths for both girls and boys intensifies in measures of more traditional regional gender norms about men's dominance in the labour market. Both effects are also fairly equally spread out across the distribution as shown in quantile regressions.

These findings imply that any positive effect from novel information on career options is outweighed by other negative effects of careers talks. To a certain extent is the effect on girls compatible with theoretical predictions regarding the effect of stereotype threat, which reduces the performance of pupils who are primed on a negative stereotype through the careers talks, for instance that women are less apt for a successful career than men. The result that the negative effect on girls with male speakers is larger in regions with more traditional gender norms would further support this hypothesis. However, I cannot rule out that the observed effect is driven by variation in regional speaker quality that is correlated with gender norms rather than with norms directly. Furthermore, since the negative effect on boys' performance also increases in regions with more traditional norms, and men are less likely to represent a negative stereotype for boys, stereotype threat triggered by male speakers is unlikely to be the sole driver of the effect.

The negative effects might also be driven by pupils to whom the male speakers seem unattainable and who in response decrease their subjective probability of a high job outcome, which reduces their optimal level of effort. According to the theoretical predictions, this effect will be largest for low performing pupils, but the quantile regressions substantiate this hypothesis only to some extent. Moreover, since speakers vary depending on the location of the school, I cannot rule out that this heterogeneity in effects is driven by differences in the quality of speakers that are correlated with regional gender norms.

These findings suggest avenues for future research. First, the remaining scope for endogeneity in the estimations implies that future research with richer data on individual level characteristics of the students including their family background and parental attitudes regarding gender and education, as well as speaker quality is needed to verify the results. This information would also allow for more insightful analysis of heterogeneous treatment effects as gender attitudes would be more accurately measured.

Additional analyses are necessary to determine the longer term effects of careers talks on course choice in secondary school and higher education as these outcomes will allow the determination of how careers talks affect actual career choices. It would be particularly interesting to examine career choices of pupils with reference to parents' careers to better gauge whether careers talks improve social mobility. Studying long term effects would also allow further insights into the mechanisms at play in this paper, as the negative effect on girls, if it is driven by stereotype threat, should affect performance only temporarily and not long term effort and it should thus wear off over time.

Acknowledgements This project has greatly benefited from discussions with Catia Batista, Tim Besley and Svetlana Chekmasova, and seminar participants at the Nova SBE research group. I am indebted to Anthony Mann and Alasdair Smith, and the entire Education & Employers Taskforce for providing me with the data for this project. I gratefully acknowledge funding for this project from the Fundaçao para a Ciencia e a Tecnologia (FCT).

Appendices

A Additional Tables

Table A.1: List of controls in estimations

Ad-hoc	Post-double LASSO			
	Without School FE		With School FE	
	Maths	Reading	Maths	Reading
mean(Y2 Maths)	mean(Y2 Maths)$_y$	mean(Y2 Maths)$_y$	(Y2 Maths)^2_y	Y2 Maths$_x$
mean(Y2 Reading)	(Y2 Maths)^2_y	(Y2 Maths)^2_y	mean(Girls)*Y2 Maths$_y$	(Y2 Reading)^2_x
sd(Y2 Maths)	mean(Y2 Maths)*sd(Y2 Maths)$_{yx}$	mean(Y2 Maths)*sd(Y2 Maths)$_y$	Y2 Maths*Y2 Reading$_y$	mean(Y2 Maths)*Y2 Maths$_x$
sd(Y2 Reading)	mean(Y2 Maths)*sd(Y2 Reading)$_{yx}$	mean(Y2 Maths)*sd(Y2 Reading)$_y$	# Invites Y4$_x$	sd(Y2 Maths)*Pupil Gender$_x$
FSM 10-y/o	mean(Girls)*Y2 Maths$_y$	FSM 10y/o*Pupil Gender$_y$	mean(Y2 Maths)^2_x	FSM 10 y/o*Pupil Gender$_x$
FSM 11-y/o	Pop. Density*Y2 Maths $_y$	Y2 Maths* Y2 Reading$_y$	sd(Y2 Reading)^2_x	Y2 Maths*Y2 Reading$_x$
# Invites Y6	Y2 Maths*Y2 Reading$_y$	mean(Y2 Maths)*sd(Y2 Maths)$_x$	(# Invites Y4)^2_x	mean(Y2 Maths)*sd(Y2 Maths)$_y$
# Invites Y5	mean(Y2 Maths)*# Invites Y4$_x$	mean(Y2 Maths)*sd(Y2 Reading)$_x$	mean(Y2 Maths)*mean(Girls)$_x$	mean(Y2 Maths)*sd(Y2 Reading)$_y$
# Invites Y4	mean(Y2 Reading)*# Invites Y4$_x$	mean(Y2 Maths)*# Invites Y4$_x$	mean(Y2 Reading)*sd(Y2 Maths)$_x$	mean(Y2 Maths)*# Invites Y4$_y$
mean(Girls)	sd(Y2 Maths)*# Invites Y6$_x$	mean(Y2 Reading)*# Invites Y4$_x$	mean(Y2 Reading*# Invites Y4)$_x$	mean(Y2 Reading)*# Invites Y4$_y$
Population Density	sd(Y2 Maths)*# Invites Y4$_x$	sd(Y2 Reading)*# Invites Y6$_x$	mean(Y2 Reading)*Pop. Density$_x$	sd(Y2 Reading)*# Invites Y6$_y$
	FSM-Eligibe 10-y/o*# Invites Y4$_x$	sd(Y2 Reading)*# Invites Y4$_x$	sd(Y2 Maths)*# Invites Y4$_x$	sd(Y2 Reading)*# Invites Y4$_y$
	# Invites Y6*# Invites Y4$_x$	FSM 10y/o*# Invites Y5$_x$	sd(Y2 Reading)*FSM 10-y/o$_x$	FSM 10 y/o*# Invites Y5$_x$
	# Invites Y4*mean(Girls)$_x$	FSM 10y/o*# Invites Y4$_x$	sd(Y2 Reading)*# Invites Y4$_x$	mean(Girls)*Pop. Density$_x$
	mean(Girls)*Pop. Density$_x$	# Invites Y6*# Invites Y4$_x$	sd(Y2 Reading)*# Invites Y4$_x$	# Invites Y6*# Invites Y4$_x$
	FSM-Eligibe 10-y/o*# Invites Y5$_x$	# Invites Y4*mean(Girls)$_x$	FSM 10-y/o*# Invites Y5$_x$	# Invites Y4*mean(Girls)$_x$
		mean(Girls)*Pop. Density$_x$	FSM 10-y/o*# Invites Y4$_x$	
			# Invites Y6*# Invites Y5$_x$	
			# Invites Y4*mean(Girls)$_x$	

Note: mean$(\cdot)$ refers to school-cohort means, and $sd(\cdot)$ to standard deviations of the school-cohort means.
y: post-double LASSO selected covariates that predict the outcome variable (*"noise reducing controls"*)
x: post-double LASSO selected covariates that predict the treatment variable (*"identifying controls"*)

Chapter 2

The Impact of Role Models on Aspirations and Educational Achievement in UK Primary Schools

Abstract
The achievement gap of pupils from disadvantaged socioeconomic backgrounds in developed countries is large and persistent. In this paper, I focus on the influence of low aspirations, arising due to a lack of exposure to role models, in reducing the demand for education from low-income families. To study how role models impact aspirations and in turn educational achievement, I run a randomised controlled trial in the year 6 classes of 48 English primary schools. I generate exogenous variation in the exposure of pupils to role models by designing a video intervention with STEM professionals to take place in class, and I analyse the role model channel explicitly by varying the gender of the speaker in the video. I test the impact of the intervention on survey measures of aspirations and performance on end-of-year tests in Maths and Reading. The estimations show a positive effect of the treatment on aspirations and educational achievement when speakers and pupils are of the same gender. This highlights the importance of same gender role models for the effective transmission of information, and shows that pupils in order to process the information and channel it into productive effort need to be able to relate to the speakers via gender.

JEL codes: C93, I24, I26, J16
Keywords: educational inequality, aspirations, role models, STEM

2.1 Introduction

In 2017 in the UK, the achievement gap for pupils eligible for free-school-meals relative to non-eligible pupils was 9.5 months at the end of primary school (Education Endowment Fund (2018)). In the OECD-EU[1], the average gap in PISA (Programme for International Student Assessment) Maths scores at age 15 was 20% in 2016 between pupils in the bottom and top quarter of an index of economic, social and cultural status (OECD (2017)). The achievement gap by socioeconomic background is associated with constraints to the supply of and the demand for quality education.

In this paper, I focus on the influence of low aspirations, arising from a lack of exposure to role models, in reducing the demand for education from low-income families. Exposure to role models in the family or social network is immediately correlated with parents' education and neighbourhood characteristics, and if exposure to role models affects educational achievement and schooling choice, the lack thereof might constitute a barrier to social mobility. Role models will matter for performance if they affect aspirations, which are the achievement related goals that pupils set for themselves as reference points, and aspirations in turn affect educational achievement. If this relationship holds, then inequality in role models reinforces educational inequality.

To study how role models impact aspirations and in turn educational achievement, I ran a randomised controlled trial (RCT) in year 6 classes of 48 English primary schools in the spring of 2017. I generate exogenous variation in the exposure of pupils to role models, by designing a video intervention to take place in class rooms of year 6 pupils, which is the final year of primary school. The video intervention

[1] Includes Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Luxembourg, Netherlands, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, United Kingdom

consists of two parts: first, an animated video illustrating the way Maths is useful in a number of careers ranging from graphic design to robotics engineering, and secondly, a TED-like talk by a speaker with a background in STEM (Science, Technology, Engineering, Mathematics) as the potential role model. To disentangle the effect of neutral transmission of information from a role model effect, the speaker in the second part of the video varies by gender, which allows examination of how the treatment effect intensifies when the genders of pupil and speaker match. The second part of the video has six different versions, three versions feature male and three feature female speakers of different ethnicities. I measure the effect of the video, and the effect of a gender match between speaker and pupil, via survey measures of aspirations immediately after the intervention in March, and test scores in nationally standardised tests (*year 6 SATs*) 1.5 to 2 months later in May 2017.

In order to measure the effect of the treatment, survey data was collected in two rounds as detailed in figure 2.2. The baseline survey took place during December 2016 and January 2017, and the endline survey was collected in treatment schools after the intervention in March 2017 and in control schools during the same time without the preceding intervention. From the surveys, I obtain measures of students' educational, career and private life aspirations and their mindset towards learning as an indicator of their Growth Mindset. The test score data to measure the impact of the intervention on educational achievement is based on administrative school data. I use progress between year 2 (age 6 -7) mandatory assessments and year 6 SATs (age 10 -11) in English and Reading as the main outcome measure of educational achievement. SATs are nationally standardised tests in Maths, Reading, Grammar, Punctuation and Spelling (GPS) and (non-standardised) Writing that all year 6 pupils take in their final year of primary school, and while they do not serve as entry-exams into secondary school, they are an important monitoring tool for the government to track school performance and are therefore taken very seriously by teachers and pupils. This implies that performance on these tests is a meaningful outcome measure of educational achievement of pupils.

Identification in this paper relies on random allocation of schools into treatment and control group. The unit of randomisation is at the school level, so that an entire school is either randomly allocated to receive the treatment or to be part of the control group. This design ensures that the SUTVA (stable unit treatment value assumption) is satisfied as it rules out spillovers from treatment group to control. No two pupils from the same school are allocated to different groups, and schools are sufficiently far apart to ensure no spillovers between schools as well. Furthermore, no pupils switched schools between treatment assignment and intervention. Balancing tests show that treatment and control groups are not significantly different based on survey characteristics at baseline and prior attainment.

The survey and test score samples differ in their composition: survey data from both rounds is available for forty schools, whilst test score data is available for thirty schools. After the data collection at baseline and randomisation of 48 schools, eight schools did not return the endline survey and are thus excluded from the final survey data sample. However, this sample is still balanced at baseline. Moreover, two of the schools that did not return the endline survey made test scores available and are included in the test score sample. Twelve of the schools that returned both surveys, did not report test scores, and in most cases, the non-reporting of data is due to staff turnover during the school summer holidays. Since schools received the test scores in July 2017, shortly before the holidays, many of them did not send them to me until after the break, and in instances where the main contact or head teacher left the school during this period, schools were reluctant to share the data. Notwithstanding, the sample of thirty schools for which test score data is available is balanced on prior attainment, and baseline survey data when available, which is the case for all but two schools. Only the average ethnic composition of schools is marginally significantly different between treatment and control groups in this sample and I will control for ethnicity in all regressions.

The analysis of gender differences in baseline survey measures and prior attainment shows that boys have lower degree aspirations but aspire to a more difficult dream job than girls and are more likely to embrace a Growth Mindset than girls. Boys also have more male centric gender stereotypes regarding professional success. Interestingly, boys have higher aspirations for their performance in Maths tests even though they have the same prior Maths attainment on average. In English, boys' aspirations are lower than girls', and so is prior attainment. These gender differences suggest that boys are overall more confident, especially in Maths, which might partly explain the the stronger treatment effect on boys' Maths scores.

Whilst the analysis does not provide evidence for a significant impact on pupils in the treatment group on average, I find a positive effect of treatment on aspirations and educational achievement when speakers and pupils are of the same gender. Boys who watch a video with a male speaker have 0.19 standard deviations higher progress between year 2 and year 6 performance on Maths tests, 0.7 standard deviations higher degree aspirations and the self reported difficulty of their dream job is 1.14 standard deviations higher than of boys with female speakers, conditional on baseline values. Girls with female speakers are 10% more likely to have a Growth Mindset than those with male speakers, and boys and girls with same gender speakers have higher aspirations for their performance on Maths tests than those with opposite gender speakers.

The fact that treatment has an effect only if the speaker is of the same gender as the pupil, highlights the importance of same gender role models for the effective transmission of information. This shows that pupils in order to process the information and channel it into productive effort need to be able to relate to the speakers via gender.

To compare how the same gender treatment effect differs between pupils with very high achieving same gender peers and those whose highest achieving classmate's performance is average compared to all other top students, I interact the same gender treatment with the test score of the highest achieving same gender peer in the class. I find that the effects on test scores of girls with female speakers vary with the performance of their female classmates as the treatment effect is smaller for girls with very high achieving peers than for girls with less high achieving girls, which suggests that female speakers and high achieving peers have substitutive effects on performance. The treatment effect of boys does not vary by peer performance.

To the best of my knowledge, this is the first paper to provide experimental evidence on the link between role models, aspirations and educational achievement in primary school pupils. The survey data collected for this experiment before and after the intervention shed light on the changes in non-cognitive skills such as aspirations that accompany and potentially drive the effects on test scores. The design allows to study the importance of relatability via gender for both boys and girls, thereby adding to the existing evidence on the impact of female role models on girls. The results illustrate that the findings for female role models do not carry over straightforwardly to boys and the widening gender gap in achievement in favour of girls highlight the importance of increasing boys' engagement at school. As the controlled video intervention ensures that the quality of speakers is uniform across schools and regions, I can study these effects in isolation from any speaker specific influences that might affect the results.

In Economics, the relevance of noncognitive skills, such as perseverance, motivation and self-control, in addition to cognitive skills are recognised in work by Heckman (2000) and Carneiro and Heckman (2003) as key to driving schooling and earnings outcomes. These studies also emphasise that both cognitive and non-cognitive skills are malleable through family and the social environment especially at a young age. See also Cunha and Heckman (2006) for a review of the literature on early childhood interventions.[2]

The concept of non-cognitive skills and aspirations in particular, as a cause and consequence of socioeconomic inequality was introduced by Appadurai (2004) and Ray (2006). The authors argue that individuals form preferences based on experiences of other people in their cognitive neighbourhood, and they call these socially dependent preferences aspirations. Subsequent papers by Genicot and Ray (2017) make further theoretical contributions regarding the circular relationship between socially determined aspirations, income and the income distribution. In Political Economy, Besley (2016) offers a theoretical perspective on the role of socially determined aspirations in influencing income inequality through voting for redistribution in a framework in which aspirations evolve endogenously.

The broader idea that individuals learn behaviour through observation of the people in their surroundings, and that learning is more likely to happen if the model is perceived as similar to themselves, goes back to the formulation of the Social Learning Theory by social psychologist Bandura (1977, 1986). In experimental tests of the theory, which I return to in section 2.2.1, he confirms the impact of role models in the transmission of information and behaviours of children.

In Economics, most empirical studies combine the effect of role models with provision of targeted information to affect aspirations and behaviour. Bernard et al. (2014) deploy a randomised video intervention in Ethiopia to show that successful entrepreneurial role models increase farmers' aspirations, savings behaviour and investments into children's education, and Batista and Seither (2018) find that in-

[2]Note that interventions at primary school age are not considered early childhood interventions in this literature.

creasing aspirations of micro-entrepreneurs in Mozambique in combination with goal setting and business skills training increases profits by 40% relative to the control group.

Focusing on the relationship between role models and educational outcomes, Dee (2005) shows that teachers perceive students of their own gender and ethnicity as being of higher ability than demographically different students, and that these gender and ethnicity dynamics between teachers and pupils contribute to the demographic attainment gaps in the US. Similarly, Lusher et al. (2018) find positive effects on performance for Asian / Non-Asian undergrad Economics students who are assigned to a TA of a similar ethnicity. Beaman et al. (2012) exploit a natural experiment in India to show that female role models affect girls' career aspirations and educational attainment, and Breda et al. (2018) show that female STEM role models raise the probability of French female high school students applying and being admitted to selective Science programmes by 30%.

I add to this empirical literature by focusing on the effect of role models and aspirations of primary age pupils before they make any irreversible education choices, which implies that policy interventions could have long lasting impact on educational outcomes. Furthermore, the design in this study allows to estimate the effect of same gender role models on boys and girls separately, while previous studies focus primarily on female role models.

