I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17261

The Impact of Role Models on Youths' Aspirations, Gender Attitudes and Education in Somalia

Elijah Kipchumba
Catherine Porter
Danila Serra
Munshi Sulaiman

AUGUST 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17261

The Impact of Role Models on Youths' Aspirations, Gender Attitudes and Education in Somalia

Elijah Kipchumba
Trinity College Dublin

Catherine Porter
Lancaster University

Danila Serra
Texas A&M University and IZA

Munshi Sulaiman
BRAC University

AUGUST 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Impact of Role Models on Youths' Aspirations, Gender Attitudes and Education in Somalia*

We evaluate the impact of a role model intervention on the gender attitudes, college aspirations and education outcomes of youths in Somalia. In 2018, we randomly selected elementary schools to receive a visit from a college student. Within each treatment school, we selected four grades, two to receive a visit from a female college student and two from a male college student. The "role models" gave unscripted talks about their personal study journeys, including challenges and strategies to overcome setbacks. Six months after the intervention we found a significant and large impact of (only) female role models on boys' and girls' attitudes toward gender equality but no impact on college aspirations. Data collected two and four years later from the cohorts graduating primary school produce smaller and non-significant treatment effects on the survey outcomes, but positive impacts on enrollment in high school and a lower probability of early marriage as reported by teachers.

JEL Classification: J16, O12, I25, C93

Keywords: role models, education, gender, aspirations, field experiment, Somalia

Corresponding author:
Danila Serra
Texas A&M University
400 Bizzell St
College Station, TX 77843
USA

E-mail: dserra@tamu.edu

* We thank Save the Children for allowing us to combine our role model intervention with their ongoing NORAD school program. The first wave of data collection received ethical clearance from the Institutional Review Board (IRB) at Southern Methodist University; the second wave from the IRB at Texas A&M University, and the third wave from the IRBs at Texas A&M University and Lancaster University. We acknowledge financial support from the Spencer Foundation through grant number 202200248. We thank seminar participants at many institutions for useful comments and discussions. We also thank participants at the 2020 NBER Gender in the Economy Summer Institute, the 2021 and 2024 CSAE conferences at Oxford University, the 2021 NEUDC conference, the 2023 Midwestern International Development conference and the 2023 Advances in Field Experiment conference.

1 Introduction

Achieving an education in the poorest countries is extremely challenging. A huge barrier to education is lack of infrastructure, including school buildings, well-functioning classrooms and sanitation facilities (Burde and Linden, 2013; Duflo, 2001; Kazianga et al., 2013). The unavailability of qualified teachers is an additional supply-side constraint to education provision (Mulkeen et al., 2007). Importantly, several studies have shown that the demand for education may also play an important role in the observed low take up of educational services and high dropout rates. This may be due to a failure to recognize the returns of education (Jensen, 2010; Nguyen, 2008), or to low aspirations of parents and/or children (Appadurai, 2004; Genicot and Ray, 2017; Ray, 2006). Moreover, the prevailing gender norms within traditional communities may dictate that it is less important for women than for men to get an education. This, in turn, leads to girls facing additional and bigger obstacles in their quest for an education, which results in their inability to access higher-return jobs in adulthood (Dhar et al., 2019; Heath and Jayachandran, 2017; Jayachandran, 2021).

We investigate whether exposure to role models could significantly impact youths' gender attitudes, as well as their aspirations and ability to enroll in secondary school and, ultimately, pursue higher education. Following Morgenroth et al. (2015), we define role models as individuals who can "influence role aspirants' achievements, motivation, and goals." We conduct a field experiment in Somalia, a very low income and fragile context. Literacy rates are among the lowest in the world; very few children are enrolled in school and even fewer complete their education (UNICEF, 2017). For example, net enrollment rate among secondary-school aged girls is 15% and 19% among boys (World Bank, 2019). The scarcity of college-educated individuals means children rarely encounter role models who could inspire them to pursue higher education, particularly female role models. Interacting with women in higher education with non-stereotypical career aspirations could also reshape children's views on gender roles and equality.

The study was rolled out in collaboration with Save the Children International, which was already involved in the implementation of an intervention aimed at improving infrastructure and teaching in 46 primary schools – primarily serving grade 1 to 8 students – in two regions. We randomly assigned schools to either be visited in person by role models, who were current college students sharing similar backgrounds as the target youths, or to act as control schools. The role model visits took place between April and May 2018. Within each treatment school, we randomly selected two grades to be visited by a male role model and two grades to be visited by a female role model. When designing the study, we hypothesized that exposure to role models would influence gender attitudes and aspirations, leading to improved education

outcomes. We predicted a greater aspiration effect for female students exposed to female role models, and a stronger gender attitude effect for both male and female students exposed to female role models.

We assess the impact of the role model intervention using survey data and exam results, after conducting three distinct waves of data collection. The initial wave, or First Endline, took place about six months after the intervention and involved students enrolled in grades 4 to 7 at the time of the role model visits (we refer to these initial grades as *baseline grades*). We administered a short survey inside the schools and elicited students' desired levels of education and plans to attend high school and college, as well as a set of questions aimed at recording attitudes toward men's and women's education and labor market participation. These measures of educational aspirations and gender attitudes are our primary survey-generated outcomes of interest throughout the different waves of data collection. The Second Endline data collection took place two years after the role model visits, in Spring 2020, at the onset of the COVID-19 pandemic, and to expedite fieldwork, only involved the oldest student cohort who were in grade 6 at the time of the role model visits, and who were about to graduate from 8th grade. We conducted a third wave of follow-up data collection in the spring and summer of 2022, four years after the intervention. This involved students who were in baseline grade 4 and were about to graduate from 8th grade at the time of the third endline survey. Additionally, we obtained the standardized grade 8 exit exam performance for all surveyed students from the graduating classes of 2020 and 2022 (baseline grade 6 and grade 4), as this exam is required for elementary school graduation.

The extremely fragile environment of our study made tracking students after they left elementary school infeasible due to instability, widespread displacement, and the lack of mobile phones. To gather information on secondary school enrollment, as well as marriage and fertility outcomes, we surveyed corresponding class teachers during the third wave of data collection in the summer of 2022. We asked them for information regarding each student who had graduated in 2020 and 2022. Specifically, we asked whether each student was currently enrolled in high school (and the name of the high school) and whether s/he was married and had children.

Our impact evaluation therefore relies on three primary sources of data. First, we have survey generated measures of education aspirations and gender attitudes. We are able to assess the impact of the intervention on these outcomes for students in baseline grades 4 to 7 six months after the intervention, for students in baseline grade 6 two years after the intervention, and for students in baseline grade 4 four years after intervention. Second, we have standardized grade 8 exit exam performance data for all surveyed baseline grade 4 and 6 students, enabling us to test the intervention's impact on academic performance. Third, the

information from class teachers on high school enrollment, marriage, and fertility for these two student cohorts allows us to evaluate whether the intervention influenced post-primary educational choices and fertility outcomes, such as delayed marriage and childbearing, which we can interpret as "revealed" gender attitudes.

The implementation of this study was particularly challenging for a number or reasons. First, we did not have access to official student rosters at the school-level as these did not exist or were not kept up to date when the study began. This means that we are unable to identify and follow the students who were enrolled and present when the role models visited the schools back in 2018. It also implies that we cannot track students over time, since we only have limited (not uniquely) identifying information from the surveys, and we cannot properly assess attrition. For the purpose of the impact evaluation, we assume student treatment status based on students' *current grade* (at the time of the follow-up data collection), which lead us to assign to each student the corresponding *baseline grade* in spring 2018, when the intervention took place. However, this assumes that students stayed enrolled at the same school and did not experience any grade retention. It follows that our data allows us to only estimate the intent-to-treat effects of the role model intervention, while acknowledging that students' retention and school transfers are likely to bias our estimates towards finding null impacts.

Six months after the intervention, we find evidence of large and significant positive impacts of the female role models' visits on both boys and girls' attitudes toward gender equality. The effects are large (and equal for boys and girls): 0.37 standard deviations from the means observed in the control schools, and are able to close the observed gap in gender attitudes. In contrast, we find no evidence of a significant impact of the role models on educational aspirations; virtually all students expect to enroll in secondary schools and about 70 percent aspire to go to college, with no significant differences by gender and by treatment status.

When surveying baseline grade 6 and 4 students, two and four years after the intervention (respectively), we see that the impact of female role models on gender attitudes is still positive but much smaller and not statistically significant. However, the number of schools in our sample declines over time, due to the fact that not all schools in our sample served grades 6 to 8. The lower number of schools, together with data availability only for one student cohort per year lowers our statistical power. While the impact of role models on college aspirations is still null in the 2020 follow-up data (for baseline grade 6 students), female role models have a positive and marginally significant impact on the college aspirations of girls in the 2022 follow-up data (baseline grade 4).

Our analysis of education outcomes for the baseline grades 4 and 6 students indicates that female role models had a positive and marginally significant impact on girls' performance in

the exit exam, and a positive impact on both boys' and girls' likelihood to enroll in secondary school. We also see a negative impact on marriage and fertility outcomes of boys and girls, of the same magnitude but statistically significant only for boys. These findings suggest that the role model intervention - and in particular exposure to female role models - not only altered gender attitudes in the short term, but it also led to improved educational outcomes for girls and possibly impacted youths' life trajectories.

This study contributes to a growing literature investigating role model effects by testing the impact of movies and TV shows on education outcomes (Riley, 2024), women's empowerment (Jensen and Oster, 2009; La Ferrara et al., 2012), entrepreneurial aspirations (Bernard et al., 2014; Bjorvatn et al., 2019) and teens' behavior (Kearney and Levine, 2015). Other related studies have examined the impact that female teachers may have on women's performance in math (Muralidharan and Sheth, 2016) and their field of study in higher education (Carrell et al., 2010; Lim and Meer, 2019). Finally, Beaman et al. (2012) and Porter and Serra (2020) have shown that exposing young women to successful and inspiring women may impact their aspirations and educational choices.[1] Similarly to Porter and Serra (2020), our experimental design relies on the direct interaction between the target sample (primary school students) and real-life role models who come from a similar background and have experienced similar challenges in life. Our work is also related to the study by Nguyen (2008), which exposed pupils in schools in Madagascar to educated individuals from the same school districts.[2] Contrary to Porter and Serra (2020) and Nguyen (2008), we randomized the gender of the role model students were exposed to. This way, we are able to assess differential impacts of male and female role models by the gender of the target students.[3]

Our study also contributes to the literature on gender norms and women's participation in the labor market. Existing investigations have highlighted the impact of family and societal characteristics on female labor force participation (Fernández et al., 2004; Olivetti et al., 2020). Studies that focus specifically on the formation of and possible change in individual attitudes toward gender equality are scarce. A field experiment conducted in the military in Norway (Dahl et al., 2021) shows that being assigned to live and work with female recruits changed men's gender attitudes. In India, an intervention targeting secondary school students and focused on classroom discussions on gender roles and gender equality significantly impacted adolescents' gender attitudes and behaviors, with the effect being larger for boys

[1] See Serra (2022) for a review of role model interventions in low-income countries.

[2] Nguyen (2008) employed a total of 72 role models chosen by a local committee consisting of the school district head, a local NGO leader, and community leaders.

[3] A recent study conducted by Patnaik et al. (2023) in a US university indicates that gender-matching is important in role model interventions. In their study of college students, they found that men were more responsive to male role models and women to female role models.

(Dhar et al., 2022). The program in India was much more intensive with a two-and-a-half year weekly treatment, whereas the role model visits we implemented were relatively short and cheap - costing approximately $53 per role model per school, with each role model visiting two classrooms.[4]

In the next section we outline the Somalia context, then describe the role model intervention and our empirical strategy. Section 4 presents our main findings and Section 5 concludes.

2 Context

Ever since the devastating civil war of 1988-91, the vast majority of school age children in Somalia have had very limited access to basic services in general and basic education in particular. Conflict mixed with recurring drought, flood and climate change and food insecurity, has been daily life for communities and their children. Over 1.5 million people have been living as Internally Displaced Persons (IDPs) for decades across the three regions of Somaliland, Puntland and South Central Somalia (SCS). The government of Somalia faces serious challenges in governing and financing public education due to a lack of financial, human and managerial resources. Parents are required to pay fees to enrol their kids in public schools, which contributes to the marginalization of children from poor and rural families. Moreover, capacity constraints have caused most education services, including examinations and certification, to be provided by private Education Umbrellas rather than the public education system, resulting in lack of regulation and standardization.