Using targeted information to influence behaviour, as this study does through video screenings, is not a new concept and has been studied in a variety of contexts.[3] Nguyen (2008), Jensen (2010), and Attanasio and Kaufmann (2014) show positive effects of information on labour market returns and returns to education on edu-

[3]See the Public Health literature (Prochaska et al. (1993), Shiffman et al. (2000), Campbell et al. (1994), Marcus et al. (1998)) for positive impacts of information on smoking cessation, dieting and exercising

cational attainment and test scores in Madagascar, the Dominican Republic and Mexico. Furthermore, other schooling decisions such as college major choices are also influenced by labour market information. Hastings et al. (2015) use data from Chile to show that students have inaccurate beliefs about major specific earnings, Baker et al. (2017) find that low-income, low-achieving students who apply to low-earning degrees in California (USA) overestimate salaries by more than 100%, and Reuben et al. (2017) document that girls have lower earnings expectations than boys. All three studies find that novel labour market information affects college major choice. A recent study by Lergetporer et al. (2018) analyses the impact of providing information about the cost and returns to university on educational aspirations of parents in Germany, and find that the information increases educational aspirations of parents with and without university degree. However, since the effect is larger on parents with university degree, they conclude that the inequality in information does not explain the socioeconomic gap in educational aspirations. Related research on the role of popular media in social change documents that behaviour change can also be achieved through indirect information through entertainment television (La Ferrara et al. (2012), DellaVigna and La Ferrara (2015) and La Ferrara (2016)).

A recent strand of the literature in Psychology extends this evidence on the impact of targeted information by showing that simple and small, but well targeted interventions can have large and long lasting effects on participants in other domains too. Walton (2014) provides an extensive overview of this literature. Two interventions that are particularly relevant and underpin the intervention in this paper are Hulleman and Harackiewicz (2009) on the Expectancy Value Theory and Blackwell et al. (2007) on the Growth Mindset of Intelligence. In the treatment group in Hulleman and Harackiewicz (2009), 9$^{\text{th}}$ graders, every 3 to 4 weeks, summarised the relevance of their science coursework in their lives. In the control group, students summarised the week's science topics without reference to its usefulness. The intervention raised science grades amongst students in the treatment group who expected to perform poorly.

Blackwell et al. (2007) follow up on earlier studies concerning the Growth Mindset of Intelligence. In this study, the treatment group was formed of 8^{th} grade students who learned in eight class room workshops that "intelligence is malleable and can grow like a muscle with hard work and help from others". In contrast, pupils in the control group learned about the relationship between brain regions and functions. The authors find that students in the treatment group improve their Maths grades whilst performance of those in the control group deteriorates.

My study builds on this evidence by designing a small and targeted intervention that relies on the predictions of the Social Learning Theory and Expectancy Value Theory to elicit the relationship between role models, aspirations and educational achievement. The intervention provides targeted information to pupils in order to change their study behaviour, and additionally allows to isolate the importance of relatability between the recipient and transmitter of information in driving these effects.

The next section provides a detailed description of the theories which underpin the experimental design, of the context, and of the intervention itself. Section 2.3 discusses the data and sample characteristics. In section 2.4, I present the methods and estimations, and in section 2.5 the main results. I discuss the implications of this research and its limitations in secion 2.6. Finally, section 2.7 concludes and provides guidance for further research.

2.2 Experimental Setup

2.2.1 Theory

Two theories, borrowed from social psychology, underpin the design of the experiment in this study. First, Wigfield and Eccles (2000) propose the Expectancy Value Theory and design experiments to test the validity of the hypothesis to show that achievement related choices depend on *subjective task value* and the *expectation of success*. The former, subjective task value, refers to the usefulness of mastering a given task, ie. how a task will help with one's future plans. The latter, as the name suggests, captures beliefs regarding one's probability of success in a given task. Taken together, the theory predicts that children's motivation increases if they believe that the task at hand will be useful in the future and that they have a high probability of being successful. The authors arrive at these conclusions by estimating a positive relationship between pupils' expectation of success and their performance in Maths, and between pupils' subjective task value and their propensity to continue taking Maths. They use longitudinal data on middle school pupils in the US (Meece et al. (1990)) and replicate the findings with data on primary school pupils (Wigfield et al. (1997)).

The second theory is the Social Learning Theory by Bandura (1977). Bandura hypothesises that behaviour is learned through observation of other people and tests this hypothesis in a series of experiments in the 1960s called the *Bobo Doll Experiments*. The experiments feature two treatment groups and one control group with 6 year old boys and girls as participants. In the first treatment group, participants watch an adult male or female model behaving aggressively towards a toy. In the second treatment, the model plays quietly and non-aggressively. Participants in the control group are not exposed to any model. The experiments show that children who observe aggressive behaviour are far more likely to behave aggressively than children in the second treatment or control group. The imitation effect of same gender models on boys is stronger than on girls. These results suggest that children learn through observing and imitating behaviour of models and are more likely to do so if they perceive the models as similar to themselves.

The video intervention in this study incorporates the predictions from both theories. First, it provides information on the usefulness of Maths in a range of careers and illustrates examples of people who overcame difficulties in acquiring their Maths skills, thus aiming to enhance the expectations of success. Secondly, it creates variation in the similarity between model and pupils via gender.

2.2.2 Context

The total sample in this study comprises 48 state-funded primary schools from across England. I recruited schools in part through a charity's network that also provided contact to speakers in the videos and partly through direct contact with schools and education trusts. The schools in the sample are randomly allocated to treatment and control group, stratified by prior attainment, location (London / non-London) and school size. The leading reason for randomising at the school level rather than class or individual level is to avoid contamination of the control group. In most schools, play time for all pupils of the same year is joint and randomisation at the class level would not have allowed to control for the spillover of information between pupils in the treatment and control group within the same school.

The intervention took place in all the year 6 classes of schools in the treatment group in March and early April 2017 while year 6 classes in the control schools did not receive any intervention. In treatment schools, pupils filled in endline surveys within two days of the intervention, and pupils in the control group around the same time of year. The main reason for the intervention taking place during primary school is that pupils, at this stage in their education, will not have made any

Figure 2.1: Education system in England

Age	Year Group	Curriculum Stage	National Curriculum Tests	
5/6	Year 1	Key Stage 1		Primary
6/7	Year 2		KS1 SATs	
7/8	Year 3	Key Stage 2		
8/9	Year 4			
9/10	Year 5			
10/11	Year 6		KS2 SATs	
11/12	Year 7	Key Stage 3		Secondary
12/13	Year 8			
13/14	Year 9			
14/15	Year 10	Key Stage 4	GCSEs	
15/16	Year 11			
16/17	Year 12	Key Stage 5	A Levels	
17/18	Year 13			

Source: UK Department for Education

irreversible specialisation choices that create past dependence in their achievement. Throughout primary school, all pupils follow a fixed curriculum, and it is not until secondary school that pupils make course choices for their end of secondary school exams in year 10 and 11 at age 14 - 16. See figure 2.1 for a graphical illustration of the English education system. Furthermore, year 6 as the final year of primary school is well suited for this experiment to take place because all pupils sit end of primary school tests in May of that year. Since the tests and assessments are nationally standardised, pupils' performance on them is comparable across schools. Even though these tests do not have a direct impact on pupils' educational future, eg. they do not serve as entry exams into secondary school, they are an important public indicator of schools' performance. As such, teachers and consequently pupils take them very seriously, which renders them a meaningful measure of educational achievement.

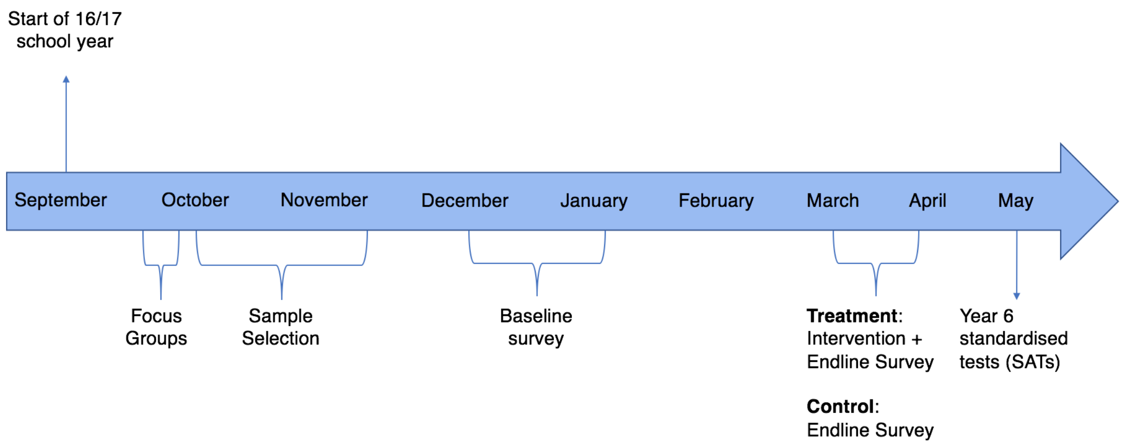
2.2.3 Intervention

The intervention consists of an approximately 10-minute-long video, made specially for this experiment. The video aims to provide information on the usefulness of

Maths and presents a TED-like speaker as potential role model. In the first part, the video illustrates the usefulness of Maths in a number of careers ranging from graphic design to robotics engineering. An animated moderator guides the pupils through an animated illustration of how graphic designers use Maths to scale comic characters using snippets from famous animation movies. The video builds on this by using the example of folding bicycles to document how bicycle engineers use geometry to calculate the correct circumference of a tyre so that it fits well with the other parts, or how robotics engineers use coordinate systems to navigate robots in robotics football. All examples are tailored to the year 6 curriculum in Maths to ensure that they cover material that is adequate for the age of pupils. This part of the video iss designed to emphasise the subjective task value of studying Maths, which, as predicted by the Expectancy Value Theory, influences motivation and achievement in Maths.

The second part of the video builds on the Social Learning Theory to generate a role model effect. This part of the video features a TED-like talk by a STEM professional on their personal "Maths story", how they use Maths in their professional and private lives, and which areas of the subject they used to struggle with and how they coped. This section has six different versions, each featuring one speaker of a different gender and ethnicity. There are three male and three female speakers of White British, South-Asian and African / Black Caribbean ethnicity. The purpose of having six different versions is to create exogenous variations in the extent to which pupils can relate to the speakers through gender and ethnicity. The analysis, however, focuses exclusively on relatability via gender, since ethnicity is unbalanced across treatment and control group in the test score data (refer to section 2.3 for more details). Having both female and male speakers, not only allows to vary the extent to which boys and girls relate to the speaker, but also to estimate the role model effect for boys and girls separately. All speakers follow the same script but give individual, genuine accounts of their experience. The occupations of speakers are balanced across gender – two bicycle engineers and one economist for each gender – to ensure that the job related examples in the videos are also balanced. The different versions of the video are randomised across treatment schools and all year 6 classes per school watch the same version of the video.

Figure 2.2: Experimental timeline



The timeline of the experiment is illustrated in figure 2.2. The baseline survey was administered in all 48 participating schools during December 2016 and January 2017. Video screenings in all but three schools took place between 14 and 31 March 2017, the three remaining interventions took place during the week commencing 17 April – this delay was due to spring term break taking place during the first two weeks of April. I collected endline survey data in all control schools during this time, and in treatment schools shortly after their respective screening dates. About two months after the intervention, between 8 and 12 May 2017, all year 6 pupils took SATs in Maths, Reading, Writing and GPS. The test scores that pupils obtained in the Maths and Reading sections of SATs will serve as my main outcome measures of achievement.

Table 2.1: All schools vs. experimental sample (for 2015 cohort)

	All Schools		Sample			
	Obs	Mean	Obs	Mean	Diff(C-T)	OLS P-val
Panel A: Test Scores						
Maths Year 6	15162	70.83	47	70.06	0.7630	0.56
Reading Year 6	15159	31.21	47	30.33	0.8853	0.12
Maths Year 2	15612	15.33	47	15.25	0.0836	0.82
Reading Year 2	15612	15.37	47	15.16	0.2112	0.59
Writing Year 2	15612	14.03	47	14.02	0.0062	0.99
Panel B: School Characteristics						
Number of pupils	15658	36.36	47	45.57	-9.2175	0.01
Girls	15658	0.48	47	0.47	0.0096	0.62
% FSM Eligible	15658	4.13	47	6.97	-2.8402	0.00
Population Density (per hectare)	15658	21.18	47	48.10	-26.9214	0.00

Note: Balancing tests based on school level data for the cohort graduating primary school in 2015, with unclustered OLS P-values reported in the last column. *% FSM Eligible* refers to an average across shares of free-school-meal eligible pupils at age 9, 10 and 11 per school. *Population Density (per hectare)* is the average number of people per hectare resident in the postcode district of the school.

2.3 Data Description & Sample Characteristics

2.3.1 Comparison of Experimental Sample to all English Primary Schools

To provide a better understanding of the sample in this study, I compare in table 2.1 the experimental sample to the universe of primary schools based on data on the cohort that graduated primary school in 2015. One sample school is not included in these statistics due to missing data, since the school recently converted to an academy and changed its school identifier code.

As balance tests, I conduct a simple T-test and report P-values of the two-sided test. The table shows that the sample is representative of all primary schools in the country in terms of achievement in year 6 and year 2 tests. P-values on the differences in achievement are far larger than conventional significance levels, except for year 6 Reading scores, which are (nearly significantly) worse in the experimental sample. However, schools in the sample are significantly larger than the average English primary school in terms of number of year 6 pupils, a larger share of pupils in the sample are eligible for free school meals, which is a measure of deprivation, and sample schools are located in more densely populated areas. These differences are not surprising as large schools in deprived, inner-city districts are more likely to be interested in raising their pupils' aspirations and therefore to respond to the invitation to join the project. Whilst these differences do not have any implications for the internal validity of the results in this paper, they are important in assessing their external validity. The findings may not translate to the average school with fewer pupils in less densely populated and less deprived locations. However, this is unlikely to be an issue if the intervention is only scaled up to schools that have identified low aspirations of pupils as a binding constraint. Since the sample was formed with an oversubscription design where all experimental schools opted into the sample, the results are likely to hold for the relevant population of schools that are interested in such an intervention.

2.3.2 Survey Data Sample

Description of Survey Data Sample

For the experimental sample, I have survey and test score data for two different sub-samples: survey data is available for 40 schools, and test score data for 30 schools. The main reason for attrition in the survey data sample is that schools failed to return the surveys after pupils had filled them in. The time of the survey coincided with the busiest period of the year with preparation for year 6 SATs. As a result, despite close monitoring, surveys from eight schools are missing. Nonetheless, attrition is random as the reduced sample is still balanced on baseline characteristics.

From the surveys, I obtain measures of aspirations in different domains. First, I measure educational aspirations as the degree a pupil aspires to. The survey question asks pupils how far they would like to go in school and possible answers range from leaving school at 16 without a degree, leaving school at 16 with some qualifications, leaving after A-levels, taking A-levels and completing vocational training to completing a university degree.

Secondly, the survey measures career aspirations by first asking pupils to list the job they would like to get in the future, and subsequently to rank the difficulty of getting that job on the Lickert scale. I interpret pupils who report a more difficult dream job to have higher aspirations, as they set themselves higher goals. Moreover, pupils are asked to list the words that come to mind when they think about an important exam. This question serves as an indicator for a Growth Mindset, which is characterised by a child's attitude to embrace mistakes as a learning opportunity and places high value on resilience. I interpret fewer negative associations, such as "scary" and "failure", as leaning more towards a Growth Mindset.

The variable *Grammar School Test* refers to whether pupils have taken an entry exam for a selective secondary school at the beginning of the school year, called grammar school tests. This variable serves as a indicator of parental aspirations. Children with parents who have higher educational aspirations are more likely to apply to a selective secondary school than other children.[4]

The survey finally measures subject specific aspirations about the performance on year 6 tests in Maths and English[5]. Pupils are familiar with being assessed in terms of age-related achievement levels, which rate their performance either below, at, or

[4]Grammar schools are state-funded so this measure of aspirations is not directly linked to parents' income. However, grammar schools are not available in all areas of England and are largely concentrated in the south-west and west of the country.

[5]English here jointly refers to tests on Reading, Writing and GPS.

above the expected level. Correspondingly, the variables *BL Maths Aspirations* and *BL Engish Aspirations* assume a value of one if a child aspires to perform below expectations, two if at expectations and three if they aspire to performing above the expected level.

The survey also elicits pupils' attitudes regarding stereotypes about the professional success of men versus women. The question asks which gender pupils think of when thinking about someone who is successful in their job. The variable *BL Job Gender Stereotype* assumes a value of zero if pupils in the baseline survey report to think of a man, and one if they think of a woman or both mean and women. Pupils also report the time they spend on homework as *Actual Homework Time* and the amount of time they think they should best be spending on their homework as *Best Homework Time*.

In addition to survey data on attitudes, I have administrative data on gender and ethnicity of children. I create categories that reflect the three main ethnicities in English primary schools: White British, South-Asian and African / Black Caribbean. All pupils of Indian, Pakistani, Sri Lankan, Bangladeshi and Nepalese ethnicities are coded as South-Asian. As African / Black Caribbean, I classify all pupils of Sub-Saharan, black Central and South American, and black Caribbean ethnicities. Finally, all children of White British descent are classified as such. A fourth category comprises all other ethnicities, including White European and any mixed ethnicities.

Summary Statistics and Balance Tests of Survey Data Sample

As balance tests, I show mean differences across treatment and control group for the two subsamples separately. In table 2.2, I show balance tests for the sample of survey data. I present OLS and randomisation inference (see section 2.4.2 for an

Table 2.2: Summary statistics for survey data sample

	Control		Treatment			P-value	
	Obs	Mean	Obs	Mean	Diff(C-T)	OLS	RI
Panel A: Aspirations							
BL Degree Aspirations	676	4.07	588	4.01	0.0644	0.60	0.59
BL Difficulty Job	655	2.91	573	2.80	0.1108	0.29	0.33
BL Growth Mindset	689	0.68	600	0.68	0.0013	0.98	0.98
Grammar School Test	688	0.49	597	0.39	0.1027	0.31	0.32
BL Maths Aspirations	664	2.19	581	2.21	-0.0187	0.61	0.73
BL English Aspirations	669	2.08	583	2.09	-0.0070	0.84	0.92
Panel B: Other Attitudes							
BL Job Gender Stereotype	656	0.59	587	0.58	0.0010	0.98	0.98
BL Act. Homework Time	676	33.95	591	31.88	2.0763	0.65	0.66
BL Best Homework Time	675	40.14	591	36.25	3.8832	0.41	0.42
Panel C: Ethnicity							
White-British Ethnicity	689	0.69	600	0.48	0.2125		
Black Ethnicity	689	0.06	600	0.14	-0.0838		
South-Asian Ethnicity	689	0.04	600	0.10	-0.0587	0.00_a***	0.23_a
Panel D: School Characteristics							
# Pupils	21	42.05	19	52.74	-10.6892	0.24	0.23
Girls	21	0.50	19	0.47	0.0264	0.45	0.45
% FSM Eligible	21	6.02	19	7.33	-1.3175	0.31	0.30
Pop. Density (per hectare)	21	40.34	19	64.45	-24.1145	0.14	0.14

Note: Balancing tests based on pupil level data. OLS P-values are clustered at the school level and reported in the penultimate column, and randomisation-t P-values from 2,000 draws are reported in the last column. *% FSM Eligible* refers to an average across shares of free-school-meal eligible pupils at age 9, 10 and 11 per school. *Population Density (per hectare)* is the average number of people per hectare resident in the postcode district of the school. The acronym *BL* in panel A and B stands for "baseline".
a: Since ethnicity shares within a cluster are not independently distributed, the F-statistics of the joint significance tests are reported.

introduction to the method) P-values from regressions of the form

$$y_i = \alpha + \beta T_i + \varepsilon_i \tag{2.1}$$

where y is the outcome and T is one if the school is in the treatment group and zero otherwise.