Unsurprisingly, literacy rates in Somalia are among the lowest in the world. The World Bank (2019) estimated that in 2018 only half of the adult population were able to read and write. Although primary education (grades 1 to 8) has been mandatory since 1975, only 24% (22% girls versus 25% boys) percent of children of primary school age (6-13) are enrolled in primary schools (Ministry of Education, 2022). Girls are equally likely to be enrolled in primary school, however a significant gap emerges at secondary level (World Bank, 2019). Among children of secondary school age (14-17), according to the most recent data, net secondary school enrollment among girls is just 15% and only 19% among boys. Gender norms are severely biased against women (UNDP, 2019). The 2022 UNDP Gender Inequality index, a composite score that assesses the status of women's reproductive health, empowerment and labor market participation, ranked Somalia third last in the entire global

[4]This was the cost of lunch and travel for each role model visiting a school. It also covers the per person cost of implementing a coaching day with all the role models. This implies that the cost of conducting the intervention in a school, with two role models - one man and one woman - each visiting two classes, was about $106.

country distribution, due to extreme gender inequality (UNDP, 2023). Entrenched traditional views assign women and girls secondary status in Somali society. This perpetuates narrow gender-based roles and inequalities. Participation of women and girls in decision-making spheres is limited. Female teachers are unlikely to serve as role models for young women, as only 14 percent of teachers in primary and 4 percent in secondary schools are women (Ministry of Education, Culture and Higher Education, 2017). Our paper is one of only a handful of field experiments conducted in Somalia, due to the difficulties of conducting research in the country.[5] A recent review of economics research on African countries in the past 20 years by Porteous (2020) shows that over 40 percent of published papers are about 5 countries accounting for less than 20 percent of the African population.[6] Somalia is one of the "scarce 7" countries,[7] which have equivalent population as the "frequent 5" but are rarely studied and account for less than 5 percent of economics publications. This makes our data collection efforts in Somalia particularly relevant.

3 The Role Model Intervention

We conducted the role model intervention in partnership with Save the Children International (SCI). SCI was already implementing a set of education programs in 46 primary schools in two Somali regions, Puntland and South-Central, in collaboration with the Ministry of Education. The programs focused on increasing access to education for marginalized children (girls, boys, pastoralists, minority groups and children with mild disabilities) in targeted areas.[8] They included upgrades and rehabilitation of classrooms and school facilities with the aim of creating safe and child-friendly learning environments. SCI also assisted schools in developing and implementing School Improvement Plans, training teachers in basic pedagogy and code of conduct and enhancing school governance, and student literacy boost among others (Save the Children, 2018).

We randomly selected 23 of the 46 primary schools[9] in which SCI was active to receive a role model intervention using pairwise randomization.[10] Given the small number of SCI schools (46), we used pairwise randomization to select the schools that would receive the

[5]Other notable recently published or working papers are: Abdullahi et al. (2023), Brar et al. (2023) and La Ferrara et al. (2021).

[6]The most commonly studied African are Ghana, Kenya, Malawi, South Africa and Uganda.

[7]The other countries are Angola, Chad, D.R. Congo, Guinea, South Sudan and Sudan.

[8]The target regions and schools were chosen based on three criteria: a) the needs of children and the extent to which their rights are violated in the region; b) the absence or limited interventions by other NGOs or agencies in addressing the plights and rights of children in a holistic and integrated manner; c) operational and technical capacity of SCI and its partners to successfully implement similar intervention in the region.

[9]Figure A1 in the Appendix provides a map depicting the locations of the treatment and control schools.

[10]We originally had 47 schools, but one school was excluded as it was an all girls school.

role model intervention. We first stratified schools by region and sorted based on number of grades offered by the schools and total enrolled students. We then conducted pairwise matching (Imbens and Rubin, 2015). Within each pair, we randomly selected one school to be treated. Abadie and Imbens (2011) show that matched-pairs randomization (and stratified randomization) can increase power in small samples. Simulation evidence presented in Bruhn and McKenzie (2009) supports these findings, though they note for large samples there is little gain from different methods of randomization over a pure single draw. Imai et al. (2009) also derive properties of matched-pair cluster randomization estimators and demonstrate large efficiency gains relative to pure simple cluster randomization. The randomization was successful with no significant pre-intervention differences at the school level in number of classrooms, number of students enrolled (and boy/girl ratio), number of available grades, school fees, length of time operational, girl friendly space availability, number of female teachers (see the top panel of Table I, which we discuss in Section 4).[11]

Each of the treated schools received a visit from one male college student and one female college student (our role models) on the *same day*. Each role model visited 2 grades, for a total of 4 treated grades per treated school. Since schools had differing number of grades, we decided which grades would receive the role model visits based on the available grades, as follows. In schools with 8 grades (46 percent of the T schools), we treated grades 2, 4, 6, 8. In schools with 7 grades (17 percent of T schools), we treated grades 2, 4, 6 and 7. In schools with 6 grades (9 percent of T schools), we treated grades 2, 4, 5 and 6. In schools with 5 grades (13 percent of T schools), we treated grades 2, 3, 4 and 5. Finally, in schools with four grades (17 percent of the treated schools) we treated all grades. In each treatment schools, we randomly selected two of the treated grades to be visited by a male role model and two to be visited by a female role model.

3.1 Implementation

The most critical aspect of the field experiment was the identification of appropriate role models, who were college students with similar backgrounds and who had experienced similar challenges as the primary school children being visited. To recruit them, we worked with head teachers from the SCI schools to identify former students who were currently attending universities in Somalia. From the long list of potential role models, we worked with the SCI program staff to short-list those college students whose life story was especially inspiring, i.e., students who had achieved academic excellence and managed to pursue university education

[11] All the schools in this study were supported by SCI through the Norad program. The majority (37) of the schools were located in urban areas. Of these schools, 14 mostly catered children from displacement camps, which are located in the periphery of urban centers.

despite having faced personal difficulties. These role models were then invited by SCI program staff to a briefing session in March 2018 to inform them about the role model intervention and asked them for their willingness to collaborate with us on the project. A total of 15 role models (9 female and 6 male) agreed to participate in the intervention. The most common fields of study being medicine and economics. Each role model visited between one and four schools.

Before the school visits, we invited the role models to a day-long coaching session where we informed them of the structure of the school visits, and we assisted them in preparing a 30-40 minute speech targeted at for primary school students. The role models were invited to include the following items in their speeches: 1) introduction (names, how old they are, which schools they went to including the current university); 2) what challenges they experienced while in school, how did they overcome them; 3) whether they experienced failures and what they would do differently if they could go back in time; 4) future plans and job prospects.[12] At the end of the coaching session, the role models chose the schools that they would visit (out of those in the treatment group) based on proximity to their place of origin. Some of the role models were alumni of the sampled schools; even those who were not grew up in nearby regions and faced similar challenges as the students they went to speak to.

The intervention took place between April and May 2018. Each role model visited one school per day and gave two speeches, mostly in the morning hours. In all occasions, two role models visited the school. This ensured that they delivered their speeches to all the four classes selected to receive the intervention. Each role models was provided with 50 USD per school visit to cover lunch and transport costs.

In all treated schools, the SCI Norad program staff sought permission for the role model visits from the school administration at least one week prior to the schedule visit day. The speeches lasted between 40 and 60 minutes and were delivered in the selected classrooms in the presence of the class teacher. The visits ended with 5 to 10 minutes of students' questions and role models' answers, facilitated by the class teacher. Since the speeches were done during normal class time, there was no contamination of students in control classes within treated schools.

3.2 Data Collection and Outcome Variables

Due to budget and logistical constraints, we were not able to collect student-level data before the intervention took place. Therefore, at baseline, we only have school-level data provided

[12]Each role model was accorded a half hour time to practice their speech in front of fellow role models and their coaches (SCI Norad program staff). We had a separate coaching session for each of the four regions that the program operated in.

by Save the Children. We discuss these data in Section 4.1.

We implemented three separate waves of endline data collection. First, in October and November 2018, we administered a short survey in the sampled schools, involving students in baseline grades 4 to 7. This constitutes our First Endline data collection. We implemented the Second Endline data collection in April and May 2020, when we surveyed students in baseline grade 6, who where in grade 8 at the time of the Second Endline and about to graduate from primary school. For these students, we were also able to obtain their performance in the standardized exit exam that all grade 8 students have to take to progress to secondary school. Finally, between May and August 2022, we implemented the Third Endline data collection. In May 2022, we surveyed the students in baseline grade 4, who were in grade 8 at the time of the Third Endline and about to graduate from primary school. For these students, we also obtained performance in the standardized exit exam. In addition, in August 2022, we implemented a short survey with the grade 8 teacher at each school (at that time). The survey aimed to record information about high school enrollment and family outcomes (marriage and children) for each of the graduating students in 2022 and 2020, i.e., for the students would have been in grades 4 and 6 at the time of the intervention.

We focus on students from baseline grade 4 and grade 6 in the long run because, following our randomization, role models visited grades 2, 4, 6, and 8 in all treatment schools with eight grades in Spring 2018 (about 50 percent of the sample).[13] This implies that our impact evaluation can rely on a clean comparison of outcomes between (*always treated*) students in the treatment schools and students enrolled in the same grades (and taking the same exit exam) in control schools.[14]

3.2.1 First Endline: October-November 2018

The First Endline data collection took place six months after the intervention, between October and November 2018. We collected survey data in 41 control and treated schools.[15] The data collection relied on a self-administered short questionnaire (two pages only), which recorded information on students' family background, aspirations concerning education and future occupation and answers to a set of questions aimed at measuring gender attitudes. An

[13]Grades 2 and 4 were treated in all schools, since all sampled schools had at least 4 grades. Grade 6 was treated in all schools with at least 6 grades, and grade 8 was treated in all schools with 8 grades.

[14]The students in baseline grade 8 graduated before the First Endline. We had originally planned to survey them just before their graduation. However, due to delays in the implementation of the intervention, we had to conduct the First Endline data collection at the beginning of the following academic year, rather than at the end of the intervention semester.

[15]Because the survey was self-administered we received mostly blank forms from three schools. Two other schools were part of another education technology evaluation within the wider program, thus were excluded from the 2018 data collection.

enumerator went through the questionnaire in class, with the teacher present, explaining the meaning of each question. The students then independently answered the questions. Due to concerns regarding younger students' ability to read and comprehend the questions, we decided to survey only students enrolled in grades 5 to 8 at the time of the data collection (i.e., baseline grades 4 to 7), conditional on being at school that day. This resulted in a sample of 1,941 students. Recall that baseline grades 4 and 6 were always treated in treatment schools, whereas baseline grades 5 and 7 were treated only in schools with 5 grades and in schools with 7 grades, respectively. This means that our sample include untreated students in control schools, students in treatment classes in treatment schools, and students in untreated classes also in treatment schools. Since the latter group may have still been affected by the role model visits through their peers or teachers, we refer to them as students in *spillover grades*.[16]

As discussed in section 2, gender norms in Somalia are severely biased against women. One of our objectives is to test whether exposing students to role models – and especially visits from young women who were pursuing a college degree – could affect girls' and boys' attitudes toward gender equality in education and job participation. To measure gender attitudes, we asked the following set of agree/disagree questions:

1. "More encouragement in a family should be given to sons than daughters to go to college."

2. "It is more important for boys than girls to do well in school."

3. "Boys are better leaders than girls."

4. "Girls should be more concerned with becoming good wives and mothers than desiring a professional or business career."[17]

We average individual answers to the four questions into an index of attitudes toward gender equality, which is measured on a scale from 1 to 4, where a higher number indicates attitudes more strongly in favor of gender equality.

Our other primary outcome variable is students' educational aspirations. In the survey, we asked students whether they intended to enroll in secondary school (and the name of the

[16]There are 559 such students for which we have survey data from the First Endline. Our primary results are unaffected by the exclusion of these students from the analysis.

[17]In the First Endline, we also included a fifth statement: "girls are as smart as boys." However, this question had a positive item-rest correlation (when negative was expected) and lower item test-retest reliability. We therefore dropped it from the index, which resulted in Chronbach's alpha for the scale of 0.66 (previously 0.61). The Second and Third Endline survey data only include the four components of the gender attitude index.

school they plan to attend). We also ask them to report the highest level of education that they would like to complete "if nothing could stop them." The answers from this question generate an indicator of college aspirations, which is equal to 1 if the student's highest level of education is a college degree (or higher), and 0 otherwise.

While we did not pre-register and pre-specify our two primary survey outcomes prior to the First Endline data collection (in 2018), we consistently employ the same survey questions, and generate the same outcome variables, in subsequent data collection waves.

3.2.2 Second Endline: April and May 2022

We implemented the Second Endline in April and May 2020, i.e., about two years after the intervention. Due to time pressures, we surveyed the students who were enrolled in grade 8 (and would have been in baseline grade 6 at the time of the intervention). At the time of the data collection, these students were about to graduate from primary school. We have data from 32 of the original 46 schools. These are the 21 schools that had all the 8 grades at time of intervention plus an additional 11 schools that had grown to include all the 8 grades at the time of the Second Endline. These 32 schools are our target set of schools also in subsequent data collection efforts (Third Endline). Half of them are control schools, and half are treatment schools. In exactly 8 of the 16 treatment schools, baseline grade 6 had been randomly assigned to the male role model visit, and in the remaining 8, grade 6 had been assigned to the female role model intervention in Spring 2018.

We have data from 829 students who were enrolled in grade 8 in the 32 schools and were in class the day of the data collection. Importantly, in 2020, for each surveyed student in all the 32 schools, we obtained performance in a standardized exam that all grade 8 students need to take at the end of the school year to be able to graduate from primary school and progress to secondary school.[18] In order to match survey data and exam data, we had the assistance of the teacher who had taught the grade 8 class in 2020. We presented the teacher with the list of surveyed students and, with the teacher help, we identified the corresponding student and retrieved his/her exit exam grade from the school records.[19] In this matching process, we infer that students whose exit exam results were missing (nearly 30%) as having

[18]The country-wide exams are coordinated by the Ministry of Education. Exams are taken across seven subjects including Islamic studies, Arabic, Somali, Social Science, Science, English and Maths. Scores from each individual subject are averaged to form a candidate's final score. A candidate is deemed to have passed upon attaining 50% in the final score. In the 2020 exams, results showed that the average national score was 59% but up to 90% of the students achieved the pass mark with hardly any gender gap (UNESCO, 2022).

[19]Given the lack of student identifiers, and the similarity in student names, it would have been otherwise impossible for us to obtain all exam records and try to do the matching ourselves. We decided to instead start with the list of surveyed students and retrieve results only for these students we had surveyed, with the help of the grade 8 teacher.

not sat for the exam.

Therefore, the Second Endline generates two sets of outcomes for students who were in grade 8 at the time of the data collection (2020), and who would have therefore been in baseline grade 6 – hence always treated in treatment schools – at the time of the role model visits. The first set of outcomes comprises our survey-generated variables of interest, i.e., gender attitudes and college aspirations collected from students in all 32 available schools. The second outcome is the grade obtained in the standardized exit exam also collected from all the 32 schools.

3.2.3 Third Endline: May-September 2022

The Third Endline took place between May and September 2022 and involved a survey of students who were in grade 8 in May 2022 and who would have therefore been in baseline grade 4 at the time of the role model visits. We targeted all the 32 schools with 8 grades in 2022. These were the same schools surveyed in the Second Endline. Of them, 16 are control schools and 16 are treatment schools. However, the within-school randomization of the gender of the role model did not yield an equal number of grade 4 classes assigned to male and female role models at the time of the intervention (2018). Purely by chance, in this case, 12 out of the 16 treated schools (with 8 grades at the time of our Second Endline) had baseline grade 4 assigned to a female role model visit, and only 4 schools had baseline grade 4 assigned to a male role model visit. [20] An additional unforeseen problem is that the 4 schools whose baseline grade 4 students were assigned to the male role model intervention only returned a total of 35 student surveys, making it impossible for us to evaluate the impact of the male role model intervention on outcomes measured in 2022.[21] We therefore have no choice but to drop these 4 schools from the analysis, leading to a working sample of 28 schools which were either assigned to the control group (16) or were in the treatment group and had their grade 4 students assigned to female role models in Spring 2018. From these schools, we have survey data from 575 students enrolled in grade 8 at the time of the Third Endline.

The survey took place shortly before the students had to take the mandatory standardized exit exam. We returned to the schools in August of 2022 to retrieve data on student performance in the exit exam. We followed the same procedure as for the Second Endline, i.e., we matched exam and student data for all surveyed students with the help of the grade 8 class teacher. Additionally, as part of the Third Endline, we surveyed the grade 8 class

[20] In the original randomization, an additional five grade 4 classes were assigned to male role models and a further two grade 4 classes were assigned to female role models. However, they were in schools with fewer than 8 grades; hence, we are unable to trace the corresponding students.

[21] We have 4, 6, 7 and 18 surveys returned from each of the 4 schools, respectively.

teacher in each school with the aim of gathering information about each of the students in our sample who graduated in 2020 (baseline grade 6) and in 2022 (baseline grade 4). Specifically, for each surveyed student in our sample, we asked the corresponding grade 8 class teacher three questions: 1) whether the student enrolled in high school; 2) whether he or she was married, and 3) whether he or she had children. The teachers were able to provide this information unless they had not taught the corresponding grade 8 class, for instance because they had recently transferred from a different school. When following up the students who graduated in 2020 (baseline grade 6), we encountered this problem in four treatment schools (two assigned to a female role model and two assigned to a male role model) and one control school, leading to a working sample (with no missing data) of 27 schools. When following up the 2022 graduating class (baseline grade 4), we encountered this problem in only one control school, reducing the working sample for this student cohort from 28 to 27 schools.

As a result of the Third Endline, we therefore have three sets of data for the 2022 graduating cohort *(baseline grade 4)*: 1) survey data collected just before graduation from grade 8; 2) performance in the standardized exit exam; 3) enrollment in high school, as reported by the grade 8 teacher; 4) marital status and fertility outcomes, as reported by the grade 8 teacher. These data are available for the surveyed students in 27 schools,[22] of which 15 control, and 12 treatment schools (all assigned to female role models). In addition, the Third Endline also generated teacher data on secondary school enrollment, marriage and fertility for the 2020 graduating cohort *(baseline grade 6)*, for students who graduated from 27 schools,[23] of which 15 control and 12 treatment schools, 6 assigned to male role models and 6 to female role models.

3.3 Ethical Considerations and Field Challenges

This study was subject to ethical approval from the Ethics Review Boards at Southern Methodist University (intervention and First Endline), Texas A&M University (Second and Third Endlines) and Lancaster University Third Endline). As part of the wider Norad Framework evaluation, administrative approval was obtained from Ministry of Education in Puntland State and South-Central Somalia State of Somalia. Data collection in that program was supervised by local Ministry of Education administration staff, respective school

[22] As explained in the text, this is a subset of 28 schools covered in the 2022 student survey after excluding (from the original 32 schools) the four schools that were allocated male role models in grade 4 and for which we received very few survey responses. We further exclude one control school whose grade 8 teacher had changed and could therefore not answer questions about the graduated students.

[23] This is a subset of the 32 schools covered in the 2020 student survey. We exclude 5 schools where the grade 8 teacher had changed between 2020 (student graduation year) and the time of the survey, leading to missing data on secondary school enrollment and marital status.

administrators, and SCI Norad program staff.

All data collectors were trained on SCI's Child Safeguarding Policy and were required to append their signatures indicating understanding of the policy as well as commitment to abide by it during data collection. Informed consent and assent was obtained in line with SCI precedents. One week before the data collection, as part of the mobilization and preparation efforts, the headteachers of respective schools were mandated to inform parents about the study and the data collection (its purpose, what information would be collected and whether there would be any benefit or risk by participating in data collection). The parents would then inform the head teacher or the school Community Education Committees (CECs) whether they did not consent to their child's participation in the study. Additionally, the day of the data collection, students were informed by their class teachers of the purpose of the data collection, and the information to be collected. Then, students were read the consent form aloud, and had a chance to read it independently, as it was attached to the survey (first page). Students were invited to return the questionnaire "empty" if they did not consent to participate in the study.

The implementation of the three waves of the survey, and the collection of exam data and teacher-generated student information in the Second and Third Endlines faced significant logistical and institutional challenges. A major difficulty was the absence of school-level student rosters, which would have enabled us to uniquely identify all students within a given grade, assess the uptake of the intervention (based on school attendance during the role model visits), and track students over time. Without these rosters, we are unable to determine whether the students for whom we have survey data represent a truly representative sample of the overall student population in the sampled schools. Importantly, in the Second and Third Endlines, we can only identify treatment status based on "current" grade at the time of the follow-up data collection. For instance, when analyzing the outcomes for the graduating class of 2022 (i.e., grade 8 at that time), we must assume that the graduating students were enrolled in grade 4 at the same school in May 2018, when the intervention took place. If grade 4 in their current school was visited by a female (or male) role model in 2018, we classify them into the "Female (Male) Role Model" treatment group. If their current school was a control school in 2018, we consider them control students, assuming they attended that school in 2018. However, these assumptions may not hold if students transferred from another school, a likely scenario given that 54 percent of the primary schools in our sample were not offering grade 8 at the time of the intervention. Additionally, in treatment schools, grade retention poses another challenge, as a grade 8 student in 2022 (for example) might have been a grade 5 (i.e., a non-treated grade) student in May 2018.

Given these caveats, we can only estimate the intent-to-treat effect of the intervention. In

addition, we recognize the potential for severe attenuation bias due to the mis-classification of an unknown proportion of students' treatment status, especially as more time passes between the intervention and the endline data collection.

Another challenge we face is the dynamic institutional structure of schools in our setting. The number of grades offered by a school can change from year to year. For instance, at the time of intervention, 21 out of the 46 schools were offering 8 grades. By the time of our second endline, this number had increased to 32 schools. Relatedly, the lack of a uniform structure among schools at baseline affected our randomization strategy. We could not pre-determine which *same* four grades would receive the interventions in all the treatment schools. This implies that we could not structure the randomization to ensure a balanced allocation of male and female role models within each treated grade. As a result, we were left with an un-balanced treatment assignment for the baseline grade 4 classes, which became apparent in the Third Endline, where we were left with only four baseline grade 4 classes, (and too few survey students from these classes) assigned to male role models. Five other grade 4 classes had been treated by male role models and a further two grade 4 classes were assigned to female role models, but they were in schools with fewer than 8 grades, hence not in our follow-up sample.

Additionally, our fragile setting – where displacement is common among students and permanent or mobile phone numbers are rare –prevents us from being able to follow and re-survey students after they graduate from primary school. To gather information on post-graduation outcomes such as high school enrollment, marriage, and fertility, surveying grade 8 teachers appeared to be the most effective strategy. Since this approach relies on teachers' familiarity with individual students (rather than school records), the post-graduation data are only available if the grade 8 teacher who was in charge of the 2020 and 2022 grade 8 cohorts was still working at the same school at the time of the survey. When the concerned class teacher is transferred to other regions or attrited for other reasons, this information is no-longer available. This affects five of the 2020 (baseline grade 6) cohort's schools and one of the 2022 cohort's schools.

3.4 Empirical Strategy

Our main interest is in estimating the impacts of male and female role models on male and female students. When analyzing survey data collected during the first endline (Fall 2018), i.e., data from students in baseline grades 4 to 7, we estimate the following equation:

$$Y_{igs,j} = \alpha + \beta_1 FRM_{gs,j} + \beta_2 MRM_{gs,j} + \beta_3 F_i * FRM_{gs,j} + \beta_4 F_i * MRM_{gs,j} + \beta_5 F_i + \gamma S_{gs,j} + \delta \mathbf{X}_i + \lambda_j + \zeta_g + \epsilon_{gs} \quad (1)$$

where Y_{igs} represents the outcome for child i in grade g in school s in pair j; $FRM_{sg,j}$ and $MRM_{sg,j}$ indicate, respectively, the Female Role Model and the Male Role Model treatment assignments for grade g in school s in pair j in Spring 2018. We include a dummy for the gender of student (F_i) and interact it with the two Role model treatment indicators. This allows us to test whether role models of different genders had differential impacts on boys and girls. We also report p-values generated by testing $\beta_1 + \beta_3 = 0$ and $\beta_2 + \beta_4 = 0$, i.e., the total impacts of female and male role models on girls, respectively. Since some students were (randomly) in untreated grades in treatment schools, we include an indicator S for being in what we call a "spillover grade" g in school s in pair j.

We control for child characteristics X that may also affect the outcomes of interest. Specifically, we control for student age, a proxy for poverty, i.e., whether the student reports having missed school in the past month due to "lack of money for fees, uniform or books" and a measure of exposure to female role models within the family, i.e., whether a female relative went to college.[24] Due to the pair matching randomization method we include pair fixed effects λ_j, as well as grade fixed effects ζ_g. Standard errors are clustered at the unit of randomization, which is the grade-school level.