The outcomes in panel A - C are based on pupil level data and OLS standard errors are clustered at the school level. In panel C, I report F-values for the joint significance of all three ethnicity shares in the last two columns as these shares of pupils in a school are inter-dependent.[6] The data in panel D are at the school level and since this is the unit of treatment, standard errors are not clustered.

The table shows that the sample contains survey data for 21 control schools and 19 treatment schools. Pupils in treatment and control schools have very similar degree aspirations with the average child aspiring to doing A-levels and completing vocational training. Career aspirations measured as the difficulty of their dream job are slightly higher in the control group by about 3%. The Growth Mindset measure is almost identical in both groups and highlights that 68% of children do not express negative associations with important exams. Pupils in the control group are ten percentage points more likely to have taken a grammar school test than treatment pupils. Aspirations about performance on year 6 Maths and English tests are similar across treatment and control group. On average, pupils aspire to performing slightly above the age related expected level, and aspirations are marginally higher for Maths than for English.

Gender stereotypes about professional success are balanced across treatment and control group. The means show that a little under 60% of children think of women or both men and women when they think of someone with a successful career. Pupils in both groups report similar time spent on homework, 34 minutes in the control group and 32 minutes in the treatment group. When asked how much time

[6]For instance, a large share of South-Asian pupils implies a relatively lower share of White-British and Black pupils by construction. I present the F-value from a regression of the form $T_i = \alpha + \beta_1 White\text{-}British_i + \beta_2 Black_i + \beta_3 South\text{-}Asian_i + \varepsilon_i$

pupils think they should spend on their homework, both groups report more time than they actually spend. Pupils in control groups consider 40 minutes the best homework time, in the treatment group it is 36 minutes on average.

The ethnic composition varies between treatment and control group. While 70% of pupils in the control group are White-British, less than 50% of the treatment group are. This implies, that the shares of pupils of Black and South-Asian ethnicities vary too. While 14% and 10% of pupils in the treatment group are of Black and South-Asian ethnicity, respectively, only 6% and 4% of the control group are. To account for the interdependence of these differences, I test the joint significance of all three indicators in predicting treatment and report the corresponding F-values in the two last columns.

In panel D, I present descriptive statistics of some school level characteristics. The average year 6 cohort in treatment schools comprises 52 pupils, ten more than in control schools. Half of all pupils in control schools are girls, the share in treatment schools is just below at 47%. The share of pupils who are eligible for free-school-meals is one percentage point higher at 7% in treatment schools compared to the control group. Furthermore, the average treatment school is located in a postcode district with slightly higher population density (64.45 per hectare) than the control group (40.34 per hectare).

The penultimate column of the table shows that based on OLS P-values the sample is balanced on variables at baseline, except ethnicity. With randomisation inference (RI) P-values, the differences in means for all baseline variables are indistinguishable from zero at all conventional levels of significance. Notwithstanding, due to the small sample size standard errors are large and some of the differences in means are substantial despite not being significantly different from zero. I deal with the imbalance in ethnicity by controlling for the three ethnicity indicators in all regressions.

Since none of the other imbalances are significant, controlling for them would not change the estimates. However, to the extent that these imbalances suggest other, unobservable differences between the two groups, they remain an issue that I cannot correct for in my estimations.

2.3.3 Test Score Data Sample

Description of Test Score Data Sample

I have administrative school records on test scores for a subsample of 30 schools, 15 in the treatment and 15 in the control group. As in the survey data sample, the primary reason for attrition is staff turnover. After year 6 SATs were taken in May and schools received the published results at the beginning of July 2017, schools shared these with me. Since schools did not receive the published results until shortly before summer break, most schools only shared them with me in the beginning of the new school year. Eighteen schools from the original sample of 48, or twelve from the sample for which survey data is available, opted out of the study during this time since the main contact had left the school over the break.[7]

For the analysis of the effect on test scores, I measure performance on year 6 SATs in Maths and Reading as scaled point scores that range from 80 to 120, where 100 refers to achievement at the expected age-related standard. To measure prior attainment, I use pupils' test scores from mandatory tests taken in Reading, Writing and Maths in year 2 at the age of 6 - 7. These scores range from 1 to 9, where 1 refers to achievement below the expected level in year 1 and 9 refers to above expected level in year 3. In the regression analysis in section 2.5.1, I standardise test scores to mean zero and standard deviation of one in the control group by gender, thereby also accounting for any gender differences in the dispersion of test scores. In this study, test scores in Maths and Reading serve as the main outcome

[7]Two schools that failed to report survey data did share test score data.

measure of educational achievement.

Summary Statistics and Balance Tests of Test Score Data Sample

In table 2.3, I show summary statistics and balance tests for the test score data sample. To test for balance, I estimate the regression in equation 2.1, and for ethnicity, I test the joint significance of all three indicators as discussed above (see footnote 6).

Prior attainment, measured as test scores in year 2 exams[8] on Maths, Reading and Writing, is very balanced across the two groups. A mean Maths score of 5.41 in the control group implies that pupils performed at the expected level in year 2. Pupils in control schools performed, on average, slightly better in Maths, and worse in Reading and Writing. Measures of aspirations and other attitudes are also similar across groups. The F-values on the difference in ethnic compositions across the two groups is significant with clustered OLS standard errors and with RI. As I discuss above, I account for this imbalance, by controlling for ethnicity in all regressions.

2.3.4 Differences by Gender at Baseline

In addition to balance tests, I present differences in attitudes and test scores by gender at baseline. I estimate these differences with a regression of the form

$$y_i = \alpha + \beta Boy_i + \varepsilon_i$$

where y_i are attitudes at baseline or performance on year 2 tests and Boy_i is an indicator equal to one if a pupil is a boy. Standard errors are clustered at the school level. The estimates in table 2.4 show differences by gender in standardised test scores in year 2 tests in Maths, Reading and Writing, and survey measures of aspirations and other attitudes at baseline.

[8]For ease of interpretation, I present here non-standardised test scores

Table 2.3: Summary statistics for test score data sample

	Control		Treatment			P-value	
	Obs	Mean	Obs	Mean	Diff(C-T)	OLS	RI
Panel A: Test scores							
Maths Year 2	586	5.41	575	5.37	0.0373	0.87	0.89
Reading Year 2	581	5.46	535	5.49	-0.030	0.90	0.91
Writing Year 2	577	4.89	535	4.98	-0.090	0.71	0.75
Panel B: Aspirations							
BL Degree Aspirations	329	4.14	423	4.17	-0.0382	0.80	0.79
BL Difficulty Job	323	2.96	412	2.88	0.0804	0.54	0.60
BL Growth Mindset	586	0.83	579	0.77	0.0556	0.36	0.39
Grammar School Test	374	0.42	370	0.39	0.0360	0.78	0.79
BL Maths Aspirations	325	2.25	423	2.24	0.0074	0.87	0.89
BL English Aspirations	328	2.15	425	2.16	-0.0137	0.75	0.84
Panel C: Other Attitudes							
BL Job Gender Stereotype	323	0.62	424	0.59	0.0265	0.58	0.59
BL Act. Homework Time	331	39.45	429	36.25	3.2034	0.69	0.72
BL Best Homework Time	332	44.24	427	42.85	1.3947	0.86	0.88
Panel D: Ethnicity							
White-British Ethnicity	586	0.70	579	0.42	0.2782		
Black Ethnicity	586	0.08	579	0.17	-0.0873		
South-Asian Ethnicity	586	0.15	579	0.20	-0.0501	0.00_a***	0.10_a*
Panel E: School Characteristics							
# Pupils	15	47.13	15	49.13	-2.0000	0.85	0.84
Girls	15	0.53	15	0.47	0.0612	0.21	0.20
% FSM Eligible	14	5.88	15	6.47	-0.5857	0.69	0.70
Pop. Density (per hectare)	15	41.86	15	66.75	-24.8933	0.22	0.21

Note: Balancing tests based on pupil level data. OLS P-values of the differences in means are clustered at the school level and reported in the penultimate column, and randomisation-t P-values from 2,000 draws are reported in the last column. *% FSM Eligible* refers to an average across shares of free-school-meal eligible pupils at age 9, 10 and 11 per school. *Population Density (per hectare)* is the average number of people per hectare resident in the postcode district of the school. The acronym *BL* in panel B and C stands for "baseline".
a: Since ethnicity shares within a cluster are not independently distributed, the F-statistics of the joint significance tests are reported.

Panel A shows that girls perform significantly better than boys on year 2 tests in Reading and Writing by about 0.6 to 0.7 points on a scale from 0 to 9, and insignificantly so in Maths. Panel B shows that girls, on average, aspire to education beyond A levels, that means either university or vocational training, whilst boys prefer not to continue after A-levels. In contrast, boys rate the difficulty of their dream job significantly higher by 4% (point estimate of the difference is 0.11) compared to girls, and they are 7% more likely to embrace a Growth Mindset, while the probability of having taken a grammar school test is not significantly different between boys and girls. On average, girls aspire to a Maths performance of 2.1 out of three, where two is the expected level and three is above expectations, and an English performance of 2.15, whereas boys' aspirations for Maths are significantly higher by 10.5% and for English lower by 5%. This is interesting and stands in contrast with the gender differences in subject specific aspirations, which indicate that boys have higher Maths aspirations than girls despite lower or equal prior attainment.

Panel C presents differences in terms of other attitudes such as gender stereotypes about career success, where roughly three quarters of girls think that women are at least as or more successful than men, whereas boys majoritively consider men more successful. The differences by gender in actual and best homework time are not significant.

2.4 Estimation & Methods

2.4.1 Estimation of Main Effect

In what follows, I estimate the effects of two treatments: (i) the effect from being in the treatment group, ie. watching the video with some speaker, regardless of whether the gender of speaker and pupil match, and (ii) the effect on children who watch a video with a speaker who is of the same gender as themselves. In the regressions, I standardise all multinomial survey measures and test scores to mean

Table 2.4: Gender differences in prior achievement and baseline survey measures

	Girls		Boys			
	Obs	Mean	Obs	Mean	Diff(G-B)	OLS P-val
Panel A: Test Scores						
Year 2 Maths	552	5.43	596	5.37	0.0549	0.46
Year 2 Reading	531	5.77	573	5.21	0.5681	0.00***
Year 2 Writing	530	5.30	569	4.61	0.6918	0.00***
Panel B: Aspirations						
BL Degree Aspirations	753	4.24	743	3.92	0.3211	0.00***
BL Diffic. Job	724	2.80	730	2.92	-0.1126	0.09*
BL Growth Mindset	1068	0.74	1097	0.82	-0.0715	0.00***
Grammar School Test	811	0.45	842	0.47	-0.0252	0.52
BL Maths Aspirations	735	2.10	739	2.31	-0.2160	0.00***
BL English Aspirations	746	2.15	738	2.05	0.0987	0.02**
Panel C: Other Attitudes						
BL Job Gender Stereotype	734	0.75	735	0.42	0.3344	0.00***
BL Act. Homework Time	753	37.08	746	33.76	3.3196	0.13
BL Best Homework Time	751	41.38	746	40.55	0.8246	0.71

Note: OLS P-values of the differences in means are clustered at the school level and reported in the last column. The acronym *BL* in panel B and C stands for "baseline".

zero and a standard deviation of one in the control group by gender to make the coefficients more meaningful to interpret.

Simple ATE

To estimate the simple ATE, I regress the outcome variable on a treatment indicator and optionally add individual level controls.

$$y_i = \beta_0 + \beta_1 T_i + \phi \tilde{y}_i + \mu X_i + \varepsilon_i \tag{2.2}$$

The treatment indicator T_i in equation 2.2 is equal to one if pupil i is in a treatment school so that β_1 estimates the effect from being treated. To interpret the estimated coefficients as measures of progress, I control for baseline values of the outcome variable, or achievement in year 2 tests where the outcome variables are test scores, as $\tilde{y}_i$. The vector of individual level controls X_i includes ethnicity indicators in all regressions.

Same Gender ATE

The estimation of the Same Gender treatment effect follows analogously. The main coefficient of interest in equation 2.3 is β_2, and $Same\ Gender_i$ is a dummy variable indicating that speaker and pupil are of the same gender.

$$y_i = \beta_0 + \beta_1 T_i + \beta_2 Same\ Gender_i + \phi \tilde{y}_i + \mu X_i + \varepsilon_i \tag{2.3}$$

I interpret the coefficient on the *Same Gender* dummy as the effect from watching a video with a speaker who is of the same gender. The counterfactual are pupils of that same gender who watched a video with a speaker of the opposite gender. The coefficient on treatment assignment, β_1, is an estimate of the effect from watching an opposite gender speaker compared to being in the control group.

To disentangle the effects on girls watching a female speaker and boys watching a male speaker, I also estimate equation 2.4. In this specification, I allow for the same gender treatment effect to vary between girls and boys.

$$y_i = \beta_0 + \beta_1 T_i + \beta_2 Girl * Female_i + \beta_3 Boy * Male_i + \phi \tilde{y}_i + \mu X_i + \varepsilon_i \tag{2.4}$$

where the interaction $Girl * Female_i$ turns one for girls with a female speaker and $Boy * Male_i$ for boys with a male speaker.

Lastly, I estimate the fully satiated model of the form

$$\begin{aligned} y_i &= \beta_0 + \beta_1 Girl * Female_i + \beta_2 Boy * Male_i + \\ &+ \beta_3 Girl * Male_i + \beta_4 Boy * Female_i + \phi \tilde{y}_i + \mu X_i + \varepsilon_i \end{aligned} \quad (2.5)$$

2.4.2 Randomisation Inference

In addition to conventional clustered standard errors, I also estimate all equations using RI to calculate the standard errors. The small sample in this analysis, and in particular the small number of clusters, can lead to downward bias in conventional clustered standard errors. Young (2017) shows that this issue is particularly salient in estimations where the dependent variable is continuous and the treatment variable is an interaction with covariates. In this study, test score estimations in equation 2.3 - 2.5 satisfy both criteria. Following Rosenbaum (2002), RI exploits the merits of randomisation to construct a test for which the distribution is known and that is resilient to outliers. This allows estimation of standard errors that are exact and thus unbiased even in a small sample. In RI, treatment allocation T_i is considered the only stochastic element in the estimation and so, the outcome variable is a deterministic function of treatment allocation. This allows to calculate a vector $\boldsymbol{\beta}$ of coefficients for any potential random allocation of T_i. The P-values used in this study reflect the null hypothesis that the $\boldsymbol{\beta}$ vector obtained in the experiment is extreme enough compared to all other realisations of $\boldsymbol{\beta}$ based on random allocations of schools to treatment.

Intuitively, it helps to think of RI as creating $n = \{1, \ldots, N\}$ different placebo treatment vectors and regressing the outcome on each of these. In each draw n, a placebo treatment vector is created that randomly assigns schools to placebo treatment and control groups to subsequently estimate the coefficients based on this placebo treatment assignment. This procedure is repeated for each draw and finally yields a set of N estimated coefficients. RI uses the distribution of the N estimated placebo treatment effects to compare the estimate obtained in the experiment. For instance, based on the distribution of these coefficients, one rejects the null hypothesis of no effect at 5% significance if the estimated coefficient obtained in the experiment lies inside the 2.5^{th} percentile of either tail of the distribution of placebo coefficients.

2.5 Results

2.5.1 Main Effects

In table 2.5 to 2.8, I show estimates of the models in equation 2.2 to 2.5 for all outcomes, and present clustered OLS P-values in round brackets and RI P-values in squared brackets below the coefficients.

Treatment Effects on Test scores

The simple average treatment effect from watching the video, regardless of the gender match between pupils and speakers, on both progress between year 2 and year 6 tests in Maths and Reading is close to zero, negative and insignificant.

When I compare the treatment effects of pupils with matching gender speakers and non-matching speakers, as shown in the second column for each outcome variable, the differences in treatment effects on performance are insignificant, but the estimate on Maths progress for pupils with a same gender speaker becomes positive whilst that on Reading scores is negative regardless of the gender match. The third column splits the same gender effect by boys and girls and shows that boys with male speakers perform significantly better in Maths by 0.19 standard deviations than those with female speakers, while for girls, the gender match does not lead to a change in performance. For the Reading outcome, the effects point in the same direction but are insignificant. Estimation of the fully satiated model in the fourth

Table 2.5: Treatment effect on test scores

	Maths Year 6				Reading Year 6			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.06	-0.09	-0.09		-0.06	-0.05	-0.05	
	(0.60)	(0.41)	(0.44)		(0.60)	(0.66)	(0.68)	
	[0.65]	[0.45]	[0.49]		[0.63]	[0.70]	[0.72]	
Same Gender Speaker		0.07				-0.04		
		(0.28)				(0.62)		
		[0.29]				[0.64]		
Girl*Female Speaker			-0.05	-0.16			-0.11	-0.17
			(0.70)	(0.36)			(0.27)	(0.25)
			[0.73]	[0.43]			[0.33]	[0.32]
Boy*Male Speaker			0.19	0.12			0.03	-0.01
			(0.08)*	(0.31)			(0.83)	(0.95)
			[0.14]	[0.40]			[0.85]	[0.96]
Girl*Male Speaker				-0.14				-0.07
				(0.29)				(0.56)
				[0.39]				[0.62]
Boy*Female Speaker				-0.04				-0.03
				(0.78)				(0.85)
				[0.82]				[0.85]
BL / Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Boy Dummy		Yes	Yes	Yes		Yes	Yes	Yes
Observations	1140	1140	1140	1140	1093	1093	1093	1093
# clusters	30	30	30	30	30	30	30	30

Note: OLS P-values are clustered at the school level and reported in round brackets below the coefficients, and randomisation-t P-values from 2,000 draws are reported in squared brackets. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring performance on year 2 tests, which are included in the regression to express treatment effects in terms of progress between year 2 and year 6 tests. Test scores are standardised to a standard deviation of one and mean zero in the control group.

column illustrates that the treatment effects by gender match on progress in comparison to the control group are qualitatively similar to those within the treatment group by gender match, but insignificant.