For the analysis of treatment effects on the baseline grade 4 and grade 6 cohorts, based on the 2022 and the 2020 follow-up data collections, respectively, we estimate a similar model with some modifications. By design, every grade 4 and grade 6 student in Spring 2018 was treated in treatment schools, yet the gender of the role model was randomized across grades within each school. Since this analysis only involves one cohort per year (which is either treated or untreated, depending on the school), we cannot exploit grade-school variation in the empirical analysis. The standard errors are clustered at the school level, which is in this case the unit of randomization. Moreover, since we survey the target students when they were in grade 8 and about to graduate, the sample size is reduced to the set of schools that offered grades 1 to 8, and for which we have survey, exam and teacher-generated data, as explained in Section 3.2.3.[25]

[24]The First Endline survey also elicited additional information on family background, i.e., whether the student live with both parents, and whether each parent completed secondary school, which we report in Table I. However, these survey questions were not included in the Second Endline due to the need of keeping the questionnaire even shorter. Our 2018 results are robust to the inclusion of the additional individual controls in the specification.

[25]When generating our working sample for the post-graduation outcomes, we are affected by the fact that data on high school graduation, marriage and fertility for the 2020 graduating class are not available for all schools. This leads to a working sample of 27 schools for which we have all sources of data for the students

Given the small sample size and the small number of clusters, when assessing the impact of the role model intervention on education, marriage and fertility outcomes, we pool the data from the Second and Third Endlines, whilst controlling for student cohort. A shortcoming of this approach is that we are unable to test for the impact of the male role model intervention (due to lack of data for the 2022 graduating cohort). In the Appendix, we report the estimates separately for the two cohorts, which allows us to test for male role model effects on the 2020 graduating cohort.

In order to account for the small number of clusters, we follow Angrist and Pischke (2008) and Cameron and Miller (2015) and apply the correction for the small number of clusters by using wild cluster bootstrapping (Cameron and Miller, 2015). In our regression tables, we report wild-bootstrapped p-values in addition to clustered standard errors. In addition, given that we have two survey-generated outcomes, and four outcomes retrieved from schools and teachers, we correct the p-values associated to individual hypotheses belonging to the same family of outcomes by employing the step-down multiple testing method developed by Romano and Wolf (2005). Finally, in the Appendix, we also present separate regressions by gender, and confidence intervals for p-values based on randomization inference (Heß, 2017).

4 Results

4.1 Balance Tests and Attrition Tests

In Table I we report descriptive statistics of baseline school and student characteristics for control and treatment schools. In Table A1 we also conduct balance tests for schools and students assigned to male versus female role models within the treatment group. We start by reporting descriptive statistics for the full sample of 46 schools (baseline), and we then restrict the analysis to the schools for which we have First, Second and Third Endline survey data respectively. In the top panel of Table I we display school characteristics in the full sample, the 2018 survey sample (First Endline), the 2020 survey sample (Second Endline) and the 2022 survey sample (Third Endline). In the bottom panel, we report survey-generated student characteristics in treatment and control schools, corresponding to the First, second and Third Endlines.

The Spring 2018 (baseline) school-level data show no significant differences between treatment and control schools in any of the available school characteristics, including the number of students, the ratio of boys to girls, student-teacher ratio and the number of female teach-

who graduated in 2020, and a working sample of also 27 schools for the 2022 graduating cohort, as detailed in Section 3.2.3.

ers as well as whether the school charges fees and whether it has girl friendly spaces. At the time of the intervention, the schools were less than 10 years old, had 8 classrooms and about 32 students per teacher on average, averaging about 230 students and 9 teachers per school. Nearly 35 percent of the schools required the payment of school fees, and fewer than 20 percent had girl-friendly spaces. About 50 percent of the schools offered grades 1 to 8, with the average number of offered grades being 6.6. Teachers are primarily men, with fewer than 2 teachers per school out of 9 on average being women.

As discussed in Section 3.3, our working sample of schools becomes smaller over time. Our First Endline survey data are generated by 41 schools, and our Second and Third Endline by 32 and 28 schools, respectively.[26] Reassuringly, the school characteristics remain balanced between treatment and control arms with the decline in the number of schools. Additionally, the characteristics of the schools retained in subsequent surveys do not drastically change, implying non-selective attrition over our survey waves. Our attrition tests, displayed in Table A3 in the Appendix, show that attrited schools are smaller than schools in the follow-up working sample. However, this is a direct result of our data collection strategy, which, for the Second and Third Endline, required sampling schools that offered 8 grades, so that we could access data on exit exam performance and post-primary education outcomes. As a result, the attrited schools have fewer grades, hence fewer students, than the not attrited schools. Among other school characteristics, only the age of the school and the charging of school fees consistently predict the likelihood of attrition. Both variables, however, may be correlated with the size of the school; in fact, schools with 8 grades are older and more likely to charge school fees.

In the bottom panel of Table I we report student characteristics as generated by our follow up surveys. Recall that the First Endline survey was conducted only on students enrolled in grades 5 to 8 in Fall 2018, while the Second and Third Endline were conducted among students enrolled in grade 8 in Spring 2020 and Spring 2022, respectively. The implicit assumption is that the students surveyed in the First Endline were enrolled in the same schools, in baseline grades 4 to 7 when the intervention took place in Spring 2018. Similarly, we assume that the grade 8 students surveyed in 2020 and 2022 were enrolled in the same schools in grades 6 and 4 at the time of the intervention. We can partly verify the assumption regarding the First Endline data using information of our survey, where we asked students for their school and grade the previous academic year. Our data show that only 1 percent of students had changed schools the summer immediately following intervention, which is reassuring. We are, however, unable to conduct the same checks in subsequent data collection waves.

[26] As previously discussed, the fall in the number of schools is due to targeting schools offering 8 grades, and the exclusion of schools that returned only a handful of student surveys.

From the First Endline, we have data for a total of 1,941 students with non-missing information for any of the variables in the sample,[27] of which 846 were in control schools and 1,095 in treated schools. From the Second Endline, we have data on 829 students (348 in control and 481 in treatment schools) enrolled in grade 8 at the time of the survey, hence likely in baseline grade 6 at the time of the intervention. From the Third Endline, we have data from 575 students (253 in control and 322 in treatment schools), all enrolled in grade 8 in Spring 2022, hence likely in baseline grade 4 at the time of the role model visits.

The descriptive statistics and corresponding tests on the First Endline data show some imbalances in students' likelihood to have missed school due to poverty[28], which is lower in the treatment schools (9 percent of students) than in the control schools (14 percent of students). The surveyed students in the treatment schools are also more likely to have a close female relative who attended college, but the percentages (and the differences) are very small for both groups, i.e., 6 percent of students in the control schools and 8 percent of students in the treatment schools. We do not see imbalances in students' age, gender composition and parents' education.[29]

The student-level balance tests for the 2020 and 2022 graduating cohorts show some small differences in student ages, with the treatment students being slightly younger (14.93 vs. 15.31 years old in the Second Endline, and 15.46 versus 15.74 years old in the Third Endline). The follow-up data also suggest that the graduating students in the treatment schools were more likely to have missed class due to poverty. In the Second Endline, we also find fewer girls in treatment schools compared to the control schools. We control for these imbalanced students' characteristics in our main regression specifications. Appendix Tables A1 and A2 report additional balance tests for students in male versus female role model classes within the treatment schools, showing no significant imbalances in student characteristics.

4.2 Impacts on Gender Attitudes and College Aspirations

We start by presenting the differences in gender attitudes and college aspirations observed in our First, Second and Third Endline data. In the top panel of Figure I, we report the simple (non-standardized) mean of the answers to the four questions constituting the gender attitude index, described in Section 3.2, by treatment status and student gender. The index

[27] We have missing age data for 251 students. We imputed their age using the average seen in the sample, hence increasing our sample size from 1690 to 1941 students.

[28] This is an indicator equal to 1 if the student stated that he/she missed school because he/she needed to work for pay or did not have money to pay for fees, books or other school costs.

[29] About 16% of students had mothers who had completed secondary school and 24% had fathers who had completed secondary school, with no significant differences by treatment status.

ranges from 1 to 4, with higher values indicating more gender equal attitudes. The bottom panel of the figure shows college aspirations, measured as the percentage of students, in each treatment group and data collection wave, who answered "college" when asked what was the highest level of education that they would like to complete "if nothing could stop them."

The figure suggests that the intervention, and in particular the classroom visits by female role models, affected boys' and girls' gender attitudes positively in 2018 and 2020, with a non-significant difference between treatment and control in 2022. We note also that gender attitudes overall are slightly more egalitarian for girls than boys, and that for both genders, attitudes become more egalitarian over time.

The bottom panel of Figure I shows that the vast majority of students in the First and Second Endline (about 70 percent with no significant differences across genders) aspired to pursue a college education. This value however dropped quite dramatically, to 41 percent, The raw data show no significant and consistent effect of the treatment on aspirations.[30]

Results from regression analysis, displayed in graphical form in Figure II as well as Table A4 in the Appendix, largely confirm these findings. The estimates for gender attitudes in Figure II are reported in standard deviations from the control mean. We find that classroom visits by female college students increased gender attitudes toward equality by about 0.36 standard deviations in the short term (First Endline in 2018), with no significant difference in the effects on male and female students. The gender gap observed in control schools, whereby female students hold attitudes that are about 0.28 standard deviations more toward equality as compared to male students, vanishes as a result of the female role model intervention (p-value = 0.251, Wald test for the sum of coefficients on Female and its interaction with Female RM). The results are robust to controlling for school and student characteristics, to including school-pair fixed effects (exploiting the randomization method we used to allocate schools to treatment and control groups), and to corrections for multiple hypothesis testing and for small numbers of clusters through wild bootstrapping. The First Endline data also show a marginally significant impact of the intervention on students in spillover classes, i.e., classes in treatment schools who did not receive role model visits.

In the middle and bottom panels of Figure II and columns 3 and 5 of Table A4 in the Appendix, we see the estimated role model impacts on the gender attitudes of the 2020 and the 2022 graduating students cohorts, who were surveyed in the Second and Third Endlines, respectively. We see that the estimated coefficient for the Female Role Model indicator remains positive but it becomes smaller over time (0.137 standard deviations in 2020 and

[30]The Second Endline data for the 2020 graduating class suggest a positive effect of female role models on boys' college aspirations.

0.10 standard deviation in 2022) and statistically insignificant.[31]

As a robustness check, we conduct randomization inference (Heß, 2017) and display the results in Table A5 (Columns 1 and 2) in the Appendix. This analysis confirms the findings that the role model visits had a significant impact on students' gender attitudes, and particularly those of boys [coefficient of 0.519 and p-value 95 percent confidence interval=0.000,0.054] six months post intervention. The coefficients become substantially smaller, and lose statistical significance, when focusing on the 2020 and 2022 graduating classes (baseline grades 6 and 4), surveyed 2 and 4 years post intervention, respectively.

Figure II and Table A4 in the Appendix report also the estimates obtained over time for college aspirations. We see no significant impacts of male and female role models on boys' and girls aspirations in the First Endline. These null results are also observed for the 2020 graduating cohort (baseline grade 6), in the Second Endline data. However, the college aspirations of the 2022 graduating girls, who were surveyed in the Third Endline (4 years post intervention) show evidence of a marginally significant impact of the female role model visits on their college aspirations.[32] The results are again confirmed using randomization inference in Table A5 (Columns 2 and 4) in the Appendix.

4.3 Impacts on Education Outcomes, Marital Status and Fertility

Our Second and Third Endline data collections allow us to examine the impact of the role model intervention on the education outcomes of the surveyed students, through access to their performance on the standardized exam that all students have to take and pass to be able to graduate from primary school. In addition, through surveys of the grade 8 teachers, we were able to record information on students' enrollment in secondary school and their marital status and fertility outcomes. However, it is important to note that this analysis is limited to two student cohorts: one surveyed in 2020 (baseline grade 6) and another in 2022 (baseline grade 4). In addition, it involves a reduced sample of schools, as we could only include schools that offered 8 grades and where we were able to survey teachers who had taught the graduated students, as detailed in Section 3.2.3.

Following the empirical strategy outlined in Section 3.4, we pool the data obtained for the 2020 (baseline grade 6) and the 2022 (baseline grade 8) graduating cohorts, and estimate

[31]The lack of impact on gender attitudes could be attributed either to the longer time-frame since the intervention, attrition, or the reduced power given the smaller sample size and only one grade surveyed.

[32]Recall that we are unable to analyse the effect of male role model in 2022 because only 4 schools with 8 grades – which we could therefore survey in an attempt to record outcomes for the 2022 graduating cohort – had their grade 4 classes assigned to a male role model at the time of the intervention. We only received 35 responses from the graduating students in these 4 schools, making it impossible to analyze male role model impacts using Third Endline data.

the impact of female role models on: i) the likelihood of sitting the grade 8 exam, ii) student performance in the exam (grade standardized around the control mean), iii) the likelihood of being enrolled in secondary school, and iv) the likelihood of being married and/or having children. In the Appendix, we report estimates obtained for the two cohorts separately (Table A8) and for the 2020 student sample augmented with students assigned to male role models and surveyed as part of the Second Endline (Table A9).

About 80 percent of students took the mandatory exit exam, and the intervention had no effect on this outcome. When looking at the 859 students who did take the exam across the two student cohorts, we see evidence of a marginally significant positive impact of the female role model visits on girls' exam performance (0.38 standard deviation increase, with wild-bootstrapped p-value = 0.084). The intervention also increased both boys' and girls' likelihood of enrolling in secondary school by 11 percentage points, i.e., a 12 percent increase over the enrollment rate observed in the control schools (88 percent). We also see in column 4 a significant and negative impact of the intervention on students' likelihood of being married or having children. While the size of the estimated coefficient is the same for boys and girls, the effect is statistically significant only for boys, and it suggests a 75 percent decline over the control mean (which is equal to 7.7 percent).[33] The randomization inference results presented in table A7 show similar coefficients, but reduced precision.

Table A8 in the Appendix presents results for the 2020 and 2022 graduating cohorts separately. Both cohorts exhibit a consistently positive coefficient for the female role model indicator on girls' exam grades and on boys' and girls' secondary school enrollment. However, these effects are not statistically significant in the subsamples, likely due to the reduced sample sizes. Table A9 expands the 2020 cohort analysis by including students exposed to male role models. Here we see some evidence of a positive impact of male role models on girls' secondary school enrollment, and a marginally significant negative impact on marriage.

5 Discussion and Conclusions

In Spring 2018, we implemented a role model intervention in primary schools (grades 1 to 8) in a fragile country. Students enrolled in selected grades in randomly chosen schools were visited by a college student who grew up in the same geographical area and had a similar

[33]We combine marriage and children outcomes, as the percentages of individuals who are married and those who have children are smaller than 10 percent, and those with children tend to be a subset of those who are married. In Table A6 in the Appendix, we analyze marriage and children outcomes separately, finding negative impacts of the intervention on both, though only the effect on marital status is statistically significant. This could however be due to power issues, as the sample size for the children outcome is reduced because not all surveyed teachers had this information on the graduated students.

background. Within the treated schools, we randomly selected grades who would receive a visit from either a male or a female role model. Six months after the intervention, at the beginning of the new school year, we conducted our First Endline survey, involving students enrolled in grades 5 and above in control and treatment schools, hence in grades 4 to 7 at the time of the role model visits. We found strong evidence that visits from female college students significantly impacted both boys' and girls' attitudes toward gender equality in education and labor market participation. The effect was large and robust. Female role models had no impact on college aspirations, and male role models had no impact on either gender attitudes or aspirations.

Two years after the intervention, in Spring 2020, we conducted our Second Endline survey, involving students who were about to graduate from primary school; these students were likely enrolled in grade 6 at the time of the role model visits. Four years after the intervention, in Spring and Summer 2024, we implemented our Third Endline, which also involved students who were about to graduate from primary school at that time, hence likely enrolled in grade 4 at the time of the role model visits. As part of our Second and Third Endlines, we also gathered students' performance in the standardized exit exam that all grade 8 students have to take in order to graduate. In addition, we surveyed the grade 8 teachers and recorded information on students' secondary school enrollment, marital status and children.

The advantage of focusing only on the graduating classes in our Second and Third Endlines is the availability of data other than those generated by our survey. The disadvantages are the smaller sample sizes and the inability to exploit grade-school variation in students' outcomes. Additionally, the absence of school-level student enrollment and attendance records prevents us from confirming whether the graduating students were enrolled in the same school during the intervention. In our analysis, we assume they were - treating the 2020 graduating class as the baseline grade 6 students and the 2022 graduating class as the baseline grade 4 students. However, school transfers are likely common, particularly since nearly 50 percent of our sampled schools in 2018 offered fewer than eight grades. As a result, our intent-to-treat estimates are likely biased downward due to potential misclassification of the treatment status of transfer students.

With these drawbacks in mind, our analysis of the longer-term impact of the role model intervention on survey-generated outcomes shows positive but smaller and statistically insignificant effects of female role models on gender attitudes in both the 2020 and 2022 graduating cohorts, two and four years after the intervention, respectively. While we continue to see no significant impacts on college aspirations in the 2020 cohort, we do find some evidence of a positive effect of female role models on girls' college aspirations in the 2022 cohort (baseline grade 4). Specifically, girls' intention to go to college, if "nothing could stop them,"

increases by 15 percentage point for girls graduating in 2022 and therefore exposed to female role models four years earlier; this is a 35 percent increase over a mean of 40 percent.

Our analysis of exit exam performance, secondary school enrollment and marital status for the 2020 and 2022 graduating cohorts shows that the intervention impacted girls' exam score (0.38 standard deviation increase), both boys' and girls' enrollment (13 percent increase), and (primarily) boys' marital status (74 percent decline). Despite the reduced sample size, our primary findings and the caveats with the analysis, are robust to corrections for multiple hypothesis testing and for small number of clusters.

In conclusion, our findings suggest that this relatively low-cost intervention could be a valuable addition to education programs in similar environments despite the logistical challenges of implementing a randomized role model intervention, and conducting follow-up fieldwork in a fragile context. The positive impacts on gender attitudes (in the short-term) and on educational aspirations and outcomes (in the longer term), particularly for girls, highlight the potential for role models to contribute to long-term educational and social gains in fragile settings.

References

Abadie, A. and G. W. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics 29*(1), 1–11.

Abdullahi, A., M. Ali, E. Kipchumba, and M. Sulaiman (2023). Supporting micro-enterprise in humanitarian programming: impact evaluation of business grants versus unconditional cash transfer. *Journal of African Economies 32*(4), 415–437.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Appadurai, A. (2004). The Capacity to Aspire: Culture and the Terms of Recognition. In V. Rao and M. Walton (Eds.), *Culture and Public Action*, Chapter 3, pp. 59–84. Palo Alto, California: Stanford University Press.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *science 335*(6068), 582–586.

Bernard, T., S. Dercon, K. Orkin, A. Taffesse, et al. (2014). *The future in mind: Aspirations and forward-looking behaviour in rural Ethiopia*. Centre for Economic Policy Research London.

Bjorvatn, K., A. W. Cappelen, L. H. Sekei, E. Ø. Sørensen, and B. Tungodden (2019). Teaching through television: Experimental evidence on entrepreneurship education in Tanzania. *Management Science 66*(6), 2308–2325.

Brar, R., N. Buehren, S. Papineni, and M. Sulaiman (2023). Rebel with a cause: Effects of a gender norms intervention for adolescents in somalia.

Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics 1*(4), 200–232.

Burde, D. and L. L. Linden (2013). Bringing education to afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics 5*(3), 27–40.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Dahl, G. B., A. Kotsadam, and D.-O. Rooth (2021). Does integration change gender attitudes? the effect of randomly assigning women to traditionally male teams. *The Quarterly journal of economics 136*(2), 987–1030.

Dhar, D., T. Jain, and S. Jayachandran (2019). Intergenerational transmission of gender attitudes: Evidence from India. *The Journal of Development Studies 55*(12), 2572–2592.

Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. *American economic review 112*(3), 899–927.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American economic review 91*(4), 795–813.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics 119*(4), 1249–1299.

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

Heath, R. and S. Jayachandran (2017). The causes and consequences of increased female education and labor force participation in developing countries. In *The Oxford Handbook of Women and the Economy*, pp. 345–367. Oxford University Press.

Heß, S. (2017). Randomization inference with stata: A guide and software. *The Stata Journal 17*(3), 630–651.

Imai, K., G. King, C. Nall, et al. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation. *Statistical Science 24*(1), 29–53.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Jayachandran, S. (2021). Social norms as a barrier to women's employment in developing countries. *IMF Economic Review volume 69*, 576—595.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Jensen, R. and E. Oster (2009). The power of tv: Cable television and women's status in india. *The Quarterly Journal of Economics 124*(3), 1057–1094.

Kazianga, H., D. Levy, L. L. Linden, and M. Sloan (2013). The effects of "girl-friendly" schools: Evidence from the bright school construction program in Burkina Faso. *American Economic Journal: Applied Economics 5*(3), 41–62.

Kearney, M. S. and P. B. Levine (2015). Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing. *American Economic Review 105*(12), 3597–3632.

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

La Ferrara, E., S. Gulesci, S. Jindani, D. Smerdon, M. Sulaiman, and H. P. Young (2021). A stepping stone approach to understanding harmful norms. Technical report, CEPR Discussion Papers.

Lim, J. and J. Meer (2019). Persistent effects of teacher-student gender matches. *Journal of Human Resources 55*(3), 809–835.

Ministry of Education (2022). Annual statistics yearbook 2021-2022. Technical report, Ministry of Education, Culture and Higher Education.

Ministry of Education, Culture and Higher Education (2017). Education Sector Strategic Plan 2018-2020. Technical report, Ministry of Education, Culture and Higher Education.

Morgenroth, T., M. K. Ryan, and K. Peters (2015). The motivational theory of role modeling: How role models influence role aspirants' goals. *Review of General Psychology 19*(4), 465–483.

Mulkeen, A., D. Chapman, J. DeJaeghere, and E. Leu (2007). *Recruiting, retaining, and retraining secondary school teachers and principals in Sub-Saharan Africa*. The World Bank.

Muralidharan, K. and K. Sheth (2016). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources 51*(2), 269–297.

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar.

Olivetti, C., E. Patacchini, and Y. Zenou (2020). Mothers, peers, and gender-role identity. *Journal of the European Economic Association 18*(1), 266–301.

Patnaik, A., G. Pauley, J. Venator, and M. Wiswall (2023). The impacts of same and opposite gender alumni speakers on interest in economics. *NBER Working Paper* (w30983).

Porteous, O. (2020). Research deserts and oases: Evidence from 27 thousand economics journal articles on africa. *Oxford Bulletin of Economics and Statistics 84*(6), 1235–1258.

Porter, C. and D. Serra (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics 12*(3), 226–54.

Ray, D. (2006). Aspirations, poverty, and economic change. *Understanding poverty 1*, 409–421.

Riley, E. (2024). Role models in movies: The impact of queen of katwe on students' educational attainment. *Review of Economics and Statistics 106*(2), 334–351.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Save the Children (2018). Review of Save the Children Norway's Education Programmes and Partnerships with National Government Institutions at Country Level. Technical report, Save the Children Norway.

Serra, D. (2022). Role models in developing countries. *Handbook of experimental development economics*.

UNDP (2019). Human development report 2019. Technical report, United Nations Development Program.

UNDP (2023). Human development report 2023/2024. *United Nations Development Programme Somalia*.

UNESCO (2022). Somalia education sector analysis, assessing opportunities for rebuilding the country through education. Technical report, Federal Government of Somalia, IIEP-UNESCO Dakar.