Table 2.6: Treatment effect on aspirations

	Degree Asp.				Diffic. Job				Growth Mindset			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment	-0.14	-0.34	-0.35		0.06	-0.21	-0.21		-0.00	-0.04	-0.04	
	(0.62)	(0.30)	(0.29)		(0.85)	(0.56)	(0.54)		(0.90)	(0.28)	(0.26)	
	[0.58]	[0.30]	[0.30]		[0.86]	[0.59]	[0.57]		[0.91]	[0.30]	[0.29]	
Same Gender Speaker		0.37				0.56				0.06		
		(0.17)				(0.04)**				(0.02)**		
		[0.20]				[0.06]*				[0.03]**		
Girl*Female Speaker			-0.01	-0.30			-0.17	-0.33			0.10	0.05
			(0.99)	(0.57)			(0.68)	(0.43)			(0.02)**	(0.34)
			[0.99]	[0.60]			[0.69]	[0.48]			[0.03]**	[0.37]
Boy*Male Speaker			0.68	0.28			1.14	0.88			0.02	-0.02
			(0.03)**	(0.34)			(0.00)***	(0.01)***			(0.63)	(0.68)
			[0.03]**	[0.37]			[0.01]***	[0.02]***			[0.65]	[0.71]
Girl*Male Speaker				-0.20				-0.08				-0.08
				(0.62)				(0.84)				(0.15)
				[0.63]				[0.84]				[0.19]
Boy*Female Speaker				-0.52				-0.37				-0.01
				(0.28)				(0.53)				(0.80)
				[0.32]				[0.56]				[0.82]
BL / Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Boy Dummy		Yes	Yes	Yes		Yes	Yes	Yes		Yes	Yes	Yes
Observations	1256	1256	1256	1256	1199	1199	1199	1199	2284	2165	2165	2165
# clusters	30	30	30	30	30	30	30	30	30	30	30	30

Note: OLS P-values are clustered at the school level and reported in round brackets below the coefficients, and randomisation-t P-values from 2,000 draws are reported in squared brackets. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring the outcome variable in the baseline survey, which are included in the regression to express treatment effects in terms of change between baseline and endline survey. Test scores are standardised to a standard deviation of one and mean zero in the control group.

Treatment Effects on Aspirations

Similar to the estimates of the effect on test performance, the simple treatment effects on measures of aspirations in table 2.6 and 2.7 are small, negative and insignificant. In particular, the effect from watching a video with a speaker of any gender does not affect average degree aspirations, the reported difficulty of the pupil's dream job, the Growth Mindset or subject specific aspirations for Maths of English.

The comparison of treatment effects by gender match within the treatment group highlights that a same gender speaker makes pupils aspire to a significantly more difficult dream jobs by 0.56 standard deviations, the probability of embracing a Growth Mindset increases by 6%, and aspirations for performance on year 6 Maths tests by 0.23 standard deviations. The difference in the effects on Maths and English aspirations, and also on test scores as discussed above, suggests that the Maths specific information in the video play a role in driving these effects.

Disentangling the effect by pupil gender shows that for boys with male speakers, degree aspirations increase by 0.68 standard deviations and the difficulty of their dream job by 1.14 compared to those with a female speaker, and girls with a female speaker are 2% more likely to embrace a Growth Mindset than those with a male speaker, whilst the same gender effects on subject specific aspirations seem to be driven by both boys and girls. Noteworthy here is also the negative, and in the case of English, borderline significant effect of male speakers on girls' test aspirations.

The positive effect on the difficulty of boys' dream jobs with male speakers remains significant when comparing them to boys in the control group (0.88 standard deviations), whereas for other outcomes the comparisons to the control group are insignificant. In fact, the satiated model reveals that some of the positive coeffi-

Table 2.7: Treatment effect on subject specific aspirations

	Maths Aspirations				English Aspirations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.07	-0.17	-0.17		-0.06	-0.11	-0.11	
	(0.71)	(0.32)	(0.32)		(0.68)	(0.53)	(0.53)	
	[0.73]	[0.37]	[0.37]		[0.70]	[0.56]	[0.56]	
Same Gender Speaker		0.23				0.12		
		(0.02)**				(0.60)		
		[0.03]**				[0.60]		
Girl*Female Speaker			0.35	0.14			0.02	-0.16
			(0.14)	(0.52)			(0.92)	(0.43)
			[0.17]	[0.56]			[0.92]	[0.46]
Boy*Male Speaker			0.14	0.01			0.20	0.16
			(0.41)	(0.98)			(0.54)	(0.57)
			[0.50]	[0.98]			[0.53]	[0.60]
Girl*Male Speaker				-0.29				-0.31
				(0.17)				(0.11)
				[0.21]				[0.14]
Boy*Female Speaker				-0.04				0.15
				(0.89)				(0.61)
				[0.90]				[0.65]
BL / Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Boy Dummy		Yes	Yes	Yes		Yes	Yes	Yes
Observations	1216	1216	1216	1216	1225	1225	1225	1225
# clusters	30	30	30	30	30	30	30	30

Note: OLS P-values are clustered at the school level and reported in round brackets below the coefficients, and randomisation-t P-values from 2,000 draws are reported in squared brackets. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring the outcome variable in the baseline survey, which are included in the regression to express treatment effects in terms of change between baseline and endline survey. Test scores are standardised to a standard deviation of one and mean zero in the control group.

cients on the same gender treatment variables in the second and third column of each variable appear to be a result of the combination of a positive same gender effect and a negative effect of the opposite gender treatment. This seems to be the case for the positive effect on Maths test scores and degree aspirations for boys and on Growth Mindset for girls, where the opposite gender treatment has a negative sign for each. Nonetheless, even compared to the control group, boys with male speakers have higher Maths test scores and degree aspirations and girls with with female speakers are more likely to have a Growth Mindset, both however insignificantly so.

Overall, the estimations document that watching a video with a speaker of the same gender leads to positive changes in Maths performance and all measures of aspirations, except aspirations for performance on English tests. This implies that pupils process the information in the video a positively or more effectively when they can relate to the speaker via gender. This suggests that the effects may be driven by a role model channel. Notwithstanding, the comparison of positive and significant effects on Maths versus insignificant effects on Reading outcomes emphasises that not only the transmitter of information but also the informational content itself plays a role in improving aspirations and performance.

For Maths test performance, degree aspirations, and the difficulty of the dream job, the positive same gender effect is stronger for boys than for girls, whilst the effect on Maths aspirations appears to be driven by both girls and boys. This suggests that the role model intervention is more effective in changing the aspirations and performance of boys than girls.

Treatment Effects on other Attitudes

The treatment does not have a significant effect on gender stereotypes about professional success. Whilst the signs of the coefficients in column 3 and 4 seem to suggest that girls with female speakers favour women and boys with male speakers favour men, the coefficients are not significant at conventional levels.

Table 2.8: Treatment effect on other attitudes

	Gender Sterotype				Actual Homework Time				Best Homework Time			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment	-0.02	-0.03	-0.02		-4.69	-5.33	-5.33		-8.77	-11.27	-11.23	
	(0.38)	(0.37)	(0.38)		(0.39)	(0.36)	(0.36)		(0.07)*	(0.01)**	(0.01)**	
	[0.40]	[0.40]	[0.41]		[0.41]	[0.39]	[0.38]		[0.09]*	[0.03]**	[0.02]**	
Same Gender Speaker		0.01				1.39				5.26		
		(0.81)				(0.75)				(0.24)		
		[0.82]				[0.77]				[0.28]		
Girl*Female Speaker			0.04	0.03			2.26	-4.50			11.89	-0.88
			(0.16)	(0.31)			(0.82)	(0.64)			(0.13)	(0.92)
			[0.18]	[0.34]			[0.84]	[0.68]			[0.19]	[0.93]
Boy*Male Speaker			-0.03	-0.06			0.70	-3.24			-0.14	-9.89
			(0.54)	(0.19)			(0.89)	(0.62)			(0.98)	(0.13)
			[0.59]	[0.23]			[0.90]	[0.65]			[0.97]	[0.17]
Girl*Male Speaker				-0.00				-9.54				-15.77
				(0.93)				(0.29)				(0.00)***
				[0.93]				[0.35]				[0.02]**
Boy*Female Speaker				-0.05				-0.34				-5.92
				(0.24)				(0.94)				(0.21)
				[0.27]				[0.95]				[0.28]
BL / Prior Attainment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Boy Dummy		Yes	Yes	Yes		Yes	Yes	Yes		Yes	Yes	Yes
Observations	1219	1219	1219	1219	1250	1250	1250	1250	1251	1251	1251	1251
# clusters	30	30	30	30	30	30	30	30	30	30	30	30

Note: OLS P-values are clustered at the school level and reported in round brackets below the coefficients, and randomisation-t P-values from 2,000 draws are reported in squared brackets. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring the outcome variable in the baseline survey, which are included in the regression to express treatment effects in terms of change between baseline and endline survey. Test scores are standardised to a standard deviation of one and mean zero in the control group.

The estimated treatment effects on the time pupils report to spend on homework (*Actual Homework Time*), are all insignificant. Interestingly, the time a pupil thinks they should spend on homework (*Best Homework Time*) decreases by 8.77 minutes on average conditional on baseline measures, and the findings show that this effect is driven primarily by girls with male speakers who report 15.77 minutes less than girls in the control group. On average, pupils with an opposite gender speaker think they should spend 11.25 minutes less on their homework than the control group. The comparison with the insignificant effects on actual homework time suggest that this change in attitudes does not translate into actual behaviour.

2.5.2 Treatment Effect Heterogeneity by Peer Performance

To examine how the same gender treatment effect varies with peer performance, I interact the same gender coefficient with the performance of the top performing student by gender per class. In table 2.9, I show estimates for the heterogeneity by peer performance for girls and boys following regressions of the form

$$\begin{aligned} y_i \;=\;& \beta_0 + \beta_1 Girl * Female_i + \beta_2 Boy * Male_i + \\ & +\beta_3 Girl * Male_i + \beta_4 Boy * Female_i + \\ & +\beta_5 Girl * Female * High\ Performing_i + \\ & +\beta_6 Boy * Male * High\ Performing_i + \\ & +\beta_7 High\ Performing_i + \mu X_i + \varepsilon_i \end{aligned} \tag{2.6}$$

where $High\ Performing_i$ refers to the test score of the highest performing girl in pupil i's class if i is female and the highest performing boy if i is male. The estimate for β_5 indicates the difference in the same gender treatment effect on girls whose highest performing female peer has a test score that is higher than the average test score of top performing girls across all classes in the sample, conditional on prior attainment. The estimate for β_1 is the same gender treatment effect of girls whose

Table 2.9: Interaction of treatment effect on test scores with peer performance

	Maths Year 6	Reading Year 6
	(1)	(2)
Girl*Female Speaker	0.38	0.21
	(0.03)**	(0.26)
	[0.12]	[0.42]
Boy*Male Speaker	0.20	0.01
	(0.25)	(0.95)
	[0.37]	[0.96]
Girl*Male Speaker	-0.15	-0.06
	(0.26)	(0.64)
	[0.33]	[0.69]
Boy*Female Speaker	-0.07	-0.04
	(0.63)	(0.77)
	[0.70]	[0.82]
Girl*Female Speaker*High Perf.	-0.32	-0.25
	(0.00)***	(0.00)***
	[0.03]**	[0.03]**
Boy*Male Speaker*High Perf.	-0.12	-0.02
	(0.33)	(0.81)
	[0.12]	[0.47]
BL / Prior Attainment	Yes	Yes
Ethnicity FE	Yes	Yes
Boy Dummy	Yes	Yes
High performing	Yes	Yes
Observations	1012	967
# clusters	30	30

Note: OLS P-values are clustered at the school level and reported in round brackets below the coefficients, and randomisation-t P-values from 2,000 draws are reported in squared brackets. The indicator *BL / Prior Attainment*, at the bottom of the table, refers to covariates measuring performance on year 2 tests, which are included in the regression to express treatment effects in terms of progress between year 2 and year 6 tests. The variable *High performing* captures test scores on year 6 tests of the highest performing girl (boy) in the class of girl (boy) i. Test scores are standardised to a standard deviation of one and mean zero in the control group.

highest performing female peer is in the bottom half of all top performing girls in the sample. Analogously, β_6 is the change in test scores due to the same gender treatment on boys with top performing male peers compared to boys with average performing male peers.

Interestingly, the negative estimates of β_5 indicate that girls with top high performing peers gain less from the same gender effect than girls with average high performing peers, and the positive estimate for β_1 that girls with low high performing female peers gain more than those with average high performing peers. This implies that girls' response to the same gender treatment varies significantly depending on the quality of their peers. Girls with very high achieving peers benefit less from female speakers than girls with less high achieving classmates, which suggests a substitutability between very high achieving classmates and the same gender treatment. For girls who are used to seeing other girls in their class achieve very high test scores, the successful female speaker in the video might represent less of a novelty than for girls whose classmates are not amongst the top performers.

For boys, there is no evidence for such heterogeneity in treatment effects by peer performance.

2.6 Discussion

2.6.1 Alternative Interventions

The findings in this paper are, to some extent, certainly specific to the type of intervention and they may vary for other interventions involving role models to raise aspirations. Therefore, hereafter I briefly discuss alternative interventions and how the treatment in this paper fits in.

The ideal setting to study the impact of role models would be one that allows to compare aspirations and educational achievement of pupils with endogenously chosen role models with whom they maintain a close and long-standing relationship, like family and friends, to pupils without such role models. However, this is naturally impossible to reconstruct in an experimental setting as these relationships form endogenously, which makes it impossible to isolate the impact of the role model from other correlated influences.

One approach to translating this setting into a controlled experiment, would be to introduce all pupils in the treatment group to a number of potential mentors and have each pupil choose one to whom they can relate, and compare outcomes of the treatment pupils to a control group without mentors. To identify the effect of relatability between pupils and mentors, half of the treatment group could be allocated a random mentor instead of the one they chose to be most relatable. Furthermore, each mentor would have to have mentees in each treatment group to ensure that there are no systematic imbalances in mentor quality between the two groups. Mentors would work with individual mentees during multiple sessions to provide information on the usefulness of Maths and raise their aspirations by showcasing that someone like themselves is capable of Maths and a career in a related field.

The intervention in this paper is an approximation of this hypothetical ideal setting within the boundaries of feasible options. Rather than allowing each pupil to choose a relatable speaker, which would have been too time consuming for both pupils and speakers in this project, I approximate variation in relatability between speakers and pupils via the gender match and assume that on average pupils with a speaker of matching gender are more likely to consider them a role model than those with an opposite gender speaker. Moreover, to ensure that the average quality of speakers is balanced between those who have a matching gender speaker and those who do not, the intervention takes place via a uniform video. There are also three speakers of each gender to reduce the likelihood that one gender is systematically of higher quality than the other.

Many organisations are running mentorship programmes and other studies have used these settings to study the impact of one-off interventions on educational outcomes (eg. Breda et al. (2018) in France). In future work, researchers could collaborate with organisations to design a longer lasting intervention according to the above criteria to study the impact of intensive mentorship from role models compared to more short lived interventions.

2.6.2 Compatibility of Results with Previous Chapter

In this section I briefly discuss the compatibility of the results in this paper with the previous chapter. In both chapters, I find an insignificant average effect of treatment on pupils. Since the sample size in the first chapter is larger than in the second, this result suggests that the insignificant effect in this chapter does not arise from to imprecise inference due to the small sample size, but rather that positive and negative effects on sub-groups of students cancel each other out on average. Albeit, this conclusion rests on the assumption that the insignificant effect in the first chapter is not driven by bias due to selection into treatment, which I cannot perfectly rule out.

The positive same gender effects on boys that I find in this chapter stand in contrast with the negative effects in the previous one, where these also vary significantly with regional views on women's role in the labour market. The heterogeneity by regional gender norms could be driven by pupils who perceive the career talks differently depending on the local norms, or by variation in the quality of speakers that is correlated with views on women's role in the labour market. In this paper, the random allocation of video interventions to schools allows to rule out such a correlation between local norms and speaker quality, and the results show no evidence for a negative effect on boys. In fact, table A.1 in Appendix A shows that treatment effects do not vary with local gender norms, or if anything the variation points in the opposite direction as in the previous chapter. This may imply that the negative effects on boys with male speakers in the previous chapter are driven by poor quality speakers in regions with traditional gender norms.

The negative effects on girls in the previous chapter are largely consistent with the findings in this paper, even though most of the effects are insignificant here. In comparison with the control group, girls with male speakers lose out in all outcome measures, and significantly so in the time they consider optimal to spend on homework (*Best Homework Time*). The findings in the previous chapter may suggest that the insignificance of most effects in this paper is due to the small sample size that makes precise inference more challenging.

2.7 Conclusion

The aim of this study is to analyse the relationship between role models, aspirations and educational achievement of pupils. The achievement gap between advantaged and disadvantaged pupils is substantial and persistent over time. This paper discusses the influence of role models and aspirations in explaining this gap. I argue that since pupils' exposure to role models is positively correlated with their social neighbourhood, and if role models affect aspirations and educational achievement, a lack of role models harms social mobility.

As economists we are interested in the efficient use of resources to optimise output. We also care about efficient allocation for reasons of equality: equality of opportunity requires for all members of society to have the same chances of success independent of their socioeconomic background. While efficiency and equity pursuits usually pose a trade-off, educational institutions that are designed so that all pupils are best able to achieve their full potential, improve upon both efficiency and equity.

This implies compensating for both constraints to the supply of education, such as school quality, and the demand for education. This paper shows that demand for education is influenced by pupils' aspirations. Aspirations are determined by pupils' social surroundings, which themselves also directly affect demand for education, eg. parents' education and neighbourhood characteristics. Therefore, understanding the influence of role models on aspirations has important implications for education policy as it can create a virtuous circle that leads to more role models for future generations. For instance, reducing the achievement gap by socioeconomic background in education and the labour market, implies a more equal distribution of role models, which may result in a self-sustained reduction of educational inequality and a more efficient use of resources.