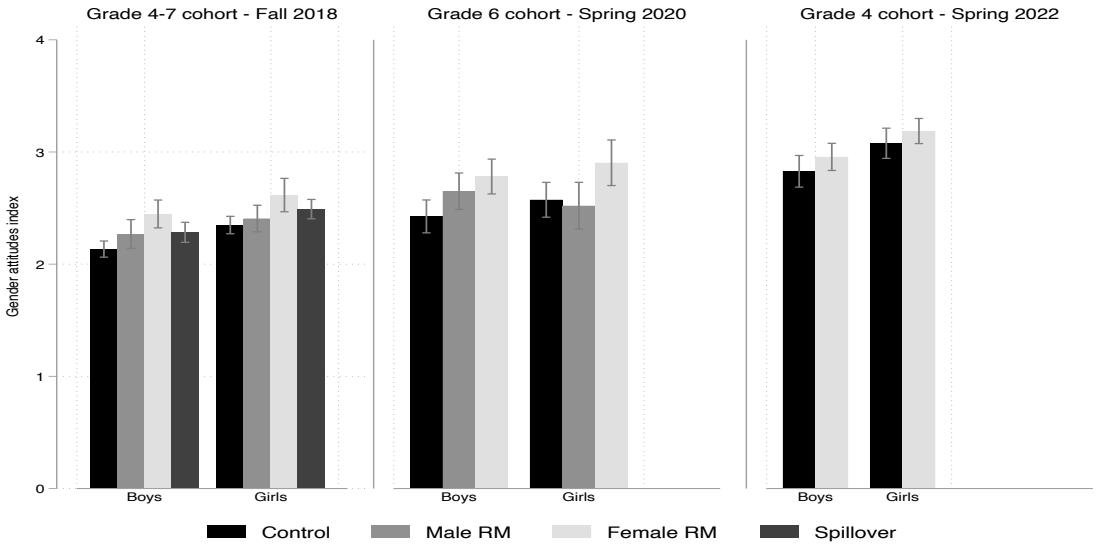
UNICEF (2017). Somalia education cluster annual report 2016. Technical report, UNICEF.

World Bank (2019). Somali poverty and vulnerability assessment. Technical Report Report No. AUS0000407, World Bank.

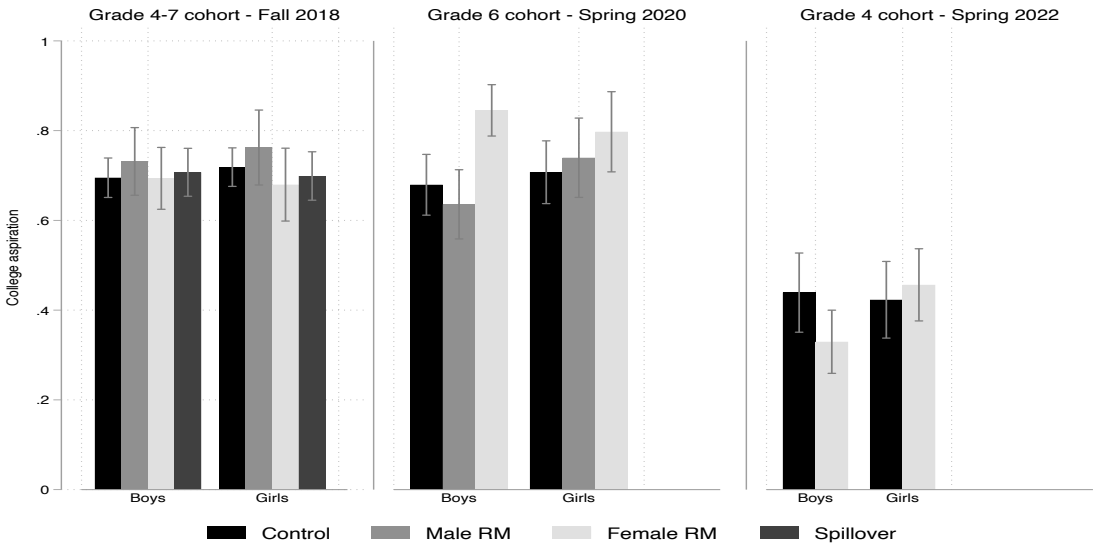
Figures and Tables

Figure I: Descriptive Statistics on College Aspirations and Gender Attitudes by Treatment

(a) Gender Attitudes

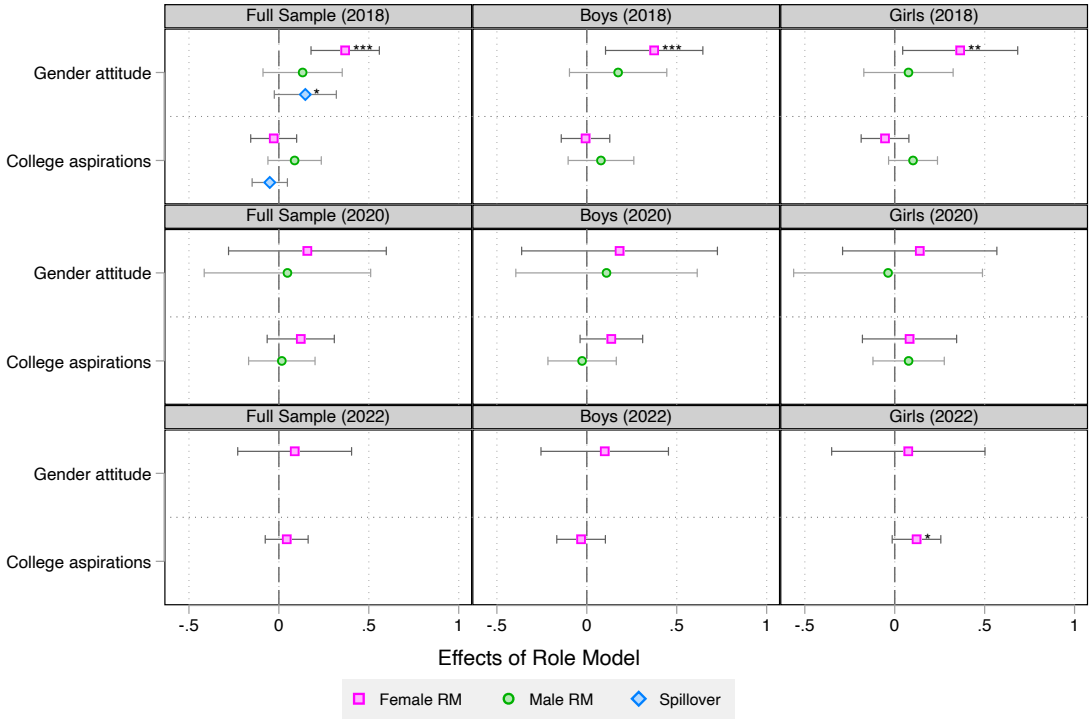


(b) College Aspiration



Note: The figure reports the outcome variables, by student gender and treatment group as measured in three separate survey waves. In the left panel, the outcomes were measured in Fall 2018, 6 months after the intervention (Spring 2018) among students who were enrolled in grades 4–7 in Spring 2018. The outcomes measured in Spring of 2020 among students who were enrolled in grade 6 in Spring of 2018 are in the middle panel while outcomes measured in Spring of 2022 are shown in the right-hand panel. The gender attitude index in the top panel is an aggregate of four survey questions, and ranges between 1 and 4, with higher numbers indicating preferences for more gender equality. College Aspirations in the bottom panel refer to the percentage of students who stated that the highest level of education they would like to achieve, if nothing would stop them, is college.

Figure II: Role Model Impacts on Gender Attitudes and College Aspirations



Note: The figure displays the estimated coefficents and 95 percent confidence intervals generated from the regression analyses reported in Table A4 in the Appendix. The gender attitude variable is standardized around the control mean, and therefore expressed in standard deviations from such mean. The top panel of the figure shows estimates obtained from the First Endline, which included students who were enrolled in grade 5 to 8 at the time of the data collection in Fall 2018, hence likely in grades 4–7 at the time of the intervention in Spring 2018. The middle panel displays estimates obtained from the Second Endline for the 2020 graduating cohort, and therefore includes student who were in grade 8 in Spring 2020 and likely in grade 6 at the time of the intervention. The bottom panel reports estimates obtained from the Third Endline for the 2022 graduating cohort, and includes students who were in grade 8 in Spring 2022 and therefore likely enrolled in grade 4 at the time of the intervention. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table I: Balance Tests

	Spring 2018			Fall 2018			Spring 2020			Spring 2022		
	Control	Treatment	Difference	Control	Treatment	Difference	Control	Treatment	Difference	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: School Characteristics												
School fees?	0.348	0.348	0.000	0.368	0.364	0.005	0.438	0.500	-0.062	0.438	0.500	-0.062
Age of school	9.087	8.217	0.870	9.316	8.455	0.861	9.812	9.562	0.250	9.812	10.833	-1.021
Has girl friendly space	0.174	0.174	0.000	0.211	0.136	0.074	0.188	0.125	0.062	0.188	0.000	0.188
Number of classrooms	6.522	8.913	-2.391	7.263	8.364	-1.100	7.125	9.125	-2.000	7.125	9.417	-2.292
Number female teachers	1.652	1.826	-0.174	1.474	1.773	-0.299	1.438	1.312	0.125	1.438	1.417	0.021
Total students enrolled	209.174	257.304	-48.130	209.158	263.773	-54.615	233.375	278.688	-45.312	233.375	297.333	-63.958
Boys enrolled	106.130	139.870	-33.739	106.947	143.545	-36.598	118.125	152.125	-34.000	118.125	161.917	-43.792
Girls enrolled	103.043	117.435	-14.391	102.211	120.227	-18.017	115.250	126.562	-11.312	115.250	135.417	-20.167
Student-teacher ratio	30.500	34.493	-3.993	28.709	35.770	-7.061	31.977	33.568	-1.591	31.977	33.579	-1.602
Number of grades	6.652	6.565	0.087	6.947	6.682	0.266	7.625	7.500	0.125	7.625	7.500	0.125
Observations	23	23	46	19	22	41	16	16	32	16	12	28
Panel B: Student Characteristics												
Baseline grade==4				0.080	0.102	-0.022*	0.000	0.000	0.000	1.000	1.000	0.000
Baseline grade==5				0.327	0.333	-0.006	0.000	0.000	0.000	0.000	0.000	0.000
Baseline grade==6				0.318	0.280	0.038*	1.000	1.000	0.000	0.000	0.000	0.000
Baseline grade==7				0.274	0.284	-0.010	0.000	0.000	0.000	0.000	0.000	0.000
Girl				0.500	0.464	0.036	0.471	0.364	0.107***	0.514	0.463	0.051
Student age				14.638	14.617	0.021	15.321	14.935	0.386***	15.743	15.456	0.287**
Missed school due to poverty				0.139	0.089	0.051***	0.164	0.222	-0.059**	0.032	0.102	-0.071***
Lives with parent(s)				0.882	0.906	-0.024*				0.874	0.907	-0.033
Father or mother completed secondary				0.242	0.260	-0.018				0.383	0.410	-0.027
Close female relative went to college				0.057	0.082	-0.025**	0.253	0.249	0.003	0.320	0.311	0.010
Observations				846	1,095	1,941	348	481	829	253	322	575

Note: Panel A shows school characteristics by treatment status using data originally obtained from Save the Children International prior to school randomization and implementation of the intervention. Panel B display student characteristics recorded through student surveys. The first three columns of Panel A refer to the entire universe of schools that were originally included in the study. In column 4-6, the data is restricted to the schools included in the First Endline, in column 7-9 to the schools included in the Second Endline, and in columns 10-11 to the schools included in the Third Endline. For each data collection wave, we report averages observed in treatment and control schools, and their differences. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. For additional balance tests for students in classes assigned to male versus female role models within the treatment group, see Tables A1 and A2 in the Appendix.

Table II: Impacts on exit exam, secondary school enrollment and marriage

	Sat exam	Exam score	Enrolled secondary	Married or has children
	(1)	(2)	(3)	(4)
Female RM	0.046	0.364	0.111	-0.057
	(0.042)	(0.294)	(0.052)**	(0.033)*
	{0.400}	{0.330}	{0.024}	{0.094}
	[0.347]	[0.208]	[0.020]	[0.089]
Girl x Female RM	-0.042	0.016	0.001	-0.000
	(0.057)	(0.185)	(0.046)	(0.033)
	{0.508}	{0.916}	{0.996}	{0.984}
	[0.594]	[1.000]	[1.000]	[1.000]
Girl	-0.003	0.061	-0.010	0.035
	(0.047)	(0.147)	(0.042)	(0.030)
Baseline grade==4	0.205	0.019	0.117	-0.082
	(0.041)***	(0.209)	(0.037)***	(0.029)***
Female RM + Girl × Female RM	0.004	0.381	0.113	-0.057
	(0.045)	(0.204)*	(0.056)*	(0.049)
	{0.938}	{0.084}	{0.068}	{0.272}
Grade fixed effects	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Control mean	0.787	0.006	0.877	0.077
Observations	1,071	859	859	859
Clusters	29	29	29	29

Note: The analysis includes the graduating cohorts of 2020 and 2022, who were likely enrolled in grades 6 and 4, respectively, at the time of the intervention in Spring 2018 and for whom we have exit exam data and teacher-generated data on secondary school enrollment and marital status. "Sat exam" is an indicator equal to 1 if the student took the grade 8 standardized exit exam, and 0 otherwise. "Exam score" refers to the grade obtained in the exit exam, conditional on having taken it. Exam scores are standardized around the cohort's control mean, and therefore expressed in standard deviations. "Enrollment secondary" is a 0-1 dummy variable equal to 1 if the student is enrolled to secondary school as reported by the corresponding grade 8 class teacher. Similarly, "married or has children" is a 0-1 dummy equal to 1 if the student is married or has had children, according to his/her grade 8 teacher. Since we cannot test for male role model effects for the 2022 graduating cohort (baseline grade 4), due to lack of data, this analysis excludes the 2020 graduating students who were assigned to male role models at the time of the intervention. Estimates obtained for the full 2020 graduating class are reported in the Appendix. Robust standard errors in parentheses, are clustered at school level. Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 APPENDIX

Figure A1: Map Showing Study Schools

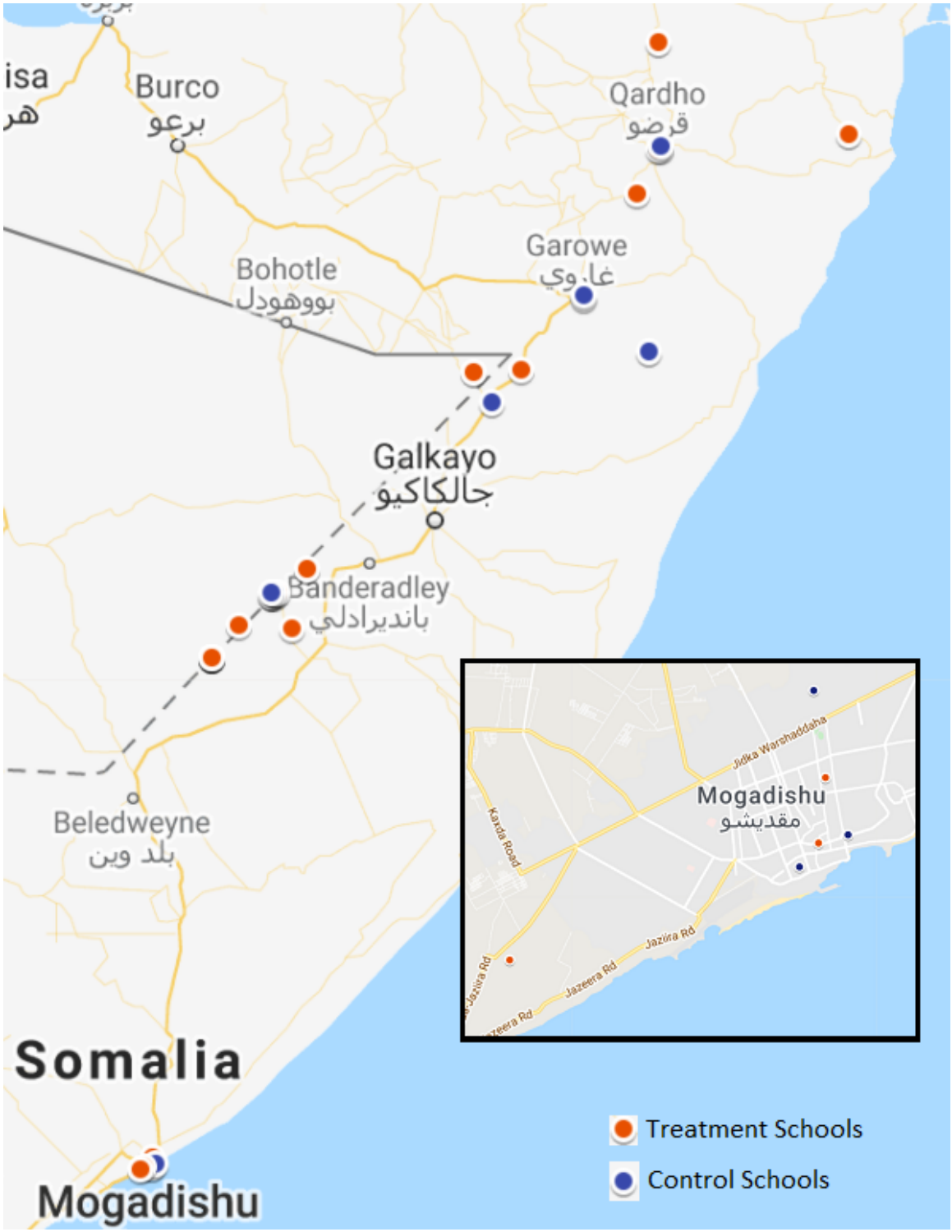


Table A1: Student Characteristics Balance by Role Model Gender (All)

	Female RM	Male RM	Spillover	P-value Female RM=Male RM	Control Mean
Panel A: Fall 2018					
Girl	-0.022(0.033)	-0.093(0.037)**	0.006(0.031)	0.100	0.500
Student age	-0.086(0.158)	0.018(0.226)	0.103(0.132)	0.675	14.638
Missed school due to poverty	-0.090(0.041)**	0.001(0.027)	-0.074(0.036)**	0.051	0.139
Lives with parent(s)	-0.003(0.021)	0.032(0.036)	0.005(0.026)	0.321	0.882
Father or mother completed secondary	0.004(0.040)	-0.057(0.052)	0.058(0.038)	0.270	0.242
Close female relative went to college	0.045(0.030)	0.000(0.022)	0.045(0.019)**	0.167	0.057
Panel B: Spring 2020					
Girl	-0.134(0.078)*	-0.083(0.067)		0.498	0.471
Student age	-0.484(0.448)	-0.293(0.414)		0.738	15.321
Missed school due to poverty	0.063(0.106)	0.055(0.120)		0.957	0.164
Close female relative went to college	-0.005(0.083)	-0.002(0.094)		0.978	0.253
Panel C: Spring 2022					
Girl	-0.051(0.036)				0.514
Student age	-0.287(0.272)				15.743
Missed school due to poverty	0.071(0.041)*				0.032
Lives with parent(s)	0.033(0.030)				0.874
Father or mother completed secondary	0.027(0.089)				0.383
Close female relative went to college	-0.010(0.055)				0.320

Note: Each row in each panel is a separate regression whereby the variable indicated in the row is the dependent variable, and treatment dummies are the key independent variables. For each regression, the omitted group is the control group. Regressions in Panel A also include baseline grade and randomization strata fixed effects. Standard errors in Panel A are clustered at school-grade level. Standard errors in Panel B and C are clustered at school level since the analysis only includes a single grade per school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Student Characteristics Balance by Role Model Gender (Working Sample)

	Female RM	Male RM	P-value Female RM=Male RM	Control Mean
Panel A: Spring 2020				
Girl	-0.235(0.091)**	-0.148(0.077)*	0.286	0.460
Student age	-0.941(0.647)	-0.279(0.458)	0.387	15.286
Missed school due to poverty	0.031(0.140)	0.134(0.166)	0.610	0.186
Close female relative went to college	-0.051(0.088)	0.111(0.104)	0.202	0.241
Panel C: Spring 2022				
Girl	-0.064(0.039)			0.518
Student age	-0.224(0.304)			15.728
Missed school due to poverty	0.073(0.040)*			0.023
Lives with parent(s)	0.033(0.033)			0.872
Father or mother completed secondary	0.015(0.093)			0.408
Close female relative went to college	-0.016(0.063)			0.317

Note: The analysis is restricted to surveyed students from schools for which we have exit exam results and teacher survey responses. These students constitute our working sample for the purpose of our evaluation of the role model intervention. Each row in each panel is a separate regression whereby the variable indicated in the row is the dependent variable, and treatment dummies are the key independent variables. For each regression, the omitted group is the control group. Standard errors in Panel A and B are clustered at school level since the analysis only includes a single grade per school. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Attrition Tests

	Spring 2020			Spring 2022		
	Not Attrited	Attrited	Difference	Not Attrited	Attrited	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Student Survey						
Treatment school	0.500	0.500	0.000	0.429	0.611	-0.183
School fees?	0.469	0.071	0.397***	0.464	0.167	0.298**
Age of school	9.688	6.286	3.402*	10.250	6.167	4.083**
Has girl friendly space	0.156	0.214	-0.058	0.107	0.278	-0.171
Number of classrooms	8.125	6.786	1.339	8.107	7.111	0.996
Number female teachers	1.375	2.571	-1.196*	1.429	2.222	-0.794
Total students enrolled	256.031	181.143	74.888*	260.786	190.389	70.397
Boys enrolled	135.125	95.286	39.839	136.893	101.389	35.504
Girls enrolled	120.906	85.857	35.049*	123.893	89.000	34.893*
Student-teacher ratio	32.772	31.865	0.908	32.664	32.236	0.428
Number of grades	7.562	4.429	3.134***	7.571	5.111	2.460***
Observations	32	14	46	28	18	46
Panel B: Exit Exam and Teacher Survey						
Treatment school	0.444	0.579	-0.135	0.444	0.579	-0.135
School fees?	0.519	0.105	0.413***	0.481	0.158	0.324**
Age of school	10.519	6.000	4.519***	10.556	5.947	4.608***
Has girl friendly space	0.148	0.211	-0.062	0.074	0.316	-0.242**
Number of classrooms	7.519	8.000	-0.481	7.963	7.368	0.595
Number female teachers	1.333	2.316	-0.982	1.296	2.368	-1.072
Total students enrolled	257.481	198.789	58.692	260.593	194.368	66.224
Boys enrolled	136.148	104.316	31.832	137.407	102.526	34.881
Girls enrolled	121.333	94.474	26.860	123.185	91.842	31.343*
Student-teacher ratio	32.291	32.788	-0.498	33.170	31.539	1.631
Number of grades	7.556	5.263	2.292***	7.593	5.211	2.382***
Observations	27	19	46	27	19	46

Note: We compare school characteristics in the schools for which we have data in the Second and Third Endline ("Not Attrited" in the table) and those for which we have no follow-up data ("Attrited"). In Panel A, attrition is defined based on the existence of grade 8 in the school, since the Second and Third Endline surveys target students graduating from primary schools. We were able to collect data from all 32 schools that offered 8 grades in 2020 and 2022. The "Not Attrited" sample fell to 28 in 2022, as we lost 4 schools from which we received only very few surveys back, as detailed in Section 3.2.3. In Panel B, attrition is defined based on the availability of exam data and teacher-generated outcomes for the surveyed graduating students. Here, the sample is further reduced due to the unavailability of the teachers who taught the graduating classes of 2020 or 2022. For more details, see the Section 3.2.3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Short-term, medium-term and long-term impacts on gender attitudes and aspirations

	2018		2020		2022	
	Gender attitude	College aspirations	Gender attitude	College aspirations	Gender attitude	College aspirations
	(1)	(2)	(3)	(4)	(5)	(6)
Female RM	0.375	-0.007	0.182	0.137	0.100	-0.032
	(0.137)***	(0.068)	(0.267)	(0.085)	(0.173)	(0.066)
	{0.016}	{0.920}	{0.578}	{0.242}	{0.586}	{0.702}
	[0.020]	[0.960]	[0.683]	[0.050]	[0.683]	[0.683]
Male RM	0.174	0.079	0.110	-0.026		
	(0.137)	(0.092)	(0.247)	(0.093)		
	{0.176}	{0.404}	{0.766}	{0.772}		
	[0.396]	[0.554]	[0.772]	[0.832]		
Girl x Female RM	-0.011	-0.047	-0.043	-0.054	-0.024	0.153
	(0.228)	(0.043)	(0.245)	(0.103)	(0.224)	(0.061)**
	{0.986}	{0.242}	{0.894}	{0.624}	{0.970}	{0.020}
	[0.970]	[0.515]	[0.832]	[0.772]	[0.842]	[0.030]
Girl x Male RM	-0.098	0.023	-0.147	0.103		
	(0.143)	(0.071)	(0.219)	(0.067)		
	{0.538}	{0.716}	{0.538}	{0.158}		
	[0.653]	[0.891]	[0.703]	[0.050]		
Spillover	0.147	-0.051				
	(0.087)*	(0.050)				
	{0.140}	{0.308}				
Girl	0.264	0.009	0.163	0.011	0.330	-0.031
	(0.070)***	(0.026)	(0.154)	(0.045)	(0.150)**	(0.047)
Female RM + Girl × Female RM	0.364	-0.054	0.139	0.082	0.076	0.121
	(0.161)**	(0.067)	(0.210)	(0.128)	(0.208)	(0.066)*
	{0.038}	{0.504}	{0.552}	{0.556}	{0.738}	{0.110}
Male RM + Girl × Male RM	0.076	0.102	-0.037	0.077		
	(0.125)	(0.069)	(0.257)	(0.097)		
	{0.556}	{0.166}	{0.888}	{0.470}		
Grade fixed effects	Yes	Yes	No	No	No	No
Pair fixed effects	Yes	Yes	No	No	No	No
School controls	No	No	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.000	0.707	-0.000	0.693	0.000	0.431
Observations	1,941	1,941	829	829	575	575
Clusters	117	117	32	32	28	28