I investigate the impact of role models on aspirations and educational achievement of primary school pupils with a randomised controlled trial. I find evidence for a positive impact of role models on aspirations and educational achievement. Whilst the video intervention in this RCT has no significant effect on pupils in the treatment group on average, it has a positive effect on pupils who watch a speaker of their own gender. Boys, in response to watching a video of a male speaker, increase Maths achievement by 0.19 standard deviations, while the effect on girls is insignificant. Boys with male speakers have higher degree aspirations by 0.7 standard deviations, and the difficulty of their dream job increases by 1.14 standard deviations, and girls with female speakers are 10% more likely to have a Growth Mindset than those with opposite gender speakers. Pupils with same gender speakers on average have higher Maths aspirations by 0.23 standard deviations. Estimates of the treatment effects on other attitudes are mostly insignificant, but girls with male speakers reduce the homework time they consider optimal by 16 minutes.

These results emphasise the importance of the gender match for the intervention to be effective, which suggests that pupils are more likely to internalise information from people who they perceive as similar to themselves. This is in line with predictions of the Social Learning Theory by Bandura (1977), which shows that behaviour is learned through observation and imitation of relatable models.

I explore how the effects on test scores vary with performance of same gender classmates and find that girls with very high achieving female peers benefit less from the same gender treatment than girls with less high achieving female classmates. The same interaction effect is not significant for boys and their same gender peers. This suggests that girls' performance is more strongly affected by same gender peer performance than boys', and that high achieving female classmates might act as substitutes for adult female role models.

The findings in this study open up avenues for further enquiry. Since inference in this analysis suffers from low power, follow-up research should rerun the experiment with current primary school pupils and pool the data to increase the sample size. This would allow to estimate the effects more precisely and and increase their value in informing policy. With ethnicity data available at baseline, the new experiment could stratify on this information to correct for the imbalances in the current sample, which would allow to study the heterogeneity in treatment effects due to ethnicity match between speakers and pupils. Moreover, since the balancing tests show insignificant, yet substantive, differences between treatment and control group, increasing the sample could also help with correcting these imbalances and confirming that the estimates in this study are unbiased.

Furthermore, to provide evidence on the long term impacts of raising pupils' aspirations on their educational attainment, future research should track long term progress and course choice of pupils involved in this and any follow-up experiments. Since the intervention in this study is small, it is conceivable that the positive effects on attitudes and test performance two months after treatment wear off over time, and more research into this area is needed to formulate meaningful policy recommendations. A better understanding of the role of parents and pupils' wider social network in influencing the long term evolution of the treatment effects should be crucial in this context.

Lastly, to further develop the research agenda on aspirations, methodological progress to develop a uniform measure of aspirations is necessary, to obtain a more comprehensive view on how aspirations interact with indicators of socioeconomic background, and demographics such as gender and ethnicity more globally. The literature documents positive effects of aspirations but all studies rely on their own measure and interpretation of aspirations. To derive relevant conclusions for the external validity from the findings in this and other studies, a more unified approach towards measuring aspirations is essential.

Acknowledgements This work has greatly benefited from invaluable discussions with Shan Aman-Rana, Nava Ashraf, Oriana Bandiera, Catia Batista, Tim Besley, Gharad Bryan, Svetlana Chekmasova, Alexia Delfino, Dita Eckhardt, Greg Fisher, Ria Ivandic, Eliana La Ferrara, Jay Lee, Martina Zanella, Celine Zipfel, and seminar participants at the Nova SBE research group, LSE Development and Growth seminar, STICERD work in progress, Warwick PhD conference, NCFAE at Pune University, and PODER Summer School at University of Namur. I am indebted to the team at the Education & Employers Taskforce, especially to Monty Karmakharm and Anthony Mann, and all the participating schools for their tireless support with this project. I am very grateful for excellent help with data entry to Rachel Fung, Pallavi Prabakhar, Jack Thiemel, and Dora Vaitsi. I greatly acknowledge funding for this project from the Fundaçao para a Ciencia e a Tecnologia (FCT).

Appendices

A Additional Tables

B Additional Figures

Table A.1: Interaction of treatment effect on test scores with regional gender norms

	Maths Year 6	Reading Year 6
	(1)	(2)
Girl*Female Speaker	3.19	2.01
	(0.07)*	(0.21)
	[0.15]	[0.30]
Boy*Male Speaker	0.17	1.91
	(0.87)	(0.36)
	[0.90	[0.47]
Girl*Male Speaker	0.17	-0.81
	(0.91)	(0.60)
	[0.93]	[0.67]
Boy*Female Speaker	1.03	1.12
	(0.48)	(0.49)
	[0.62]	[0.58]
Girl*Female Speaker*Work Rights	-4.84	-3.11
	(0.05)*	(0.14)
	[0.16]	[0.88]
Boy*Male Speaker*Work Rights	-0.14	-2.85
	(0.92)	(0.31)
	[0.88]	[0.10]*
Girl*Male Speaker*Work Rights	-0.52	0.95
	(0.80)	(0.64)
	[0.99]	[0.97]
Boy*Female Speaker*Work Rights	-1.58	-1.65
	(0.41)	(0.45)
	[0.55]	[0.32]
BL / Prior Attainment	Yes	Yes
Ethnicity FE	Yes	Yes
Boy Dummy	Yes	Yes
Work Rights	Yes	Yes
Observations	1139	1092
# clusters	30	30

Note: OLS P-values are clustered at the school level and reported in round brackets below the coefficients, and randomisation-t P-values from 2,000 draws are reported in squared brackets. The indicator *BL/Prior Attainment*, at the bottom of the table, refers to covariates measuring performance on year 2 tests, which are included in the regression to express treatment effects in terms of progress between year 2 and year 6 tests. Test scores are standardised to a standard deviation of one and mean zero in the control group. *Work Rights* refers to a binary variable based on question V44 in the 2005 survey wave of the World Values Survey (Inglehart et al. (2014)) (*Do you agree, disagree or neither agree nor disagree with the following statements? - When jobs are scarce, men should have more right to a job than women.*), where I obtain a binary measure by treating *neither agree nor disagree* as missing values. However, the estimations are not sensitive to this interpretation.

Figure B.1: Baseline questionnaire

Name: ____________________

Class teacher: ____________________

School Name: ____________________

SURVEY

1) How much time (in minutes) do you usually spend on homework per day? _______min

2) How much time (in minutes) do you think you **should** spend on homework per day? _______min

3) Did you take a Grammar school exam this year? Yes ☐ No ☐

How important is it for you to…	Very important 1	2	3	4	Not important at all 5
3) have a well-paying job when you are grown up?	☐	☐	☐	☐	☐
4) do very well in exams?	☐	☐	☐	☐	☐
5) be successful in your job when you are grown up?	☐	☐	☐	☐	☐

6) How far would you **LIKE** to go in school? (*Please tick **one** of the options*)

- ☐ I'd like to **leave school at 16 as soon as I have the chance.**
- ☐ I'd like to **leave school at 16 but with some qualifications.**
- ☐ I'd like to **stay in school until I'm 18 and do A levels.**
- ☐ I'd like to **stay in school until I'm 18 and do a course to prepare me for a job.**
- ☐ I'd like to **go to university.**

7) How far do you **THINK** you will actually end up going in school? (*Please tick **one** of the options*)

- ☐ I think I will **leave school at 16 as soon as I have the chance.**
- ☐ I think I will **leave school at 16 but with some qualifications.**
- ☐ I think I will **stay in school until I'm 18 and do A levels.**
- ☐ I think I will **stay in school until I'm 18 and do a course to prepare me for a job.**
- ☐ I think I will **go to university.**

During the next section, please think about yourself when you are grown up.

How likely is it that you will…	Very Likely 1	2	3	4	Not likely 5
8) have a well-paying job when you are grown up?	☐	☐	☐	☐	☐
9) do very well in exams?	☐	☐	☐	☐	☐
10) be successful in your job when you are grown up?	☐	☐	☐	☐	☐
11) get married and have a family?	☐	☐	☐	☐	☐
12) have children?	☐	☐	☐	☐	☐

13) If you would like to have children, have you thought about **how many**? ____________

14) **Which job would you like to get when you grow up**? ____________________

15) **Why** would you like to get this job? ____________________

____

How difficult will be for you to…	Very easy 1	2	3	4	Very Difficult 5
16) get this job? (the one you wrote in the last question)	☐	☐	☐	☐	☐
17) become a doctor?	☐	☐	☐	☐	☐
18) become a cafe worker?	☐	☐	☐	☐	☐
19) become a teacher?	☐	☐	☐	☐	☐
20) become an author?	☐	☐	☐	☐	☐
21) become a carer in a hospital?	☐	☐	☐	☐	☐

22) Which **words come to your mind** when you think about an **important exam**?

____

23) Do you think of **a man or woman** when thinking about someone successful in their job?

Man ☐ Woman ☐

24) **Why**? ____________________

____

25) What do you think will be **more important to you when you are grown up, family or job**?

Family ☐ Job ☐

Now, a few questions about the SATs exams you will be sitting this school year:

Which level would you **LIKE** to get in your…	Above expected	Expected	Below expected
26) English SAT exam?	☐	☐	☐
27) Maths SAT exam?	☐	☐	☐

Which level do you **THINK** you will get in your…	Above expected	Expected	Below expected
28) English SAT exam?	☐	☐	☐
29) Maths SAT exam?	☐	☐	☐

Finally, a few questions about yourself

30) Are you a **girl or boy**? ____________________

31) Which **languages** do you speak at home? ____________________

32) Which jobs do your **parents work**? ____________________

THANKS A LOT FOR YOUR ANSWERS!!

Figure B.2: Endline questionnaire

Name: ____________________

Class teacher: ____________________

School Name: ____________________

2. SURVEY

1) How much time (in minutes) do you usually spend on homework per day? _______min

2) How much time (in minutes) do you think you **should** spend on homework per day? _______min

3) Did you take a Grammar school exam this year? Yes ☐ No ☐

How much do you agree or disagree with the following statements?	Strongly agree 1	Agree 2	Neither agree nor disagree 3	Disagree 4	Strongly disagree 5
4) There is really no way I can do well in Maths.	☐	☐	☐	☐	☐
5) I have little control over how well I do in Maths.	☐	☐	☐	☐	☐
6) I can achieve just about any mark in Maths if I really set my mind to it.	☐	☐	☐	☐	☐
7) I often feel helpless when trying to become better at Maths.	☐	☐	☐	☐	☐
8) I am confident that I can do well in Maths if I work hard.	☐	☐	☐	☐	☐
9) There is little I can do to change how well I do in Maths.	☐	☐	☐	☐	☐

How important is it for you to...	Very important 1	Important 2	Neither important nor unimportant 3	Not important 4	Not important at all 5
10) have a well-paying job when you are grown up?	☐	☐	☐	☐	☐
11) do very well in exams?	☐	☐	☐	☐	☐
12) be successful in your job when you are grown up?	☐	☐	☐	☐	☐

How much do you agree or disagree with the following statements?	Strongly agree 1	Agree 2	Neither agree nor disagree 3	Disagree 4	Strongly disagree 5
13) I want to become better at Maths.	☐	☐	☐	☐	☐
14) Maths is my most dreaded subject.	☐	☐	☐	☐	☐
15) I don't feel confident about doing new things in Maths.	☐	☐	☐	☐	☐
16) I can think of many ways I can use Maths outside of school.	☐	☐	☐	☐	☐
17) I think studying Maths is useful.	☐	☐	☐	☐	☐

18) How far would you **LIKE** to go in school? (*Please tick **one** of the options*)

☐ I'd like to **leave school at 16 as soon as I have the chance.**
☐ I'd like to **leave school at 16 but with some qualifications.**
☐ I'd like to **stay in school until I'm 18 and do A levels.**
☐ I'd like to **stay in school until I'm 18 and do a course to prepare me for a job.**
☐ I'd like to **go to university.**

19) How far do you **THINK** you will actually end up going in school? (*Please tick **one** of the options*)

☐ I think I will **leave school at 16 as soon as I have the chance.**
☐ I think I will **leave school at 16 but with some qualifications.**
☐ I think I will **stay in school until I'm 18 and do A levels.**
☐ I think I will **stay in school until I'm 18 and do a course to prepare me for a job.**
☐ I think I will **go to university.**

During the next section, please think about yourself when you are grown up.

How likely is it that you will...	Very Likely 1	Likely 2	Neither likely nor unlikely 3	Unlikely 4	Not likely 5
20) have a well-paying job when you are grown up?	☐	☐	☐	☐	☐
21) do very well in exams?	☐	☐	☐	☐	☐
22) be successful in your job when you are grown up?	☐	☐	☐	☐	☐
23) get married and have a family?	☐	☐	☐	☐	☐
24) have children?	☐	☐	☐	☐	☐

25) If you would like to have children, have you thought about **how many**? ____________

26) **Which job would you like to get when you grow up?** ____________

27) **Why** would you like to get this job? ____________

How difficult will it be for you to...	Very easy 1	Easy 2	Neither easy nor difficult 3	Difficult 4	Very Difficult 5
28) get this job? (the one you wrote in the last question)	☐	☐	☐	☐	☐
29) become a doctor?	☐	☐	☐	☐	☐
30) become a cafe worker?	☐	☐	☐	☐	☐
31) become a scientist?	☐	☐	☐	☐	☐
32) become an engineer?	☐	☐	☐	☐	☐
33) become a teacher?	☐	☐	☐	☐	☐
34) become a computer programmer?	☐	☐	☐	☐	☐

35) Which **words come to your mind** when you think about an **important exam**?

____

36) Do you think of **a man or a woman** when thinking about someone who is successful in their job?

man ☐ woman ☐ both ☐

37) **Why**? ____________

38) Do you think of **a man or a woman** when thinking about someone who is good at Maths?

man ☐ woman ☐ both ☐

39) **Why**? ____________

40) What do you think will be **more important to you when you are grown up, family or job**?

family ☐ job ☐ both ☐

Now, a few questions about the SATs exams you will be sitting this school year:

Which level would you **LIKE** to get in your...	Above expected	Expected	Below expected
41) Reading SAT exam?	☐	☐	☐
42) Writing SAT exam?	☐	☐	☐
43) GPS SAT exam?	☐	☐	☐
44) Maths SAT exam?	☐	☐	☐
Which level do you **THINK** you will get in your...	**Above expected**	**Expected**	**Below expected**
45) Reading SAT exam?	☐	☐	☐
46) Writing SAT exam?	☐	☐	☐
47) GPS SAT exam?	☐	☐	☐
48) Maths SAT exam?	☐	☐	☐

49) In **the video you just watched**, what was the **name of the speaker**? ____________

Finally, a few questions about yourself

50) Are you a **girl or boy**? ____________

51) Which **languages** do you speak at home? ____________

52) Which jobs do your **parents work**? ____________

THANKS A LOT FOR YOUR ANSWERS!!

Note: Endline survey for the treatment group. The version of the control group is identical but misses question 49.

Figure B.3: Video part I

Note: Snippets from part one of the intervention video.

Figure B.4: Video part II

Note: The six speakers that each appear in part two of one of the six versions of the video.

Chapter 3

The Interplay of Risk Heterogeneity and Risk Sharing

Abstract
Subsistence farmers, without meaningful opportunities to save, rely heavily on steady agricultural yields for food consumption and income. Since formal insurance products are largely unavailable to smallholder farmers, informal risk sharing arrangements with other farmers are often the only means to compensate for shocks and to smooth consumption. Risk sharing arrangements allow households to pool part of their income to insure each other against adverse shocks, but gains from insurance crucially depend on the composition of the risk sharing group. In this paper, I focus on the role of households' risk profiles in determining the benefits from risk sharing. I develop a model in which the costs and benefits from risk sharing are summarised in the risk premium that a utility maximising agent is willing to pay to join the group. The model yields two testable implications: Risk sharing partners will have similar idiosyncratic shock variances, and idiosyncratic shocks of risk sharing partners will have negative covariance. To test these theoretical implications, I use data on the insurance network of a rural Tanzanian village consisting of 119 households. The empirical analysis confirms the theoretical implications that risk sharing partners will have similar shock variances and the idiosyncratic shocks have negative covariance. However, when testing for the joint correlation of the similarity in shock variances and their covariance, the correlation with shock variances remains significant, while the coefficient on the covariance of shocks is imprecisely estimated.

JEL codes: I3, O12, O17
Keywords: risk sharing, development, agriculture

3.1 Introduction

Subsistence farmers, without meaningful opportunities to save, rely heavily on steady agricultural yields for food consumption and income. However, without irrigation, agricultural production is subject to large variation due to changing weather conditions. Since formal insurance products are largely unavailable to smallholder farmers, informal risk sharing arrangements with other farmers are often the only means to compensate for shocks and to smooth consumption. Risk sharing arrangements work by allowing households to pool part of their income to insure each other against adverse shocks, but gains from insurance for risk sharing households crucially depend on the composition of the group. Therefore, to study the benefits arising from informal insurance it is important to define an optimal allocation of risk sharing partners in equilibrium, and then to test empirically whether insurance groups form according to these guidelines.

Risk sharing groups only persist if both households involved benefit from the agreement. The gains from insurance, in this paper, are defined based on households' idiosyncratic shock variances and the covariances of shocks between households. Intuitively, a household will want to share risk with another household if they are less risky or at most as risky as themselves. This arises since a household's own risk decreases through risk pooling if their risk sharing partner is less risky. Furthermore, insurance partners' idiosyncratic risks must be insurable by the arrangement, ie. they must not experience the same type of shock simultaneously. If two risk sharing households were to experience negative shocks at the same time, the risk pooling agreement would be of no use in smoothing consumption. In this paper, I formally derive the equilibrium allocation of risk sharing partners in terms of shock variances and covariances to capture both effects. I then conduct an empirical analysis of insurance links in a rural Tanzanian village to find that these indeed conform to the theoretical implications.

In the theoretical part of the analysis, I develop a model in which the costs and benefits from risk sharing are summarised in the risk premium that a utility maximising agent is willing to pay to join the group. The framework builds on Jaramillo et al. (2015) by relaxing the assumption of independently distributed idiosyncratic shocks to account for correlations between shocks that are likely to arise due to shared environmental conditions between farming households in a small village. Accordingly, the model shows that, in equilibrium, the risk premium increases in the shock variances and in the covariances of shocks of the two households. Households will choose to be in risk sharing groups for which the risk premium is smaller or equal to that of the single state where a household does not share risk. This yields two testable implications: Risk sharing partners will have similar idiosyncratic shock variances, and idiosyncratic shocks of risk sharing partners will have negative covariance.

To test these theoretical implications, I use data on the insurance network of a rural Tanzanian village consisting of 119 households. Exploring all possible bilateral, undirectional connections between households, this yields a dyadic data set of 7021 dyads in which each observation describes the relationship between two households in the village. Using information on the income sources of each household member, I construct proxies of the shock variance of a household and the covariance of shocks between any two households in the data. Based on the rationale that the shock variance of a household captures their level of riskiness, which is inversely related to the level of diversification of a household's income, I use the number of income sources per household member to approximate shock variances in the analysis. Similarly, I compute the covariance of two households' idiosyncratic shocks as the covariance of their income profiles weighted by the number of household members active in each income category. In the analysis, I regress a binary variable indicating whether two households share risk on the difference in the number of income categories and the covariance separately and jointly, controlling for dyad level characteristics like kinship and geographical distance.

The empirical analysis confirms the theoretical implications that risk sharing partners will have similar shock variances and the idiosyncratic shocks have negative covariance. In particular, the probability of two households being risk sharing partners is positively and significantly correlated with the similarity of, or negatively with the difference in, the number of income sources per household member. This correlation is robust to controlling for measures of wealth like land owning of both households and dyad level characteristics. This suggests that the correlation is not driven by the fact that wealthier households are more attractive and are therefore more likely to share risk which each other. Furthermore, I find that the probability of risk sharing is significantly negatively correlated with a covariance of greater than zero between idiosyncratic shocks. However, when testing for the joint correlation of the similarity in shock variances and their covariance, the coefficient on the covariance remains negative but loses significance, while the coefficient on the similarity of shock variances remains significant and barely changes in magnitude. I do not find evidence that this is driven by a lack of supply of potential risk sharing partners whose shocks have negative covariance.

The main contributions of this paper are to develop a model of the endogenous formation of risk sharing groups that allows for non-zero covariances between idiosyncratic shocks, and to test whether the equilibrium implications of the model hold empirically. In the theoretical analysis, I extend the model by Jaramillo et al. (2015) in two ways, first, by allowing for non-zero covariances between idiosyncratic shocks, and by assuming a network structure that arises from multiple, possibly overlapping, bilateral risk sharing groups per agent. By allowing for non-zero covariance between shocks, I am able to account for any interdependence of shocks due to a similar living and working environment. Contributions by Fafchamps and Lund (2003) and Fafchamps and Gubert (2007) discuss this hypothesis but this study is the first to formalise it and test the theoretical implications empirically.

This work fits into the literature on the endogenous formation of informal insurance networks. After early contributions by sociologists (amongst others J. Clyde Mitchell (ed.) (1969), and Raub and Weesie (1990)), seminal papers by economists Jackson and Wolinsky (1996) and Bala and Goyal (2000) marked the starting point of theoretical research on strategic network formation. The former present a theoretical analysis of the stability and efficiency of networks when these are formed endogenously. The latter paper models the decision to form networks as a trade-off between costs and benefits derived from the resulting links and consequently from any indirect links. However, these analyses do not focus on the peculiarities of informal insurance networks.

To fill this gap, Bramoullé and Kranton (2007) study specifically the formation of informal insurance networks. By exploring the effect of cross-community links on the shape and efficiency of risk sharing, they show that the existing empirical evidence for inefficient risk sharing at the village level, which I return to in the next paragraph, might be due to ignoring these inter-village arrangements. Finally, the model by Jaramillo et al. (2015) is an important building block for this analysis as they explore the effect of risk heterogeneity on the efficiency of risk pooling. They find that Nash bargaining leads to sorting into networks with low risk heterogeneity among members. I extend their model by relaxing the assumption that idiosyncratic shocks are independently distributed which offers theoretical insights into sorting along the dimension of covariances between idiosyncratic shocks.

In the absence of detailed information on network structures, early empirical studies assumed insurance networks to be exogenous, for instance Townsend (1994) used villages, and Grimard (1997) cultural identifiers such as ethnicity or caste, to proxy for social networks. The former finds that idiosyncratic shocks explain little of the variation in household consumption and suggests that this is achieved through mutual insurance in the village, and Grimard (1997) also finds similar evidence for households linked through ethnic ties.

Fafchamps and Lund (2003), Fafchamps and Gubert (2007), Udry and Conley (2004), and De Weerdt and Dercon (2002) use micro data on reported links between individuals to obtain a more accurate mapping of networks. The first paper uses data on informal insurance networks from the Philippines to examine co-movements between individual shocks and transfers that network members send or receive. The authors find strong evidence for a causal relationship between individual shocks, and gifts and loans, and further that risk sharing takes place within networks of friends and family rather than at the village level. Fafchamps and Gubert (2007) use the same data to study how individual characteristics affect the probability of a link and find strong evidence that geographical and social proximity, age and wealth are important predictors of network links, but not negative correlations in occupations and income, which they measure as farming versus non-farming. Notwithstanding, this paper presents some evidence that a negative covariance between income categories within farming is correlated with risk sharing. Furthermore, Fafchamps and Gubert (2007) contribute methodologically by developing a covariance formula that allows to calculate standard errors that account for autocorrelation in the error terms of dyads that contain the same household, a method that I use to calculate standard errors in the empirical analysis of this paper.

Udry and Conley (2004) map networks among farmers in Ghana to find that the probability of a link is higher for individuals who reside within short geographical distance and follow the same matrilineage. Finally, De Weerdt and Dercon (2002), using the same data that I use, find consistent results with Udry and Conley (2004) that kinship, geographical proximity, cultural background and wealth increase probabilities for network links among individuals. This paper builds on this literature by explicitly relating risk sharing to the shock variances and covariances of households.

Finally, three papers are to be noted that discuss insurance mechanisms in the context of poverty alleviation and development of the least developed. Dercon and Christiaensen (2011) and Karlan et al. (2014) emphasise the importance of lifting insurance constraints in order to foster technology adoption and business investments. The former find that a higher propensity to experience low harvests impedes technology adoption and dampens agricultural risk taking. The authors highlight that insurance arrangements can increase the propensity to innovate by reducing risk due to income volatility. Karlan et al. (2014) focus specifically on the distinction between cash and insurance constraints in restraining investment in agricultural production and they find that constraints to risk insurance pose the greater obstacle on investment than the pure lack of financial liquidity. These results emphasise the relevance of research on risk insuring mechanisms in order to derive policies aimed at promoting rural and agricultural development of the poorest. Batista and Vicente (2018) study the introduction of mobile money in Mozambique and find that offering insurance through mobile money decreases investments in agricultural production and likely leads to occupational change. As this effect appears to be driven by credit constraints, it is in line with the theoretical predictions in Karlan et al. (2014), but it illustrates that insurance can have adverse effects.

I proceed by presenting the theoretical framework in section 3.2 and deriving testable implications in section 3.2.2. A detailed description of the network data from Tanzania follows in section 3.3, and I illustrate the estimation and methods in section 3.4. The empirical findings are presented in section 3.5 and discussed in section 3.6. Section 3.7 concludes by summarising the main findings and suggesting avenues for future research.

3.2 Theoretical Framework

The framework in this paper is an extension of the model by Jaramillo et al. (2015), which models the role of idiosyncratic shock variances in the optimal formation of risk sharing agreements under the assumption that covariances between idiosyncratic shocks are zero. In a variant of this model, I relax this assumption and define the optimal risk sharing rule when idiosyncratic shocks are not independent.

3.2.1 Utility from Risk Sharing

Agents form bilateral risk sharing agreements in order to reduce the effect of adverse shocks on utility via consumption.[1] Following Jaramillo et al. (2015), agent i in risk sharing group j solves a CARA expected utility maximisation problem with exponential utility.

$$\max_{\{c_{ij}(\varepsilon_j)\}} U_{ij} = -E\left[\frac{1}{\alpha} exp\left\{-\alpha c_{ij}(\varepsilon_j)\right\}\right]$$

The parameter α is the coefficient of absolute risk aversion. Consumption is denoted as c_{ij} and is a function of the realisation of shocks ε_j, and the state of nature in group j is defined as $\varepsilon_{jt} \equiv (v, e_i, e_{-i})$. The common income shock that the group faces is $v \sim (0, \sigma_v^2)$, the idiosyncratic income shock of agent i is $e_i \sim (0, \sigma_{e_i}^2)$, and $e_{-i} \sim (0, \sigma_{e_{-i}}^2)$ is the income shock of the risk sharing partner, where σ^2 denotes the variance. Both idiosyncratic shocks are normally distributed.

Consumption in any given realisation immediately depends on the risk sharing rule between partners. I assume an equal sharing rule that is exogenously given by

$$c_j = \kappa_i + \frac{e_i + e_{-i}}{2} + v \tag{3.1}$$

Jaramillo et al. (2015) show that this risk sharing rule can be obtained through

[1] I discuss one possible matching mechanism in section A of the appendix.

Nash bargaining.[2] The fixed portion of income that an agent allocates to the risk sharing group is κ_i. Regardless of the number of risk sharing groups that agent i forms, the portion of income κ_i is fixed for all agents. To ensure that an agent can meet all their liability at any time, κ_i needs to be small enough such that $\kappa_i < \frac{1}{p_i}$, where p_i is the number of risk sharing agreements.[3] The expected indirect utility from sharing risk in group j under the insurance rule in equation 3.1 is given by

$$V_{ij} = -E\left[\frac{1}{\alpha}e^{-\alpha(\kappa+\frac{e_i+e_{-i}}{2}+v)}\right] \tag{3.2}$$

To express the resulting risk from this insurance agreement in monetary terms, I rewrite it as the risk premium that the utility maximising agent is willing to pay to enter the agreement. The premium attaches a price to the risk of an agreement, and the higher the risk premium of a risk sharing group, the higher the associated risk. Arrow (1963) and Pratt (1964) define the risk premium of group j as

$$\Pi_j = \frac{1}{2}\alpha\sigma^2_{\varepsilon_j}$$

where $\sigma^2_{\varepsilon_j}$ is the variance of the shocks in the insurance rule. Since I assume CARA utility and normally distributed shocks, the Arrow-Pratt formula for the risk premium applies without approximation error and the indirect utility from sharing risk in j becomes

$$\begin{aligned} V_{i,j} &= -\frac{1}{\alpha}e^{\alpha\left[-\kappa+\frac{\alpha}{2}\left(\frac{1}{4}\sum_{j=\{i,-i\}}\sigma^2_{e_j}+\frac{1}{2}\mathrm{Cov}(e_i,e_{-i})+\sigma^2_v\right)\right]} && (3.3)\\ &= -\frac{1}{\alpha}e^{\alpha(-\kappa+\Pi_j)} && (3.4)\end{aligned}$$

The risk premium increases in the variances of idiosyncratic shocks of both risk sharing partners, in the variance of the common shock, and in the covariance between the two idiosyncratic shocks.

[2] The solution is conditional on (i) commitment in the allocation of resources once a group has been formed, and (ii) pareto optimal allocation of resources amongst members of a group.

[3] I discuss the implications of this assumption in appendix A.

Utility maximising agents will enter a risk sharing agreement if their utility from sharing risk is higher than in the single state. The single state refers to the situation in which an agent does not share risk and instead bears their risk alone. Equation 3.4 shows that in this model, the utility from sharing risk in a given risk sharing group is fully described by its risk premium. To illustrate how the decision to share risk changes when the premium accounts for non-zero covariances between idiosyncratic shocks of insurance partners, it is therefore sufficient to compare the risk premia in equation 3.5 - 3.7.[4]

$$\Pi_j^{single} = \frac{\alpha}{2}\left[\sigma_i^2 + \sigma_v^2\right] \tag{3.5}$$

vs.

$$\Pi_j^{id} = \frac{\alpha}{2}\left[\frac{1}{4}\sum_{j=\{i,-i\}}\sigma_j^2 + \sigma_v^2\right] \tag{3.6}$$

vs.

$$\Pi_j^{cov} = \frac{\alpha}{2}\left[\frac{1}{4}\sum_{j=\{i,-i\}}\sigma_j^2 + \frac{1}{2}\mathrm{Cov}(e_i, e_{-i}) + \sigma_v^2\right] \tag{3.7}$$

In the single state, in equation 3.5, the agent does not share their idiosyncratic risk σ_i, whereas equation 3.6 and 3.7 show the risk premia when two agents do enter a risk sharing agreement. The difference between these two equations is that Π_j^{id} ignores the covariance between idiosyncratic shocks $\mathrm{Cov}(e_i, e_{-i})$, whereas Π_j^{cov} accounts for it. Since the risk premium increases in the covariance between idiosyncratic shocks, the utility from sharing risk decreases in the covariance. This means that two agents who face shocks that have positive covariance are less suitable as risk sharing partners as the probabilities of experiencing a negative shock are positively correlated. Ideally, risk sharing partners should have negatively related shocks so that when one faces a negative shock, the risk sharing partner can share their positive income shock to compensate. Furthermore, the comparison of risk

[4]Note that since by definition the common shock is independently distributed from idiosyncratic shocks, the covariances between the common and idiosyncratic shocks are zero $Cov(v, e) = 0$.

premia shows that the variance of the village-wide shock, σ_v^2, which is defined as the intersection of all idiosyncratic shocks, is common to all states, which implies that it is uninsurable through mutual insurance as it affects all agents.

Equation 3.7 illustrates that an agent picks their risk sharing partner based on the variance of the partner's idiosyncratic shock, and on the covariance between the partner's and their own idiosyncratic shock. Since the utility from risk sharing can be summarised by the risk premium alone, an agent will enter a risk sharing group if the risk premium of the group is lower than that of the single state $\Pi^{cov} < \Pi^{single}$.

Proposition 3.2.1 *In equilibrium with all other variables equal, risk premia in risk sharing groups for individuals $i \in j$ and $i' \in j'$, with $\sigma_i^2 < \sigma_{i'}^2$, will be such that $\Pi_j^{cov} \leq \Pi_{j'}^{cov}$*

The proof follows directly from equation 3.7.

Proof With all other variables equal, $\sigma_i^2 < \sigma_{i'}^2$ implies $\sum_{j=\{i,k\}} \sigma_j^2 < \sum_{j'=\{i',k\}} \sigma_j'^2$, where k is any third agent in the economy. Hence, if i and i' are members of two distinct risk sharing groups j and j', then $\Pi_j^{cov} \leq \Pi_{j'}^{cov}$ follows. ■

Since the utility of an agent is highest when the risk premium is small, an agent will seek to enter the risk sharing group in which the sum of shock variances is smallest. Whilst any agent will always prefer to share risk with agents whose shock variance is small, the agent themself will only be accepted in a risk sharing group if the risk premium from sharing for the partner is lower than in their single state. Therefore, matching between agents following this premise yields risk sharing groups that consist of members with similar shock variance or riskiness.

The following proposition concerns the role of covariances between idiosyncratic shocks.

Proposition 3.2.2 *In equilibrium with all other variables equal, risk sharing coalitions will be such that for $i \in j$ and $i' \in j'$, with* $\mathrm{Cov}(e_i, e_k) < \mathrm{Cov}(e_{i'}, e_k)$*, then* $\Pi_j^{cov} \leq \Pi_{j'}^{cov}$*.*

Proof $\mathrm{Cov}(e_i, e_k) < \mathrm{Cov}(e_{i'}, e_k)$ directly implies $\Pi_{i,k}^{cov} \leq \Pi_{i',k}^{cov}$. Hence, k prefers i to i' as a risk sharing partner. If i and i' are members of two distinct risk sharing groups, the group of i will be the one that generates the lower risk premium in comparison. ■

Utility maximising agents prefer risk sharing partners whose idiosyncratic shock is non-positively correlated with their own shock. If shocks are positively correlated, the materialisation of one idiosyncratic shock increases the probability of the other one materialising too. In this case risk sharing would be futile. However, in a risk sharing group where the idiosyncratic shocks of the two partners have zero or negative covariance, the two shocks are independent or counteract each other. The risk premium reflects this relationship by increasing in a positive, and decreasing in a negative covariance between idiosyncratic shocks.

3.2.2 Testable Implications

The model suggests that risk sharing partners who match according to utility maximisation principles will have the following characteristics:

Testable Implications

I *Similar variances of idiosyncratic shocks*, and

II *Negatively correlated idiosyncratic shocks.*

The first implication arises from proposition 3.2.1. Intuitively, it draws on the rationale that an agent prefers to share risk with another agent who exhibits a similar level of riskiness as expressed by the idiosyncratic shock variance. Two similarly risky agents can reduce their risk premium by sharing risk relative to the single state. However, forming a risk sharing group with a partner who is substantially more risky will lead to a risk premium that is higher than that in the single state. The second implication arises from proposition 3.2.2 and it concerns insurability. The ideal risk sharing partner experiences economic prosperity when the agent themself faces a negative shock. This is to ensure that risk sharing partners are capable of compensating one another for shock related losses.

3.3 Data & Sample Characteristics

I test these implications using data from De Weerdt (2004), which contain a complete map of the insurance network of a small Haya village in Tanzania. The dyadic data, in which each observation is a description of the relationship between two households, have information on all insurance links between households within the village. Figure 3.1 illustrates the network structure graphically.

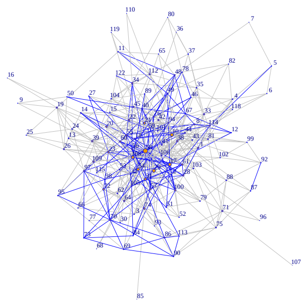
3.3.1 Data Description

The strengths of links in the data vary from 'no link', 'unilateral link', to 'reciprocal link'. Since the network is undirectional, which implies that the link of A to B is identical to the one of B to A, 'unilateral' and 'reciprocal' do not indicate the direction of a link but rather the strength of the tie. This is due to the phrasing of the network identification question, which infers reciprocity: *"Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour"*. This insurance link variable serves as the endogenous variable in the empirical analysis. The dyadic data also inform about kinship ties and geographical distance between any two households.