Note: The estimates in columns 1 and 2 are generated by the First Endline data and includes students who were enrolled in grades 5 to 8 at the time of the survey (Fall 2018), hence in grades 4–7 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated by the Second Endline data, which includes the 2020 graduating cohort, i.e., students likely enrolled in grade 6 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated by the Second Endline data, which includes the 2022 graduating cohort, i.e., students likely enrolled in grade 4 at the time of the intervention, in Spring 2018. The gender attitude index averages 4 survey questions and is standardized around the control mean. Therefore, the associated estimates are expressed in standard deviations from the control mean. College aspirations are measured through a dummy variable, which is equal to 1 if the highest level of education that a student would like to complete if nothing could stop them is college. Robust standard errors in parentheses are clustered at the grade–school level (Column 1–2) and at the school level (Column 3–6). Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Randomization Inference: Short-term, medium-term and long-term impacts on gender attitudes and aspirations

	2018		2020		2022	
	Gender attitude	College aspirations	Gender attitude	College aspirations	Gender attitude	College aspirations
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Boys			
Female RM	0.519	0.011	0.009	0.113	0.083	-0.025
	(0.119)	(0.077)	(0.205)	(0.095)	(0.188)	(0.068)
	[0.000, 0.054]	[0.861, 0.971]	[0.901, 0.989]	[0.248, 0.442]	[0.611, 0.796]	[0.708, 0.873]
Male RM	0.059	0.043	0.118	0.009		
	(0.125)	(0.109)	(0.282)	(0.096)		
	[0.731, 0.890]	[0.643, 0.823]	[0.600, 0.788]	[0.874, 0.978]		
			Panel B: Girls			
Female RM	0.282	-0.195	0.135	0.030	0.165	0.111
	(0.133)	(0.065)	(0.222)	(0.103)	(0.219)	(0.066)
	[0.102, 0.258]	[0.022, 0.126]	[0.590, 0.779]	[0.708, 0.873]	[0.428, 0.631]	[0.094, 0.247]
Male RM	0.055	0.173	-0.108	0.110		
	(0.148)	(0.072)	(0.231)	(0.085)		
	[0.590, 0.779]	[0.029, 0.139]	[0.664, 0.840]	[0.257, 0.452]		

Note: The table displays estimates obtained from randomization inference, using the *ritest* command in Stata (Heß, 2017). The analysis is conducted separately for boys and girls. The estimates in columns 1 and 2 are generated from the First Endline, which targeted students who were enrolled in grades 5 to 8 at the time of the survey (Fall 2018), hence in grades 4–7 at the time of the intervention in Spring 2018. The estimates in columns 3 and 4 are generated by the Second Endline, which targeted the 2020 graduating cohort, i.e., students likely enrolled in grade 6 at the time of the intervention, in Spring 2018. The estimates in columns 3 and 4 are generated from the Third Endline, which targeted the 2022 graduating cohort, i.e., students likely enrolled in grade 4 at the time of the intervention in Spring 2018. All regressions include student-level controls. Standard errors in parentheses are clustered at grade–school level in Columns 1 and 2, and at the school level in Column 3 to 6, since the analysis in this case only includes a single grade per school. In square brackets, we report 95 percent confidence interval of p-value for treatment effect coefficient based on randomization inference.

Table A6: Separate Impacts on Marriage and Children

	Married	Has children
	(1)	(2)
Female RM	-0.058	-0.050
	(0.033)*	(0.046)
	{0.084}	{0.264}
	[0.030]	[0.178]
Girl x Female RM	-0.007	-0.001
	(0.031)	(0.035)
	{0.854}	{0.940}
	[0.861]	[0.960]
Girl	0.036	0.028
	(0.030)	(0.030)
Baseline grade==4	-0.082	-0.130
	(0.029)***	(0.061)**
Female RM + Girl × Female RM	-0.065	-0.051
	(0.048)	(0.063)
	{0.202}	{0.430}
Grade fixed effects	Yes	Yes
Pair fixed effects	Yes	Yes
School controls	No	No
Student controls	Yes	Yes
Control mean	0.077	0.069
Observations	859	627
Clusters	29	28

Note: The analysis includes the graduating cohorts of 2020 and 2022, who were likely enrolled in grades 6 or 4, respectively, at the time of the intervention in Spring 2018 and for whom we have exit exam data and teacher-generated data on secondary school enrollment and marital status. "Married" is a 0-1 dummy equal to 1 if the student is married according to his/her grade 8 teacher. "Has children" is a 0-1 dummy equal to 1 if the student is married according to his/her grade 8 teacher. Since we cannot test for male role model effects for the 2022 graduating cohort (baseline grade 4), due to lack of data, this analysis excludes the 2020 graduating students who were assigned to male role models at the time of the intervention. Robust standard errors in parentheses, clustered at school level. Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Randomization Inference: Exit exam, secondary school enrollment and marriage

	(1) Sat exam	(2) Exam score	(3) Enrolled secondary	(4) Married or has children
		Panel A: Boys		
Female RM	0.056	0.402	0.110	-0.058
	(0.037)	(0.340)	(0.057)	(0.027)
	[0.195, 0.379]	[0.204, 0.389]	[0.064, 0.200]	[0.056, 0.188]
		Panel B: Girls		
Female RM	-0.024	0.336	0.097	-0.042
	(0.050)	(0.201)	(0.054)	(0.051)
	[0.548, 0.743]	[0.071, 0.212]	[0.212, 0.400]	[0.788, 0.929]

Note: The table displays estimates obtained from randomization inference, using the *ritest* command in Stata (Heß, 2017). The analysis is conducted separately for boys and girls. The analysis employs data generated from the Second and Third Endlines, which followed the 2020 and 2022 student graduating cohorts (baseline grades 6 and 4), respectively. All regressions include student-level controls. Standard errors in parentheses are clustered at the school level, since the analysis only includes a single grade per school. In square brackets, we report 95 percent confidence interval of p-value for treatment effect coefficient based on randomization inference.

Table A8: Impacts on exam, secondary school enrollment and marriage by cohort

	Sat exam	Exam score	Enrolled secondary	Married or has children
	(1)	(2)	(3)	(4)
Panel A: Grade 6 cohort				
Female RM	0.042	-0.105	0.048	-0.048
	(0.086)	(0.299)	(0.056)	(0.047)
	[0.139]	[1.000]	[0.040]	[0.040]
Girl x Female RM	-0.091	0.354	0.063	0.000
	(0.137)	(0.536)	(0.108)	(0.088)
	[0.594]	[0.683]	[0.832]	[1.000]
Girl	-0.011	-0.031	-0.082	0.090
	(0.067)	(0.159)	(0.088)	(0.072)
Female RM + Girl × Female RM	-0.048	0.249	0.110	-0.047
	(0.121)	(0.505)	(0.082)	(0.081)
Grade fixed effects	No	No	No	No
Pair fixed effects	No	No	No	No
School controls	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Control mean	0.690	0.010	0.810	0.135
Observations	515	348	348	348
Clusters	21	21	21	21
Panel B: Grade 4 cohort				
Female RM	0.021	0.414	0.043	0.003
	(0.031)	(0.330)	(0.054)	(0.004)
	[0.921]	[0.723]	[0.772]	[0.921]
Girl x Female RM	-0.023	-0.105	0.014	-0.018
	(0.045)	(0.112)	(0.029)	(0.022)
	[0.861]	[0.634]	[0.921]	[0.772]
Girl	0.010	0.103	-0.009	0.026
	(0.029)	(0.090)	(0.027)	(0.021)
Female RM + Girl × Female RM	-0.002	0.309	0.057	-0.015
	(0.043)	(0.310)	(0.039)	(0.020)
Grade fixed effects	No	No	No	No
Pair fixed effects	No	No	No	No
School controls	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Control mean	0.936	0.001	0.950	0.014
Observations	556	511	511	511
Clusters	27	27	27	27

Note: Panel A reports the analysis conducted on the 2020 graduating students, who were likely enrolled in grades 6 at baseline and for whom we have survey data, exit exam data and teacher-generated data. Panel B reports the analysis conducted on the 2022 graduating students, who were likely enrolled in grades 4 at baseline and for whom we have survey data, exit exam data and teacher-generated data. "Sat exam" is an indicator equal to 1 if the student took the grade 8 standardized exit exam, and 0 otherwise. "Exam score" refers to the grade obtained in the exit exam, conditional on having taking it. Exam scores are standardized around the cohort's control mean, and therefore expressed in standard deviations. "Enrollment secondary" is a 0-1 dummy variable equal to 1 if the student is enrolled to secondary school as reported by the corresponding grade 8 class teacher. Similarly, "married or has children" is a 0-1 dummy equal to 1 if the student is married or has had children, according to his/her grade 8 teacher. Since we cannot test for male role model effects for the 2022 graduating cohort (baseline grade 4), due to lack of data, this analysis excludes the 2020 graduating students who were assigned to male role models at the time of the intervention. Robust standard errors in parentheses, clustered at school level. Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Baseline Grade 6 – Impacts on exam, secondary school enrollment and marriage

	Sat exam	Exam score	Enrolled secondary	Married or has children
	(1)	(2)	(3)	(4)
Female RM	0.036	-0.108	0.047	-0.053
	(0.080)	(0.279)	(0.057)	(0.045)
	{0.758}	{0.772}	{0.466}	{0.276}
	[0.931]	[0.931]	[0.931]	[0.554]
Male RM	0.013	0.179	0.090	-0.048
	(0.065)	(0.227)	(0.050)*	(0.034)
	{0.840}	{0.508}	{0.132}	{0.138}
	[0.960]	[0.931]	[0.109]	[0.317]
Girl x Female RM	-0.093	0.339	0.058	0.002
	(0.138)	(0.532)	(0.106)	(0.087)
	{0.594}	{0.732}	{0.624}	{0.990}
	[0.931]	[0.931]	[0.931]	[0.960]
Girl x Male RM	-0.022	-0.197	0.121	-0.104
	(0.106)	(0.248)	(0.089)	(0.070)
	{0.852}	{0.488}	{0.236}	{0.206}
	[0.960]	[0.931]	[0.317]	[0.257]
Girl	-0.014	-0.021	-0.080	0.090
	(0.064)	(0.169)	(0.086)	(0.071)
Female RM + Girl × Female RM	-0.057	0.232	0.105	-0.051
	(0.123)	(0.489)	(0.082)	(0.083)
	{0.688}	{0.756}	{0.226}	{0.586}
Male RM + Girl × Male RM	-0.008	-0.017	0.211	-0.152
	(0.071)	(0.308)	(0.079)**	(0.077)*
	{0.992}	{0.994}	{0.054}	{0.108}
Grade fixed effects	No	No	No	No
Pair fixed effects	No	No	No	No
School controls	Yes	Yes	Yes	Yes
Student controls	Yes	Yes	Yes	Yes
Control mean	0.690	0.010	0.810	0.135
Observations	709	476	476	476
Clusters	27	27	27	27

Note: The table reports estimates generated for the full sample of the 2020 graduating students (baseline grade 6) for whom we have survey data, exit exam data and teacher-generated data. These include students assigned to male role models while in grade 6, at the time of the intervention. "Sat exam" is an indicator equal to 1 if the student took the grade 8 standardized exit exam, and 0 otherwise. "Exam score" refers to the grade obtained in the exit exam, conditional on having taking it. Exam scores are standardized around the cohort's control mean, and therefore expressed in standard deviations. "Enrollment secondary" is a 0-1 dummy variable equal to 1 if the student is enrolled to secondary school as reported by the corresponding grade 8 class teacher. Similarly, "married or has children" is a 0-1 dummy equal to 1 if the student is married or has had children, according to his/her grade 8 teacher. Robust standard errors in parentheses, clustered at school level. Wild cluster bootstrap p-values in curly brackets. Romano-Wolf corrected p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$