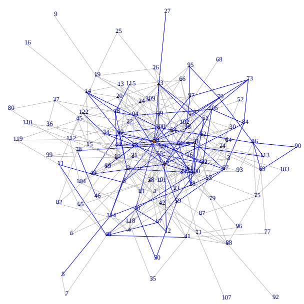
In addition to the dyadic data, I have survey data on each household in the village. These include information on clan and religious membership of the household,

Figure 3.1: Risk sharing network structure

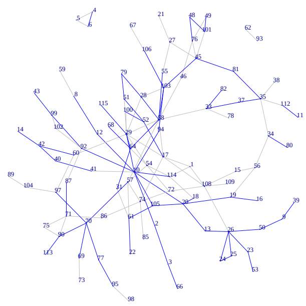
(a) All links



(b) Unilateral links



(c) Reciprocal links



Note: Illustration of the informal insurance network split base on all links in panel A, only unilateral links in B, and only reciprocal links in C.

weekly consumption, land ownership, and the value of livestock. Furthermore, for each adult in the household there is data on age, education, sex and their income portfolio. The income portfolio data consists of seven categories covering income from livestock, trade, assets, processing of agricultural produce, casual labour and income from other off-farm work. Each category is represented by a dummy variable that indicates whether a household member accrues income from it. This information allows to derive a measure of income diversification to proxy for the shock variance of a household and the covariance between the idiosyncratic shocks of two households.

3.3.2 Sample Characteristics

In table 3.1, I present descriptive statistics for the sample at the household level, which consists of 119 households. On average, a household consists of 2 adult household members, and in 11% of households, none of the members have education. In 17% of households, the most educated household member started primary school, and in 57% the most educated household member finished primary school education. In 8% of households, at least one member has secondary school education, and for 8%, information on education is missing. On average, the oldest household member in a household is 46 years old. Mean weekly food consumption across five rounds of surveys is TZS 1308 (approx. EUR 1.64 on 1 Jan 2000), non-food consumption is TZS 314.25 and total consumption is TZS 1622.71. The average household owns 1.35 acres of land and livestock worth TZS 54659.66. In total, a household is active in three income categories, and each household member in two on average. 50% of households earn income from casual labour, 36% from trade, 93% from crops, 27% from livestock, 0.6% from other assets, 35% from processing agricultural produce, and 34% from other off farming activities. Finally, each household has on average 5.38 bilateral risk sharing agreements.

Table 3.1: Descriptive Statistics

	Mean	Std. Dev.
Panel A: Demographics		
# Households	119	
# Adult HH Members	1.77	0.65
No Education	0.11	0.31
Started Primary School	0.17	0.38
Completed Primary School	0.57	0.50
(Some) Secondary School	0.08	0.28
Age Oldest HH Member	45.72	16.50
Panel B: Consumption		
Food Consumption (TZS)	1301.14	462.83
Non-Food Consumption (TZS)	310.55	341.67
Total Consumption (TZS)	1611.83	675.84
Panel C: Wealth		
Land owning (acres)	1.34	1.26
Livestock (TZS)	53747.46	187187.34
Panel D: Income		
# Income Categories	2.80	1.14
# Income Categories PP	2.04	0.74
Other Off Farm	0.35	0.48
Casual Labour	0.50	0.50
Trade	0.35	0.48
Crop	0.93	0.26
Livestock	0.27	0.44
Assets	0.06	0.24
Processing	0.35	0.48
Panel E: Network		
# Links	5.38	3.02

Note: Consumption in panel B and value of livestock in panel D are expressed in terms of Tanzanian Shillings.

3.4 Estimation & Methods

3.4.1 Proxies for Shock Variances and Shock Covariances

In the empirical analysis, I examine whether the theoretical results regarding the role of risk heterogeneity in the formation of risk sharing agreements is borne out by the data. The model suggests that in equilibrium only those risk sharing agreements that make the partners better off than the single state exist. All else equal, this will be the case when risk sharing partners have similar idiosyncratic shock variances, ie. face similar levels of riskiness, and their shocks have negative covariance. I test whether two households who identify as risk sharing partners have on average more similar shock variances than those households who do not, and if a negative covariance of shocks predicts risk sharing.

I construct proxies to analyse each of the two predictions. First, as a measure of a household's idiosyncratic shock variance, I use data on income categories to build a proxy based on the income diversification of a household. I argue that in the context of rural farmers the riskiness of a household is reflected in the diversification of their income profile. The more diversified the composition of the income profile, the less sensitive the household is to shocks that affect individual income categories. I measure the level of diversification as the average number of income sources across all household members.

Secondly, to measure the covariance of idiosyncratic shocks, I compute the covariance of income profiles weighted by the number of household members active in each category. I compute the covariance between the income profile of household A and B as

$$\mathrm{Cov}_{A,B} = \sum_{c=1}^{C} \frac{(n_{A,c} - \bar{n}_A)(n_{B,c} - \bar{n}_B)}{C}$$

where c refers to the different income categories, $n_{.,c}$ to the number of people per household working in income category c, and $\bar{n}$ is the average number of people working in a category per household.

This variable is a measure of the degree to which the income profiles of two households are interdependent. If the covariance is positive for two households, their income profiles are related as they rely on the same income source to a similar extent in terms of total income shares. If the covariance is negative, the opposite is true and if the covariance equals zero, there is no relationship between the two income profiles. To proxy for the probability that two households are likely to suffer similar shocks, I use a binary variable that is equal to one if the covariance is greater than zero, and zero otherwise.

As the outcome variable in all regressions, I use the binary link variable in the dyadic data set that indicates whether a link between two households exists, treating both unilateral and reciprocally links as identical.

3.4.2 Methods

Fafchamps and Gubert (2007) identify two issues regarding identification and inference with estimating a dyadic model like this one. In what follows, I use largely the same notation as the authors for the ease of reference.

Identification

In a dyadic data set, each observation is a description of the link between two households k and l, and contains information on link attributes W_{kl}, and information on each household in the dyad (Z_k, Z_l). The regressors must enter the dyadic regression in a way that ensures that the effect of dyad attributes (Z_k, Z_l) on y_{kl} is the same as the effect of (Z_l, Z_k) on y_{lk}. Since this paper uses undirectional data, which means that $y_{kl} = y_{lk}$, I can rewrite this requirement as $\beta Z_{kl} = \beta Z_{lk}$. The regression satisfies this criterion by including both the sum and the difference of the household attributes as regressors, which leads to estimating an equation of the form

$$y_{kl} = \alpha + \beta_1 |Z_k - Z_l| + \beta_2 (Z_k + Z_l) + \phi |W_{kl}| + \eta_{kl}$$

Even though the main coefficients of interest in this study are β_1 and ϕ and the identification issue concerns β_2, I briefly discuss it for completeness. Identification of β_2 hinges on variation in the number of insurance links that a household maintains as the coefficient on the sum of household attributes is only identified if the number of links varies across the sample. To provide an intuitive illustration of this issue, I paraphrase the example given in Fafchamps and Gubert (2007): If the dyadic data in this sample consisted entirely of risk sharing households with strictly one risk sharing partner, I could estimate with β_1 whether households with more similar shock variances are more likely to share risk, but not whether households with higher shock variance are more likely to share risk, since all households have only one risk sharing partner. In other words, the estimation would allow to identify the influence of the differences in attributes but not the influence of the sum. As shown in table 3.1, the number of links in this sample varies, which implies that both β_1 and β_2 are identified in my regressions.

Inference

The second issue arises due to the fact that there are idiosyncratic factors affecting the outcome that are uniform for all observations involving a individual. This implies that $E(\eta_{kl}, \eta_{km})$ is likely not zero for all k. Similarly, I have to expect that $E(\eta_{kl}, u_{ml}) \neq$ for all l, $E(\eta_{kl}, \eta_{mk}) \neq 0$, and $E(\eta_{kl}, \eta_{lm}) \neq 0$. I correct for this correlation by calculating the standard errors using a covariance matrix that accounts for cross-observation correlation as proposed by Fafchamps and Gubert

(2007).[5]

3.4.3 Estimation

This leads to the following set of models that I use to test the implications from the model in section 3.2.2. I estimate equation 3.8 to test whether risk sharing partners have similar shock variances. The outcome variable y_{kl} indicates whether households k and l form a risk sharing group and the coefficient β_1 estimates the correlation between the difference in the number of income categories per person of the two households k and l, and β_2 the sum of categories. The coefficient β_1 is the main coefficient of interest in this equation and the model predicts that it assumes a negative sign. The control vector W_{kl} contains dyad level controls like geographical distance and kinship relationship between household k and l, and η_{kl} is the error term.

$$\begin{aligned}
y_{kl} &= \alpha + \beta_1 |\text{No. income sources}_k - \text{No. income sources}_l| + \qquad (3.8)\\
&\quad +\beta_2(\text{No. income sources}_k + \text{No. income sources}_l) + \phi|W_{kl}| + \eta_{kl}\\
y_{kl} &= \alpha + \gamma_1 |\text{Covariance}_{kl}| + \phi|W_{kl}| + \eta_{kl} \qquad (3.9)\\
y_{kl} &= \alpha + \delta_1 |\text{No. income sources}_k - \text{No. income sources}_l| + \qquad (3.10)\\
&\quad +\delta_2(\text{No. income sources}_k + \text{No. income sources}_l) +\\
&\quad +\delta_3 \text{Covariance}_{kl} + \phi|W_{kl}| + \eta_{kl}
\end{aligned}$$

Similarly, in equation 3.9, I test the second theoretical implication by estimating the correlation between the probability of the household k and l forming an insurance group and the indicator $Covariance_{kl}$ that equals one if the covariance of the income

[5] Fafchamps and Gubert (2007) provide a STATA command to estimate the standard errors according to the covariance

$$\text{A Var}(\hat{\beta}) = \frac{1}{N-M}(X'X)^{-1}\left(\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\frac{g_{klmn}}{2N}X_{kl}\eta_{kl}\eta'_{mn}X_{mn}\right)(X'X)^{-1}$$

The STATA command is called *nreg* and is available as ado file from their website.

profiles of the two households is greater than zero. The model predicts that γ_1 assumes a negative value. Finally, I estimate 3.10 to test whether the similarity of shock variances and the covariance of shocks are both jointly significant predictors of risk sharing.

3.5 Results

The results from estimating equation 3.8 are presented in table 3.2, and from equation 3.9 and 3.10 in table 3.3. I estimate each equation with OLS and unclustered standard errors in the top panel, and with dyadic standard errors following Fafchamps and Gubert (2007) in the bottom panel. I also estimate all the equations with a Probit model in table B.1 in appendix B to show that the results are qualitatively similar to the OLS.

The results in table 3.2 document that the likelihood of two households sharing risk increases in the similarity of their shock variances. In particular, households with a similar number of income sources per adult household member are more likely to form risk sharing groups. Since wealthier households are less sensitive to idiosyncratic income shocks as they are better able to compensate with other means, I also show that this relationship is significant when I use measures of wealth, such as land owning, as an alternative proxy for the shock variance. Interestingly, the coefficient on the difference in land owning switches signs when I control for the sum. This suggests that the positive effect of the difference, without controlling for the sum, is driven by omitted variable bias, which likely stems from the fact that wealthier households are more attractive as risk sharing partners.

To show that the correlation of risk sharing and similar levels of diversification is not driven by wealth effects, I also present estimations of the joint model of income diversification and land owning with dyad level controls in the last column. Since

Table 3.2: Differences in shock variances

	Probability of Risk Sharing				
	(1)	(2)	(3)	(4)	(5)
Panel A: Unclustered Std. Errors					
Δ(# Income Sources)	-0.0089*	-0.0113**			-0.0106**
	(0.005)	(0.005)			(0.005)
$\sum$(# Income Sources)		0.0050			0.0039
		(0.003)			(0.003)
Δ(Land Owning)			0.0071***	-0.0150***	-0.0123***
			(0.002)	(0.004)	(0.004)
$\sum$(Land Owning)				0.0213***	0.0211***
				(0.003)	(0.003)
Observations	6328	6328	7021	7021	6328
Panel B: Dyadic Std. Errors					
Δ(# Income Sources)	-0.0089*	-0.0113*			-0.0106*
	(0.005)	(0.006)			(0.005)
$\sum$(# Income Sources)		0.0050			0.0039
		(0.005)			(0.004)
Δ(Land Owning)			0.0071	-0.0150**	-0.0123*
			(0.006)	(0.007)	(0.006)
$\sum$(Land Owning)				0.0213***	0.0211***
				(0.006)	(0.006)
Observations	12656	12656	14042	14042	12656
Kinship					Yes
Geo. Distance					Yes

Note: The outcome variable is a binary measure of insurance links that treats unilateral and reciprocal links as equivalents. Δ(# Income Sources) is the difference in the average number of income sources per adult household member between the two households in the dyad, $\sum$(# Income Sources) is the sum of the average number of income sources per adult household member. Analogously for Δ(Land Owning) and Δ(Land Owning). Standard errors are reported in brackets below the coefficients. In panel A, standard errors are unclustered, and in panel B, I follow Fafchamps and Gubert (2007) in accounting for any correlation between the error terms of all observations involving one household by calculating standard errors according to the variance formula they propose, which also requires to double the dyadic data by adding the sample in reversed dyad sorting.

both coefficients remain significant, I conclude that similar levels of diversification in income sources and similar wealth levels jointly predict risk sharing conditional on covariates. These findings represent strong support for the first testable implication that risk sharing partners in equilibrium will have similar shock variances.

In the first two columns of table 3.3, I present results for the second implication of the model that the covariance of the idiosyncratic shocks of risk sharing partners is negative. The estimates show a negative correlation between the covariance of income profiles and the probability of risk sharing, thereby confirming the prediction of the model. In the top panel, the coefficient becomes larger in absolute value and significant at 10% when controlling for kinship relationship and geographical distance between the households.

In the last two columns of table 3.3, I estimate equation 3.10 by including the difference in the number of income sources and the sum thereof as independent variables. I find that the estimated relationship between the covariance and the probability of risk sharing remains negative, though the estimates become insignificant. The estimated correlation between the difference in number of income sources and the outcome variable in this table is almost identical compared to the estimations in table 3.2. The estimates in the bottom panel with dyadic standard errors show that the findings are largely robust even though some lose significance due to the clustering. Overall, these findings provide some evidence that households sort into risk sharing groups with households such that their income profiles have negative covariance but the correlation is imprecisely estimated.

To understand whether the lack of sorting to achieve a negative shock covariance arises due to a shortage in supply of adequate risk sharing partners, I test if wealthier, and thus more attractive, households who have a greater choice of risk sharing partners, are more likely to adhere to the criterion. By interacting the covariance

Table 3.3: Covariances of risk profiles, and differences in shock variances

	Probability of Risk Sharing			
	(1)	(2)	(3)	(4)
Panel A: Unclustered Std. Errors				
Δ(# Income Sources)			-0.0113**	-0.0114**
			(0.005)	(0.005)
$\sum$(# Income Sources)			0.0049	0.0048
			(0.003)	(0.003)
Income Covariance	-0.0094	-0.0118*	-0.0048	-0.0104
	(0.007)	(0.007)	(0.008)	(0.007)
Observations	7027	6670	6328	6328
Panel B: Dyadic Std. Errors				
Δ(# Income Sources)			-0.0113*	-0.0114*
			(0.006)	(0.006)
$\sum$(# Income Sources)			0.0049	0.0048
			(0.005)	(0.005)
Income Covariance	-0.0093	-0.0118	-0.0048	-0.0104
	(0.008)	(0.008)	(0.007)	(0.007)
Observations	14042	13340	12656	12656
Kinship		Yes		Yes
Geo. Distance		Yes		Yes

Note: The outcome variable is a binary measure of insurance links that treats unilateral and reciprocal links as equivalents. Δ(# Income Sources) is the difference in the average number of income sources per adult household member between the two households in the dyad, $\sum$(# Income Sources) is the sum of the average number of income sources per adult household member. Standard errors are reported in brackets below the coefficients. In panel A, standard errors are unclustered, and in panel B, I follow Fafchamps and Gubert (2007) in accounting for any correlation between the error terms of all observations involving one household by calculating standard errors according to the variance formula they propose, which also requires to double the dyadic data by adding the sample in reversed dyad sorting.

Table 3.4: Interaction of covariance with wealth and education

	Probability of Risk Sharing	
	(1)	(2)
Panel A: Unclustered Std. Errors		
Δ(# Income Sources)	-0.0105**	-0.0101**
	(0.005)	(0.005)
$\sum$(# Income Sources)	0.0045	0.0035
	(0.003)	(0.003)
Income Covariance	-0.0189	-0.0384
	(0.013)	(0.037)
Covariance*$\sum$(Land Owning)	0.0037	
	(0.004)	
Covariance*$\sum$(Education)		0.0054
		(0.007)
Observations	6328	6105
Panel B: Dyadic Std. Errors		
Δ(# Income Sources)	-0.0105*	-0.0101*
	(0.006)	(0.006)
$\sum$(# Income Sources)	0.0045	0.0035
	(0.004)	(0.005)
Income Covariance	-0.0189	-0.0384
	(0.012)	(0.028)
Covariance*$\sum$(Land Owning)	0.0037	
	(0.004)	
Covariance*$\sum$(Education)		0.0054
		(0.006)
Observations	12656	12210
Kinship	Yes	Yes
Geo. Distance	Yes	Yes
$\sum$(Land Owning)	Yes	
$\sum$(Education)		Yes

Note: The outcome variable is a binary measure of insurance links that treats unilateral and reciprocal links as equivalents. Δ(# Income Sources) is the difference in the average number of income sources per adult household member between the two households in the dyad, $\sum$(# Income Sources) is the sum of the average number of income sources per adult household member. Analogously for $\sum$(Land Owning) and $\sum$(Education). Standard errors are reported in brackets below the coefficients. In panel A, standard errors are unclustered, and in panel B, I follow Fafchamps and Gubert (2007) in accounting for any correlation between the error terms of all observations involving one household by calculating standard errors according to the variance formula they propose, which also requires to double the dyadic data by adding the sample in reversed dyad sorting.

of risk profiles with total land owning and education of the dyad in table 3.4, I estimate whether the correlation between risk sharing and a negative shock covariance is stronger for wealthier and more educated dyads. The estimations show that neither of these two interactions are significant, which implies that there is no heterogeneity in the correlation by wealth and education.

3.6 Discussion

3.6.1 Interpretation of Main Results

The theoretical discussion regarding the role of risk profiles in the endogenous formation of risk sharing groups suggests two equilibrium results: Agents will share risk with partners who exhibit a similar level of riskiness, and the idiosyncratic shocks of risk sharing partners will have negative covariance. The empirical analysis shows that these predictions hold true in the case of the insurance network in a village in rural Tanzania. Controlling for dyad characteristics that influence the probability of a link regardless of the risk profile, like kinship and geographical distance, I find that households with similar levels of riskiness and negative risk covariance are more likely to share risk.

In the context of rural subsistence farmers, like the individuals in this sample, idiosyncratic risk is largely related to diversification of income. Households who are unable to save part of their income have little or no wealth to draw on during times of economic hardship. In order to minimise risk exposure, households will therefore endeavour to spread their income over as many sources as possible. The more diversified the income, the less they will suffer from an adverse shock on one of their income sources. To account for this relationship between riskiness and income diversification, I proxy for the idiosyncratic shock variance by using the number of income sources per adult household member, and find a significant correlation with the probability of risk sharing.

While the correlation between the difference in the number of income sources per household member and risk sharing could also be driven by a wealth effect, the estimations in table 3.2 confirms that the correlation remains significant even when I control for the sum of income sources and wealth via land owning.

Similarly, I compute a measure of the covariance of idiosyncratic shocks based on households' income profiles and show that there is a significant negative correlation between risk sharing and the covariance of household income as expected. However, significance is not robust to including the difference in shock variances as a regressor, which suggests that households are less likely to adhere to the theoretical implication that risk sharing should optimally happen between households whose shocks have negative covariance.

This may be due to a variety of reasons. First, risk sharing relies on trust and so risk sharing partners will likely want to know each other well. Outside of kinship and neighbourhood ties, those people might be business partners or colleagues - people whose income profiles will inevitably be positively correlated. Or the pool of potential risk sharing partners is limited and not all households can choose risk sharing partners who satisfy both criteria. The interactions in table 3.4 do not substantiate this hypothesis. If wealthier households are more attractive as risk sharing partners, and are therefore the first to choose from the pool of potential partners, one would expect the interaction to be negative and significant. However, the correlation between income covariance and risk sharing does not vary significantly with wealth or education.

3.6.2 Alternative Measures of the Shock Variance and Covariance

The empirical proxies of shock variances and covariances are based on income categories of household members in the village. I select these measures to reflect the level of income diversification in a household to capture the shock variance, and the covariance of income profiles to grasp the covariance between shocks. However, I do not claim that these are the only conceivable proxies, or the best ones.

An arguably better approximation would be one that uses the monetary income shares of each income category for each household member to measure the intensive margin of diversification in addition to the extensive margin measured by binary income category indicators. In the absence of data on relative income from each category, the current measures might be misleading if income shares within households are very skewed so that the number of categories does not adequately reflect the diversification of income, and the covariance does not capture the actual covariance between income profiles.

To test whether the measures in the empirical analysis are appropriate approximations of the true shock variances and covariances, network data with information on both the income and risk profiles would be needed, or panel data to reconstruct measures of riskiness from acyclical variation in income. The data could be used to estimate the predictive value of the proxies in this study on actual shock variances and covariances. In the absence of such data, I cannot rule out that the correlations in the estimations could also be driven by correlated factors other than the variance and covariance.

3.7 Conclusion

In the absence of formal insurance products, informal risk pooling agreements are often the only means of insurance for smallholder farmers in order to smooth income and consumption. In this paper, I analyse the role of risk heterogeneity in the formation and final composition of such informal risk sharing groups.

The theoretical framework in this paper is a variant of the model by Jaramillo et al. (2015) and extends their analysis to account for the influence of the covariance between idiosyncratic shocks in addition to the individual shock variances. In the model, I derive two testable implications regarding the role of shock variances and the covariance between idiosyncratic shocks in the decision to endogenously form risk sharing groups. I find that in equilibrium, risk sharing partners will have similar shock variances and negative or zero covariances of idiosyncratic shocks. This paper is the first to consider this relationship theoretically and explicitly test for it empirically.

In the empirical analysis, I find that the insurance network in a small rural village in Tanzania is largely maintained along those equilibrium predictions. In particular, I find that similar shock variances are positively correlated with the probability of two households being risk sharing partners. This correlation remains significant even after controlling for dyad characteristics, wealth and the covariance of shocks. Furthermore, I find that households are more likely to form risk sharing agreements if their idiosyncratic shocks have negative covariance. This correlation is significant after controlling for dyad characteristics that are likely to influence risk sharing decisions like kinship and geographical distance. However, the correlation seems weaker than the one with similar shock variances. This suggests that zero or negative covariance between idiosyncratic shocks plays a less important role in forming and maintaining risk sharing agreements.

These findings provide avenues for future research. At present, the model does not account for the influence of secondary links on the equilibrium allocation of risk sharing partners. However, in reality the suitability of risk sharing partners might also depend on other risk sharing agreements a household maintains and this might influence the equilibrium allocation. Moreover, this model is static and does not allow for renegotiation of risk sharing agreements, and the predictions of the model may change if it allowed for sequential formation of risk sharing agreements and renegotiation.

Empirically, future research should build on these findings by confirming that they hold for other proxies of agents' risk profiles as discussed in section 3.6.2. Panel data with information on shocks could be used to measure shock variances and co-variances more directly. Furthermore, information on transfers between risk sharing partners would allow to disentangle the relative importance of each theoretical implication for links of different strengths.

Acknowledgements I am grateful for the comments that I have received from Catia Batista, Svetlana Chekmasova, Susana Peralta, and the seminar participants in the Nova SBE research group. I am indebted to Joachim DeWeerdt for making the data available for me and allowing me to use it in the empirical analysis. I greatly acknowledge funding for this project from the Fundaçao para a Ciencia e a Tecnologia (FCT).

Appendices

A Matching Mechanism

In this section, I use the basic intuition of many-to-many matching models to demonstrate a possible matching mechanism of risk sharing partners. This allows to illustrate matching with several bilateral matches per agent.

I assume a frictionless risk sharing market, in which each agent has perfect information over the characteristics of all N agents in the economy. Every agent ranks each of the $N-1$ agents relative to the single state. The single state refers to the situation in which an agent does not share risk and bears their own risk entirely by themself. These rankings are subjective to each agent and they are determined in accordance with the risk profile of the potential risk sharing partner and the correlation between the risk profiles of the agent and the potential partner. The risk profile of agent i depends on the agent's idiosyncratic shock $e_i \sim (0, \sigma_{e_i}^2)$ with mean zero and variance $\sigma_{e_i}^2$, and on a shock that is common to all agents $v \sim (0, \sigma_v^2)$ with mean zero and variance σ_v^2. An agent's riskiness is defined by their production technology, and I assume that the agents have no control over their own position in the ranking of other agents. This is to account for the fact that agents are unlikely to factor in the effect on their ranking when maximising profits from production. This is particularly convincing as there is no aggregate ranking over all agents in the economy. Instead, each of the N agents has an idiosyncratic ranking for each of the other $N-1$ agents relative to the single state. Any production decision can thus affect one's position in the rankings of different agents in opposite ways.

In equilibrium, risk sharing groups are obtained through simultaneous proposals from every agent to all those agents that rank better in their individual rankings than the single state. This proposal procedure may result in the least risky agent receiving up to (N-1) proposals and the riskiest ones receiving none. Any time two agents propose a link to each other the link between these two nodes forms. If only one agent proposes the link, no risk sharing agreement is formed as the receiving agent does not improve upon their single state by sharing risk with the proposing agent. Depending on their risk profile an agent can have 0 - (N-1) bilateral risk sharing partners in equilibrium. Since I assume no limit to the number of risk sharing partners, there is no need for sequential selection of partners in order to generate an efficient assignment of risk sharing partners.

I further assume that there are no fixed costs involved in entering a risk sharing agreement. Sharing risk with many partners is no more costly than sharing risk with only a few. I assume that regardless of the number of risk sharing partners, p, an agent allocates a fixed insurance amount κ_i to any risk sharing agreement. In order to guarantee that an agent can meet all their liabilities from risk sharing at any time, $\kappa_i \times p < 1$ needs to be satisfied. Total income of any agent is identical for all agents and normalised to one. The solution of the model shows that the decision whether or not to propose to an agent is independent of κ_i as κ_i does not affect the ranking. This assumption simplifies the analysis by allowing to abstract from the effect of second order links between agents on the risk sharing proposal decision. The decision whether or not to establish a risk sharing agreement is independent of any other risk sharing agreement that either node might have joined since neither the total income of a partner nor the share allocated to the agreement matter for the ranking.

B Additional Tables

Table B.1: Probit estimations

	Probability of Risk Sharing										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Δ(# Income Sources)	-0.0664*	-0.0819**			-0.0957**			-0.0821**	-0.1013**	-0.0910**	-0.0963**
	(0.037)	(0.038)			(0.042)			(0.038)	(0.041)	(0.042)	(0.043)
$\sum$(# Income Sources)		0.0354			0.0388			0.0350	0.0393	0.0396	0.0322
		(0.024)			(0.026)			(0.024)	(0.026)	(0.026)	(0.026)
Δ(Land Owning)			0.0477***	-0.0878***	-0.0836***						
			(0.016)	(0.024)	(0.027)						
$\sum$(Land Owning)				0.1308***	0.1600***					0.0915***	
				(0.017)	(0.019)					(0.027)	
Income Covariance						-0.0684	-0.0965*	-0.0341	-0.0834	-0.1777	-0.3398
						(0.053)	(0.058)	(0.054)	(0.058)	(0.110)	(0.309)
Covariance*$\sum$(Land Owning)										0.0303	
										(0.031)	
Covariance*$\sum$(Education)											0.0484
											(0.055)
Observations	6328	6328	7021	7021	6328	7027	6670	6328	6328	6328	6105
Kinship					Yes		Yes		Yes	Yes	Yes
Geo. Distance					Yes		Yes		Yes	Yes	Yes
$\sum$(Education)											Yes

Note: The outcome variable is a binary measure of insurance links that treats unilateral and reciprocal links as equivalents. Δ(# Income Sources) is the difference in the average number of income sources per adult household member between the two households in the dyad, $\sum$(# Income Sources) is the sum of the average number of income sources per adult household member. Analogously for Δ(Land Owning) and $\sum$(Land Owning), and $\sum$(Education). Column 1-5 correspond to table 3.2, column 6-9 correspond to table 3.3, and column 10-11 to table 3.4. Unclustered standard errors are reported in brackets below the coefficients.

Bibliography

Appadurai, A. (2004). The Capacity to Aspire: Culture and the Terms of Recognition. In Rao, V. and Walton, M., editors, *Culture and Public Action*, pages 59–84. Stanford University Press, Palo Alto.

Arrow, K. J. (1963). Liquidity Preference. Lecture IV. *Lecture Notes for Economics 285, The Economics of Uncertainty*, (undated):33–53.

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

Autor, D. H. (2014). Skills, Education, and the Rise of Earnings Inequality Among the "Other 99 Percent". *Science*, 344(6186).

Baker, R., Bettinger, E., Jacob, B., Marinescu, I., and Jacob, B. (2017). The Effect of Labor Market Information on Community College Students' Major Choice. *NBER Working Paper*, 23333.

Bala, V. and Goyal, S. (2000). A Noncooperative Model of Network Formation. *Econometrica*, 68(5):1181–1229.

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84(2):191–215.

Bandura, A. (1986). Social Foundations of Thought and Action: A Social Cognitive Theory. *Englewood Cliffs, NJ: Prentice-Hall.*

Batista, C. and Seither, J. (2018). Reference Points and Entrepreneurship: Evidence from a Field Experiment on Aspirations, Goal Setting, and Business Skills.

Batista, C. and Vicente, P. C. (2018). Is Mobile Money Changing Rural Africa? Evidence from a Field Experiment.

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science*, 335.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, 70(5):9–49.

Becker, G. S. (1964). *Human capital : A Theoretical and Empirical Analysis,with Special Reference to Education.* New York : Columbia University Press for National Bureau of Economic Research, New York.

Becker, G. S. and Chiswick, B. R. (1966). Education and the Distribution of Earnings. *The American Economic Review*, 56(1):358–369.

Becker, G. S. and Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, 87(6):1153–1189.

Becker, G. S. and Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, 4(3):1–39.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *Review of Economic Studies*, 81:608 – 650.

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*.

Besley, T. (2016). Aspirations and the Political Economy of Inequality. *Oxford Economic Papers*, 89(1):1 – 35.

Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, 78(1):246–263.

Bramoullé, Y. and Kranton, R. (2007). Risk Sharing across Communities. *The American Economic Review*, 97(2):70 – 74.

Breda, T., Grenet, J., Monnet, M., and Effenterre, C. V. (2018). Can Female Role Models Reduce the Gender Gap in Science? Evidence from Classroom Interventions in French High Schools. *PSE Working Papers*, 2018 - 06.

Campbell, M. K., DeVellis, B. M., Strecher, V. J., Ammerman, A. S., DeVellis, R. F., and Sandler, R. S. (1994). Improving dietary behavior: the effectiveness of tailored messages in primary care settings. *American Journal of Public Health*, 84(5):783–7.

Carneiro, P. and Heckman, J. J. (2003). Human Capital Policy. In Heckman, J. J., Krueger, A. B., and Friedman, B. M., editors, *Inequality in America: What Role for Human Capital Policies?* MIT Press, Cambridge, MA.

Chiswick, B. R. (1969). Minimum Schooling Legislation and the Cross-Sectional Distribution of Income. *The Economic Journal*, 79(315):495–507.

Cunha, F. and Heckman, J. J. (2006). Investing in our Young People.

De Weerdt, J. (2004). Risk-Sharing and Endogenous Network Formation. In Dercon, S., editor, *Insurance against Poverty*, chapter 10, pages 197–216. Oxford University Press.

De Weerdt, J. and Dercon, S. (2002). Risk-Sharing and Endogenous Network Formation. *Foreign Affairs*, pages 197–216.

Dee, T. S. (2005). A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review*, 95(2):158–165.

DellaVigna, S. and La Ferrara, E. (2015). Economic and Social Impacts of the Media. In *Handbook of Media Economics*, chapter 19, pages 724 – 766. Elsevier, 1b edition.

Dercon, S. and Christiaensen, L. (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of Development Economics*, 96(2):159–173.

Education Endowment Fund (2018). The Attainment Gap 2017.

Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350.

Fafchamps, M. and Lund, S. (2003). Risk-sharing networks in rural Philippines. *Journal of Development Economics*, 71(2):261–287.

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

Grimard, F. (1997). Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire. *Journal of Development Economics*, 53(2):391–422.

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.

Hanushek, E. A. and Woessmann, L. (2011). How Much Do Educational Outcomes Matter in OECD Countries? *Economic Policy*.

Hastings, J., Neilson, C. A., and Zimmerman, S. D. (2015). The Effects of Earnings Disclosure on College Enrollment Decisions. *NBER Working Paper*, 21300.

Heckman, J. J. (2000). Policies to Foster Human Capital. *Research in Economics*, 54:3–56.

Hulleman, C. S. and Harackiewicz, J. M. (2009). Promoting Interest and Performance in High School Science Classes. *Science*, 326:1410 – 1412.

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., and Puranen, B., editors (2014). *World Values Survey: Round Five - Country-Pooled Datafile Version: www.worldvaluessurvey.org/WVSDocumentationWV5.jsp*. JD Systems Institute, Madrid.

J. Clyde Mitchell (ed.) (1969). *Social networks in urban situations: analyses of personal relationships in central African towns*. Manchester University Press for the Institute for Social Research, Manchester.

Jackson, M. O. and Wolinsky, A. (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71(1):44–74.

Jaramillo, F., Kempf, H., and Moizeau, F. (2015). Heterogeneity and the Formation of Risk-Sharing Coalitions. *Journal of Development Economics*.

Jensen, R. (2010). The (Perceived) Returns to Education and The Demand for Schooling. *Quarterly Journal of Economics*.

Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural Decisions after Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics*, 129(2):597–652.

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

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

Lergetporer, P., Werner, K., and Woessmann, L. (2018). Does Ignorance of Economic Returns and Costs Explain the Educational Aspiration Gap? Evidence from Representative Survey Experiments. *CESifo Working Paper*, 7000.

Lusher, L., Campbell, D., and Carrell, S. (2018). TAs like me: Racial interactions between graduate teaching assistants and undergraduates. *Journal of Public Economics*, 159:203 – 224.

Marcus, B. H., Emmons, K. M., Simkin-Silverman, L. R., Linnan, L. A., Taylor, E. R., Bock, B. C., Roberts, M. B., Rossi, J. R., and Abrams, D. B. (1998). Evaluation of Motivationally Tailored vs. Standard Self-Help Physical Activity Interventions at the Workplace. *American Journal of Health Promotion*, 12(4).

Meece, J. L., Wigfield, A., and Eccles, J. S. (1990). Predictors of Math Anxiety and Its Influence on Young Adolescents' Course Enrollment Intentions and Performance in Mathematics. *Journal of Educational Psychology*, 82(1):60–70.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4):281 – 302.

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

OECD (2017). Understanding the Socioeconomic Divide in Europe. *COPE (Centre for Opportunity and Equality)*.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica*, 32(1):122–136.

Prochaska, J. O., DiClemente, C. C., Velicer, W. F., and Rossi, J. S. (1993). Standardized, individualized, interactive, and personalized self help program for smoking cessation. *Health Psychology*.

Raub, W. and Weesie, J. (1990). Reputation and Efficiency in Social Interactions: An Example of Network Effects. *American Journal of Sociology*, 96(3):626–654.

Ray, D. (2006). Aspirations, Poverty and Economic Change. In Benabou, R., Banerjee, A., and Mookherjee, D., editors, *What Have We Learnt About Poverty*. Oxford University Press, Oxford.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*, 127(604):2153–2186.

Rosenbaum, P. R. (2002). Covariance Adjustment in Randomized Experiments and Observational Studies. *Statistical Science*, 17(3):286–327.

Shiffman, S., Paty, J. A., Rohay, J. M., Di Marino, M. E., and Gitchell, J. (2000). The Efficacy of Computer-Tailored Smoking Cessation Material as a Supplement to Nicotine Polacrilex Gum Therapy. *Archives of Internal Medicine*, 160(11):1675.

Social Mobility Report (2017). Social Mobility, the Class Pay Gap and Intergenerational Worklessness: New Insights from The Labour Force Survey. Technical report, Social Mobility Commission, London.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 35:4 – 28.

Steele, C. M. and Aronson, J. (1995). Stereotype Threat and the Intellectual Test Performance of African Americans. *Journal of Personality and Social Psychology*, 69(5):797 – 811.

Townsend, R. M. (1994). Risk and Insurance in Village India. *Econometrica*, 62(3):539–591.

Udry, C. and Conley, T. (2004). Social Networks in Ghana. *Yale University Economic Growth Center Discussion Paper No. 888.*

Walton, G. M. (2014). The New Science of Wise Psychological Interventions. *Current Directions in Psychological Science*, 23(1):73 – 82.

Wigfield, A. and Eccles, J. S. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25:68–81.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., and Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, 89(3):451–469.

Young, A. (2017). Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *Working paper*.